12-1-2005

Testing Linearity in Cointegrating Relations with an Application to Purchasing Power Parity

Seung Hyun Hong

Peter C.B. Phillips

TESTING LINEARITY IN COINTEGRATING RELATIONS WITH AN APPLICATION TO PURCHASING POWER PARITY

By

Seung Hyun Hong and Peter C. B. Phillips

December 2005

COWLES FOUNDATION DISCUSSION PAPER NO. 1541

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.econ.yale.edu/

Testing Linearity in Cointegrating Relations with an Application to Purchasing Power Parity

Seung Hyun Hong*
Department of Economics
Concordia University
and
Peter C. B. Phillips†
Cowles Foundation, Yale University
University of Auckland and University of York

November 13, 2005

Abstract

This paper develops a linearity test that can be applied to cointegrating relations. We consider the widely used RESET specification test and show that when this test is applied to nonstationary time series its asymptotic distribution involves a mixture of noncentral χ^2 distributions, which leads to severe size distortions in conventional testing based on the central χ^2. Nonstationarity is shown to introduce two bias terms in the limit distribution, which are the source of the size distortion in testing. Appropriate corrections for this asymptotic bias leads to a modified version of the RESET test which has a central χ^2 limit distribution under linearity. The modified test has power not only against nonlinear cointegration but also against the absence of cointegration. Simulation results reveal that the modified test has good size in finite samples and reasonable power against many nonlinear models as well as models with no cointegration, confirming the analytic results. In an empirical illustration, the linear purchasing power parity (PPP) specification is tested using US, Japan, and Canada monthly data after Bretton Woods. While commonly used ADF and PP cointegration tests give mixed results on the presence of linear cointegration in the series, the modified test rejects the null of linear PPP cointegration.

JEL Classification: C12, C22

Key words and phrases: nonlinear cointegration, specification test, RESET test, noncentral χ^2 distribution

*Department of Economics, Concordia University, 1455 de Maisonneuve Blvd. West, Montreal, Quebec H3G 1M8, Canada; e-mail: shhong@alcor.concordia.ca

†Cowles Foundation for Research in Economics, Yale University, Box 208281, New Haven, CT 06520-8281, USA; e-mail: peter.phillips@yale.edu. Partial support from NSF grant 3 04-142254 is acknowledged.

1 Introduction

Economic time series are often believed to exhibit nonlinear behavior and economists usually formulate this nonlinearity in one of the following two ways: (i) by building nonlinear dynamics into the model for an individual time series; or (ii) by allowing for nonlinearity in the relationship between time series. This paper investigates issues associated with the second approach and does so from a model specification perspective. Since the introduction of the cointegration concept, linear models have dominated practical work in cointegration analysis. This emphasis has arisen, not so much because the underlying economic theory suggests linearity, but rather because the cointegration concept and associated econometric methodology has been developed very largely for linear models of integrated processes. Correspondingly, the tools of econometric analysis are available in this case and there is great convenience in computation for applied work. In contrast, until recently, there has been a lack of appropriate analytic tools for considering nonlinearly transformed integrated time series and an absence of an asymptotic theory of inference.

Empirical applications often stimulate an interest in nonlinear specifications and, in consequence, many nonlinear models (and almost as many specification tests) have been developed for stationary time series modeling. Many recent nonlinear model applications of nonstationary time series have focused on nonlinear short-run dynamics around linear long-run equilibria in error correction models (ECM), as in Berben & Dijk (1999), Lo & Zivot (2001), and Teräsvirta & Eliasson (2001) among others. Few attempts have been made to study nonlinear cointegrating relations directly and the methods that have been tried in practical work often require restrictive conditions on the DGP (e.g. Basher & Haug, 2003). Such extensions also await a corresponding development in tests of specification.

Neglecting the possibility of nonlinearity in a long-run relationship can be particularly detrimental in nonstationary cases. For stationary time series, linear models can often provide workable approximations at least locally to nonlinear models. Unlike mean-reverting stationary processes, nonstationary time series have a tendency to wander with no fixed mean or locality in the sample space and, like random walks, revisit points distant from the origin an infinite number of times. In such cases, local linear approximations can only poorly represent the global characteristics of the process, producing a high risk of faulty inference about a misspecified long-run equilibrium.

Considerations of the possibilities suggest three cases – linear cointegration, some form of nonlinear cointegration, or complete absence of cointegration. Existing cointegration tests essentially presume a form of linear cointegration and do not effectively distinguish between linear and nonlinear cointegration (Granger & Hallman, 1989). So, linear cointegration analysis requires an additional test of specification to address this particular issue. Existing linearity tests also fail to provide any guidance concerning the type of relationship that may be present between nonstationary time series if it were nonlinear (Granger, 1995). Accordingly, it is not surprising to find that linearity tests which have been developed for stationary processes work poorly with nonstationary time series. This was well recognized earlier in the case of the RESET test.

The RESET test (Ramsey, 1969) is a convenient device for testing general misspecification (e.g. Vitaliano, 1987; Baghestani, 1991; Peters, 2000, among others), but is known not to be robust to autocorrelated disturbances, especially when the regressor is itself highly autocorrelated (Porter & Kashyap, 1984) or contains a deterministic time trend (Leung & Yu, 2001). Using simulation experiments, Porter & Kashyap (1984) show that the presence of serially correlated disturbances combined with an $AR(1)$ regressor leads to size distortions, and the more autocorrelated is the regressor the less robust the RESET test is to error autocorrelation. Naturally, we might expect this size distortion problem to become worse in the cointegrating case where the regressor has an autoregressive unit root and the errors are typically serially dependent.

In the absence of more appropriate specification tests, applied economists have treated existing cointegration tests as tests for *linear cointegration*. In other words, if evidence for cointegration is found, it is conventionally assumed without further testing that such cointegration is linear in nature and all subsequent analysis rests on this assumption. The present paper develops a direct testing method that can answer the simple but important question: do the data support a linear cointegrating specification? To do so, we modify the widely used regression error specification test (RESET), which is a linearity test based on general approximation theory. The RESET test implicitly uses a Taylor series approximation to capture unspecified nonlinear forms by seeking to detect nonlinearities that remain in the linear regression residuals using a linear combination of polynomial functions. Our approach here is to use recently developed asymptotic tools for nonlinearly transformed integrated time series from Park & Phillips (1999, 2001) to modify the RESET test in an appropriate manner so that it can be applied to nonstationary time series to test directly for linearity in cointegrating relations.

The rest of this paper is organized as follows. The next section introduces the model and the maintained assumptions and shows how the testing method is related to an underlying theory of nonlinear approximation. In addition, we show how the nonstationarity of the data changes the limiting theory of the existing test using sample covariance asymptotics of nonlinearly transformed integrated processes. Section 3 discusses the modifications that are needed when the RESET test is applied to cointegrating relations. The asymptotic distribution of the modified test statistic is discussed both for the null and various alternative hypotheses. Simulation results on the finite sample size and the power of the modified test are reported in Section 4. Section 5 presents an empirical application of the modified test to purchasing power parity (PPP). Section 6 concludes and additional assumptions, lemmas and proofs are collected together in the Appendix.

2 Model and Background Ideas

The RESET test utilizes an approximate representation based on a power series expansion to determine whether the linear specification leaves anything unexplained in the regression residuals that can be detected by a linear combination of polynomial basis functions. The idea can be naturally extended to other families of basis functions and, as we will show, utilized in the context of nonstationary data applications.

For an arbitrary function $f(x)$ lying in a suitably defined L_2 function space over a certain domain, it is possible to construct an orthogonal series representation of the following form

$$f(x) \approx \sum_{j=1}^{\infty} \beta_j F_j(x) \quad \text{or in two parts} \quad f(x) \approx \sum_{j=1}^{k} \beta_j F_j(x) + \sum_{j=k+1}^{\infty} \beta_j F_j(x)$$
$$= \hat{f}_k(x) + error$$

in terms of some basis functions $\{F_j(x)\}$ that form a complete set and where $\approx$ signifies L_2 convergence. In the special case of a convergent power series (Taylor) representation, we may use simple polynomials as a basis. Given $f(x)$, the accuracy of the *finite sum approximation* $\hat{f}_k(x) = \sum_{j=1}^{k} \beta_j F_j(x)$ or the size of the "*error*" term depends on the number of terms (k) included in the sum and the properties of the function f, on which there is a huge literature in Fourier series analysis (e.g. Tolstov, 1976).

Suppose that we want to test the linear conditional mean specification $H_0 : P[E(Y_t|X_t) = \theta_1 X_t] = 1$, $\forall t$ using a regression specification

$$Y_t = \theta_1 X_t + \theta_2 f(X_t) + u_t. \tag{1}$$

If one has a specific nonlinear alternative model in mind, such as (1) for some given $f(X_t)$, and wants to test that specific model against the linear model, then one can use tests such as a Wald or LM test of $H_0 : \theta_2 = 0$ to decide which model fits the data better. In many practical cases, however, theory fails to provide a specific functional form, and while it is possible to come up with alternate nonlinear models, these often seem rather arbitrary. Also, if the focus of attention is some convenient linear model (such as that implied by purchasing power parity considerations) with no specific nonlinear alternative, then it is of particular interest to test whether the linear specification is "acceptable". A linearity test based on approximation theory seems appropriate in these situations.

Replacing the unspecified nonlinear function $f(X_t)$ with a partial sum approximation $\hat{f}_k(X_t)$, we may proceed to test the validity of the linear specification by testing whether a linear combination of a finite number of suitable basis functions $\{F_j(\cdot)\}_1^k$ can detect any nonlinearity in the regression residuals. This procedure involves testing $H_0 : \beta_j = 0$, $\forall j$ in the regression

$$Y_t = \theta X_t + u_t \quad \text{and} \quad \hat{u}_t = \sum_{j=1}^{k} \beta_j F_j(X_t) + e_t, \tag{2}$$

where the $\hat{u}_t$ are the regression residuals and the basis functions are chosen to be $F_j(X_t) = X_t^{j+1}$ for the RESET test[1]. As is apparent, using this approach there can be as many tests of linearity as there are approximation methods[2]. Here we will focus on the RESET test in view of its popularity in applied work.

Note that estimation of (2) involves working with the sample moments of nonlinearly transformed integrated time series of the form $\sum_t F_j(X_t)\hat{u}_t$ and $\sum_t F_j(X_t)F_i(X_t)$, whose asymptotic behavior must be characterized. Before examining these quantities, we first specify the data generating processes and some assumptions that will facilitate the development of a limit theory.

Assumption A: *Let $\triangle X_t = v_t$ and u_t be general linear processes satisfying the following*

[1] While the original test by Ramsey (1969) and the similar tests by Keenan (1985) and Tsay (1986) use the fitted value $\hat{Y}_t$, Thursby & Schmidt (1977) propose using the polynomials of X_t instead for a higher power.

[2] For example, DeBenedictis & Giles (1998) use a Fourier series approximation idea that has better global fitting capability than a Taylor series approximation, while Keenan (1985), Tsay (1986) and Barahona & Poon (1996) use variations of Volterra series expansions. White (1989) presents a neural network (NN) test based on the cdf of the logistic distribution and Blake & Kapetanios (2000) develop another NN test using radial basis functions for artificial neural networks.

conditions.

$$u_t = \sum_{j=0}^{\infty} c_j \varepsilon_{t-j} = C(L)\varepsilon_t \qquad v_t = \sum_{j=0}^{\infty} d_j \eta_{t-j} = D(L)\eta_t$$

$$\tilde{\zeta}_t = \begin{pmatrix} \eta_{t+1} \\ \varepsilon_t \end{pmatrix} \text{ is a stationary and ergodic martingale difference sequence}$$

$$\text{with natural filtration } \mathcal{F}_t = \sigma\left(\{\tilde{\zeta}_s\}_{-\infty}^{t}\right) \text{ and variance matrix } \Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}$$

and where $\{c_j, d_j\}$ *satisfy the conditions*

$$D(1) \neq 0 \ , \ \sum_{j=0}^{\infty} j|d_j| < \infty \ , \text{ and } \sum_{j=0}^{\infty} j^{1/2}|c_j| < \infty$$

These assumptions on the innovation processes are fairly standard, although in some cases below the linear process X_t is assumed to be predetermined in the sense that $E(F_j(X_t)|\mathcal{F}_{t-1}) = F_j(X_t)$. Conditions similar to these assumptions and Assumption A1 in the Appendix (an additional technical moment condition) are employed in deriving the results of Park & Phillips (1999, 2000, 2001) and Chang, Park & Phillips (2001). However, De Jong's (2002) more relaxed conditions are sufficient for the modification of the RESET test presented in this paper.

Under Assumptions A and A1, the following invariance principle holds

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]} \tilde{\zeta}_t \Rightarrow^d W(r) \equiv \begin{pmatrix} W_1(r) \\ W_2(r) \end{pmatrix} \equiv BM(\Sigma) ,$$

and using the Beveridge-Nelson decomposition (Phillips & Solo (1992)), we can show that a similar result holds for the time series $\zeta_t = [\,v_t, u_t]'$.

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]} \zeta_t \Rightarrow^d B(r) \equiv \begin{pmatrix} B_x(r) \\ B_u(r) \end{pmatrix} \equiv BM(\Omega) .$$

Here the covariance matrix $\Omega = \sum_{h=-\infty}^{\infty} \Gamma_\zeta(h)$, where $\Gamma_\zeta(h) = E(\zeta_0 \zeta_h')$. It is convenient to partition Ω conformably with ζ_t as

$$\Omega = \begin{pmatrix} \Omega_{vv} & \Omega_{vu} \\ \Omega_{uv} & \Omega_{uu} \end{pmatrix} , \tag{3}$$

and to define and partition the one-sided long-run covariance matrix

$$\Lambda = \sum_{h=1}^{\infty} \Gamma_\zeta(h) = \begin{pmatrix} \Lambda_{vv} & \Lambda_{vu} \\ \Lambda_{uv} & \Lambda_{uu} \end{pmatrix} , \tag{4}$$

$$\triangle = \Gamma_\zeta(0) + \Lambda = \begin{pmatrix} \Delta_{vv} & \Delta_{vu} \\ \Delta_{uv} & \Delta_{uu} \end{pmatrix} . \tag{5}$$

Among the wide variety of possible nonlinear functions, Park & Phillips (1999, 2001) provide asymptotic tools for certain classes of functions (of integrated processes) satisfying some regularity conditions. The simple basis functions $\{X_t^j\}$ of a Taylor series expansion fall within the H-regular (or Class H) class, which is defined as follows.

Definition 1 *A transformation $F(\cdot)$ is said to be H-regular iff*

$$F(\lambda x) = \kappa(\lambda) H(x) + R(x, \lambda)$$

where $H(\cdot)$ is locally integrable, and $R(\cdot, \cdot)$ is such that

- $|R(x, \lambda)| \leq a(\lambda) P(x)$, *where* $\limsup_{\lambda \to \infty} a(\lambda)/\kappa(\lambda) = 0$ *and $P(\cdot)$ is locally integrable, or*
- $|R(x, \lambda)| \leq b(\lambda) Q(\lambda x)$, *where* $\limsup_{\lambda \to \infty} b(\lambda)/\kappa(\lambda) < \infty$ *and $Q(\cdot)$ is locally integrable and vanishes at infinity, i.e. $Q(x) \to 0$ as $|x| \to \infty$*

Functions in this class have homogenous, asymptotically dominating components $\kappa(\lambda)H(x)$ that are locally integrable. $H(x)$ is referred as the *asymptotic homogenous function of* $F(x)$ and $\kappa(\lambda)$ as the *asymptotic order* of $F(x)$. Park & Phillips (1999) provide various examples that belong to this class, such as finite order polynomials, logarithmic functions, and distribution-like functions, including their linear combinations and products. The basis functions $\{F_m = X^{m+1}\}$ from a Taylor series expansion belong to this class with $H(x) = x^{m+1}$ and $\kappa(\lambda) = \lambda^{m+1}$.

Another important class of nonlinear transformation is the I-regular (or Class I) transformation. Roughly speaking, functions in this class are bounded, integrable and (piecewise) smooth (see Park & Phillips (1999) for further details). All pdf-like functions belong to this class.

2.1 A Linearity Test and Sample Covariances of Nonlinearly Transformed X_t

As a first step in the development, consider the simplest case where X_t is *strictly exogenous* so that Assumption A holds with $E(v_t u_s) = 0$ for all t, s and the long-run covariances $\Omega_{uv} = \Lambda_{uv} = 0$. Both OLS and FM-OLS (Phillips and Hansen, 1990) estimators of θ in (2) then yield consistent and asymptotically mixed normally distributed $\hat{\theta}$, and the RESET test statistic for $H_0 : \beta_j = 0, \forall j$ follows a limiting central $\chi^2(k)$ distribution, as we now show. That is

$$R_n = \left(\sum_{t=1}^{n} \hat{u}_t F_t\right)' \left(\hat{\Omega}_{uu.v} \sum_{t=1}^{n} \tilde{F}_t \tilde{F}_t'\right)^{-1} \left(\sum_{t=1}^{n} \hat{u}_t F_t\right) \overset{A}{\sim} \chi^2(k), \tag{6}$$

with

$$\tilde{F}_t = F_t - X_t \left(\sum_t F_t F_t'\right)^{-1} \sum_t X_t F_t \quad \text{for} \quad F_t = \left[\, X_t^2 \;\cdots\; X_t^{k+1} \right]', \tag{7}$$

and where $\hat{\Omega}_{uu.v}$ is a consistent estimator of Ω_{uu} under exogeneity. The test statistic R_n is a quadratic form of sample covariances between a nonlinearly transformed integrated process and the fitted residuals, viz.,

$$\sum_{t=1}^{n} \left(\frac{X_t}{\sqrt{n}}\right)^m \frac{\hat{u}_t}{\sqrt{n}} = \sum_{t=1}^{n} \widetilde{\left(\frac{X_t}{\sqrt{n}}\right)^m} \frac{u_t}{\sqrt{n}} \Rightarrow^d \int \widetilde{B_x^m}\, dB_u, \tag{8}$$

where the tilde over a variable implies that it is the residual from a linear regression of that variable on X_t for a finite sample, and equivalently for the limit processes where we write the projection residuals as $\widetilde{B_x^m} = B_x^m - B_x(\int B_x^2)^{-1} \int B_x^{m+1}$, where $\int$ denotes integration over $[0, 1]$

with respect to Lebesgue measure. The latter notation is used throughout the paper. The limit (8), conditional on $\mathcal{F}_x = \sigma(B_x(r), 0 \le r \le 1)$, is a mean-zero Gaussian mixture

$$\int \widetilde{B_x^m}\, dB_u \Big|_{\mathcal{F}_x} \sim \mathcal{N}\left(0, \Omega_{uu} \int \widetilde{B_x^m}^2\right),$$

so that, combined with the limit of the sample variance

$$\frac{1}{n}\sum_{t=1}^{n} \widetilde{\left(\frac{X_t}{\sqrt{n}}\right)^m} \widetilde{\left(\frac{X_t}{\sqrt{n}}\right)^m} \Rightarrow \int \widetilde{B_x^m}^2,$$

and a consistent estimator for Ω_{uu}, the RESET test statistic (6) has the following limit

$$R_n \Rightarrow^d \left(\int \widetilde{B_x^m}\, dB_u\right)' \left(\Omega_{uu} \int \widetilde{B_x^m}^2\right)^{-1} \left(\int \widetilde{B_x^m}\, dB_u\right)\Big|_{\mathcal{F}_x} \sim \chi^2(k)$$

conditional on $\mathcal{F}_x$. Since the limit distribution is independent of $\mathcal{F}_x$, we deduce that $R_n \Rightarrow \chi^2(k)$ unconditionally.

Next, we discard the "strong exogeneity" condition on X_t so that Assumption A holds with $\Omega_{uv} \neq 0$. Abandoning the strong exogeneity condition changes the previous result in several ways. First, the least squares estimator of θ now has two second order bias terms in the limit, viz.,

$$\begin{aligned} n\left(\hat{\theta} - \theta\right) &= \left(\frac{1}{n^2}\sum_{t=1}^{n} X_t^2\right)^{-1} \frac{1}{n}\sum_{t=1}^{n} X_t u_t \\ &\Rightarrow^d \left(\int B_x^2\right)^{-1} \left\{\int B_x\, dB_{u.x} + \Omega_{uv}\Omega_{vv}^{-1} \int B_x\, dB_x + \triangle_{uv}\right\}, \end{aligned} \tag{9}$$

where the Brownian motion $B_{u.x} = B_u - \Omega_{uv}\Omega_{vv}^{-1} B_x$ is independent of B_x and has variance $\Omega_{uu.v} = \Omega_{uu} - \Omega_{uv}\Omega_{vv}^{-1}\Omega_{vu}$. In (9) the term $\int B_x\, dB_{u.x}$ is a *mean zero* Gaussian mixture, but the other two terms in braces shift the mean of the limit distribution away from zero. These terms correspond to the so-called *endogeneity bias* and *serial correlation bias* of linear cointegration theory (Phillips & Hansen, 1990), and stem from the nonstationarity of X_t. Similar effects also generally arise with the sample covariance of u_t with nonlinear transformations of X_t. The following lemma summarizes the effects when the sample covariances involve polynomial functions.

Lemma 2 *Under Assumptions A and A1, the sample covariance between X_t^m and u_t satisfies*

$$\frac{1}{n^{(m+1)/2}}\sum_{t=1}^{n} X_t^m u_t \Rightarrow^d \int B_x^m dB_{u.x} + \Omega_{uv}\Omega_{vv}^{-1} \int B_x^m dB_x + m\Delta_{vu} \int B_x^{m-1} \tag{10}$$

with $\Delta_{vu} = \sum_{h=0}^{\infty} E\left(v_0 u_h\right)$.

Lemma 2 shows that, as in the linear cointegration case (9), sample covariances of nonlinear functions and stationary processes have limits that also involve two bias terms that produce nonzero location effects in the limit distribution of $\{\hat{\beta}_j\}$ in (2). We refer to these effects as "bias" terms because both terms shift the location of the limit distribution of $\hat{\beta}_j$ away from the true value β_j. The three components in the limit distribution in (10) are a mean-zero Gaussian mixture and two bias terms stemming from endogeneity and serial correlation effects, the latter being random when $m > 1$.

De Jong (2002) examines this nonlinear sample covariance asymptotics under conditions that are less strict on the innovation processes, but more restrictive in terms of functional forms. He shows that for an H-regular function $F(\cdot)$ with *continuously differentiable* asymptotic homogeneous function $H(\cdot)$, the sample covariance of $F(X_t)$ with u_t satisfies

$$\frac{1}{n^{1/2}\kappa_n}\sum_{t=1}^{n} F(X_t)u_t \Rightarrow^d \int H(B_x)dB_u + \triangle_{uv}\int H'(B_x) \tag{11}$$

where the asymptotic order of $H(\cdot)$ is $\kappa_n = \kappa(\sqrt{n})$. With $F(z) = z^m$, we have $H(z) = z^m$ and $\kappa_n = n^{m/2}$, so that (11) then reduces to (10) in Lemma 2. A recent and much more general semimartingale approach to establishing limit results such as (11) is developed in Ibragimov and Phillips (2004).

The effects from the bias terms in Lemma 2 can be substantial in the RESET test statistic (6). As is well known (e.g., Muirhead, 1982, theorem 1.4.5), a necessary and sufficient condition for the quadratic form $x'A\,x$ in the Gaussian random vector $x \equiv \mathcal{N}(\xi, V)$, where V is nonsingular, to follow a noncentral χ^2 distribution is that AV be idempotent. In this event $x'A\,x$ is noncentral $\chi^2(k,\nu)$, where $k = rank\,(AV)$ is the degrees of freedom and $\nu = \xi' A\,\xi$ is the noncentrality parameter. Here, we can replace x with the limit of $\sum_{t=1}^{n}\hat{u}_t F_t$ after an appropriate normalization and suitable conditioning, and thereby show that the test statistic R_n in (6) follows a mixture of noncentral χ^2 distributions, giving the following asymptotic result for the test when the exogeneity condition on X_t does not hold.

Theorem 3 *Under Assumptions A and A1, the RESET test statistic R_n has asymptotically a mixture noncentral χ^2 distribution with k degree of freedom and random noncentrality parameter $\nu = \xi' A\,\xi$. That is,*

$$\begin{aligned} R_n &= \left(\sum_{t=1}^{n}\hat{u}_t F_t\right)' \left(\hat{\Omega}_{uu.v}\sum_{t=1}^{n}\tilde{F}_t\tilde{F}_t'\right)^{-1}\left(\sum_{t=1}^{n}\hat{u}_t F_t\right) \\ &= \hat{u}'F\left(\hat{\Omega}_{uu.v}\tilde{F}'\tilde{F}\right)^{-1}F'\hat{u} \overset{A}{\sim} \chi^2(k,\nu) \end{aligned}$$

for $\tilde{F}_t$ defined in (7), $F = [\,F_1, \cdots, F_n]'$, $\tilde{F} = \left[\,\tilde{F}_1, \cdots, \tilde{F}_n\right]'$, and where $\hat{\Omega}_{uu.v}$ is a consistent estimator of $\Omega_{uu.v}$. The random noncentrality vector ξ is $k \times 1$ with $(m-1)^{th}$ element defined as

$$\xi(m-1) = \Omega_{uv}\Omega_{vv}^{-1}\int B_x^m dB_x + m\triangle_{vu}\int B_x^{m-1} - \Lambda_{vu}\left(\int B_x^2\right)^{-1}\int B_x^{m+1},$$

and A is a $k \times k$ covariance matrix

$$A = \left[\Omega_{uu.v} \begin{pmatrix} \int \widetilde{B_x^2}^2 & \cdots & \int \widetilde{B_x^2} \widetilde{B_x^{k+1}} \\ \vdots & \ddots & \vdots \\ \int \widetilde{B_x^{k+1}} \widetilde{B_x^2} & \cdots & \int \widetilde{B_x^{k+1}}^2 \end{pmatrix} \right]^{-1}$$

with $\widetilde{B_x^m} = B_x^m - B_x (\int B_x^2)^{-1} \int B_x^{m+1}$.

Remarks

1. In general, the mean and variance of a quadratic form $x' A\, x$ with a noncentral $\chi^2(k, \nu)$ are (e.g., Johnson & Kotz, 1978)

$$E(x' A\, x) = k + \frac{1}{2} \xi' A\, \xi \quad \text{and} \quad var(x' A\, x) = 2(k + \xi' A\, \xi).$$

So, conditional on $\mathcal{F}_x = \sigma(B_x(r), 0 \leq r \leq 1)$, the first two moments of R_n can be written as $k + \frac{1}{2}\nu$ and $2(k + \nu)$ respectively, and they are greater than the central $\chi^2(k)$ counterpart. This implies a higher probability of Type I errors, which explains the large size distortions observed in the simulation work by Porter & Kashyap (1984). We can check this by approximating the noncentral $\chi^2(k, \nu)$ distribution by a multiple of a central χ^2 distribution, $a\chi^2(b)$, where the two constants are given by (see Johnson & Kotz, 1978)

$$a = 1 + \frac{\nu}{k + \nu} \geq 1 \quad \text{and} \quad b = k + \frac{\nu^2}{k + 2\nu} \geq k.$$

Therefore, conditional on $\mathcal{F}_x$, the probability of rejecting the linearity null hypothesis can be written approximately as

$$P[R_n > \chi^2_\alpha] \sim P\left[\left(1 + \frac{\nu}{k + \nu}\right) \chi^2(b) > \chi^2_\alpha \right] \geq P\left[\chi^2 \left(k + \frac{\nu^2}{k + 2\nu}\right) > \chi^2_\alpha \right] \geq \alpha,$$

which is always at least as great as the nominal size α.

2. Originally, Ramsey (1969) introduced the RESET test as an F-test, where the test statistic can be written as

$$F_n = \frac{(\hat{u}'\hat{u} - \hat{e}'\hat{e})/k}{\hat{e}'\hat{e}/(n - K - k)} = \frac{\hat{u}'\,(I - M_F)\,\hat{u}/k}{\hat{u}' M_F \cdot \hat{u}/(n - K - k)}, \tag{12}$$

where $\hat{u}$ and $\hat{e}$ are the vectors of regression residuals from (2), K is the column number of X_t, and the $k \times k$ projection matrix $M_F = I - M_X F(F' M_X\, F)^{-1} F' M_X$ with $M_X = I - X(X'X)^{-1}X'$ and $X = [X_1', \cdots, X_n']'$. The numerator in (12) is equal to the χ^2 version of the RESET test R_n in Theorem 3 multiplied by $\hat{\Omega}_{uu.v}/k$. That is,

$$\frac{1}{k} \hat{u}'\,(I - M_F)\,\hat{u} = \frac{\hat{\Omega}_{uu.v}}{k}\, \hat{u}' F \left(\hat{\Omega}_{uu.v} F' M_X F \right)^{-1} F' \hat{u} \overset{A}{\sim} \frac{\Omega_{uu.v}}{k} \cdot \chi^2(k, \nu).$$

The denominator of (12) also can be shown to converge to $\Omega_{uu.v}/(n-K-k)$ times a mixture of noncentral χ^2 random variable $\chi^2(n-K-k,\nu')$ with noncentrality parameter $\nu' = E(\hat{u}' M_F \hat{u}|X)$. Therefore, conditional on $\mathcal{F}_x$, the ratio of two normalized noncentral χ^2 random variables follows a doubly noncentral F-distribution, denoted as $F(k, n-K-k;\nu,\nu')$, and

$$F_n \Rightarrow F(k, n-K-k;\nu,\nu') \quad \text{with } \nu = \xi' A\,\xi \text{ and } \nu' = \xi'(I-A)\,\xi, \tag{13}$$

for ξ and A defined in Theorem 3. Johnson & Kotz (1978) show that this doubly noncentral F-distribution can be approximated by a central F-distribution

$$F(k, n-K-k;\nu,\nu') \approx \left(\frac{1+\nu/k}{1+\nu'/(n-K-k)}\right) F(df_1, df_2),$$

with two degrees of freedom defined by

$$df_1 = \frac{(k+\nu)^2}{k+2\nu} \text{ and } df_2 = \frac{(n-K-k+\nu')^2}{n-K-k+2\nu'}.$$

Using this approximation, the probability of rejecting the linearity null hypothesis can be written as

$$P[F_n > F_\alpha] \sim P[F(df_1, df_2) > F_\alpha/C] \quad \text{with } C = \left(\frac{1+\nu/k}{1+\nu'/(n-K-k)}\right),$$

with random noncentrality parameters ν and ν'. Note that as long as $\xi \neq 0$, $C \Rightarrow 1+\nu/k$ as $n \to \infty$ and the noncentrality can therefore produce a substantial size distortion in the test.

3 Bias Correction and a Modified RESET Test

The previous section shows that nonstationarity of X_t introduces two bias terms in the limit distribution of the sample covariance between X_t^m and $\hat{u}_t$, so that the RESET statistic R_n is a limiting mixture of *noncentral* χ^2 distributions. These bias terms are the main source of the large size distortions in the test and we now present a method to remove them, leading to the modified RESET test whose limit distribution is central χ^2. The correction method is similar to that used in FM regression (Phillips and Hansen, 1990). After identifying sample quantities that converge to the bias terms, nonparametric corrections are implemented in the test statistic to eliminate them. For the first step, the following lemma introduces sample quantities that have the same limits as the bias terms.

Lemma 4 *Let Assumptions A and A1 hold. For $m \geq 1$,*

$$\hat{\triangle}_{vu} \frac{m}{n} \sum_{t=1}^{n} \left(\frac{X_t}{\sqrt{n}}\right)^{m-1} \Rightarrow^d m \triangle_{vu} \int B_x^{m-1}, \tag{14}$$

and

$$\sum_{t=1}^{n} \left(\frac{X_t}{\sqrt{n}}\right)^{m} \frac{v_t}{\sqrt{n}} - \hat{\triangle}_{vv} \frac{m}{n} \sum_{t=1}^{n} \left(\frac{X_t}{\sqrt{n}}\right)^{m-1} \Rightarrow^d \int B_x^m dB_x, \tag{15}$$

where $\hat{\triangle}_{vu}$ and $\hat{\triangle}_{vv}$ are consistent estimates of the long run covariance quantities $\triangle_{vu}$ and $\triangle_{vv}$.

Remarks

1. Lemma 4 provides sample quantities that converge to the asymptotic bias terms shown in Lemma 2. In linear cointegration, the corresponding terms for the two biases in (9) are given by

$$\hat{\Omega}_{uv}\hat{\Omega}_{vv}^{-1}\left(\frac{1}{n}\sum_{t=1}^{n}X_t v_t - \hat{\triangle}_{vv}\right) \Rightarrow^d \Omega_{uv}\Omega_{vv}^{-1}\int B_x dB_x \quad \text{and} \quad \hat{\triangle}_{uv} \Rightarrow^p \triangle_{uv}$$

   Denoting these two components as E_n and S_n, respectively, the sample covariance now becomes a mean zero Gaussian mixture in a linear case

$$\frac{1}{n}\sum_{t=1}^{n}X_t u_t - E_n - S_n \Rightarrow^d \int B_x dB_{u.x},$$

   and FM estimation simply applies these corrections, giving

$$\begin{aligned} n\left(\hat{\theta}_{FM} - \theta\right) &= \left(\frac{1}{n^2}\sum_{t=1}^{n}X_t^2\right)^{-1}\left\{\frac{1}{n}\sum_{t=1}^{n}X_t u_t - E_n - S_n\right\} \\ &\Rightarrow^d \left(\int B_x^2\right)^{-1}\int B_x\, dB_{u.x}\Big|_{\mathcal{F}_x} \sim \mathcal{N}\left(0, \Omega_{uu.v}\left(\int B_x^2\right)^{-1}\right) \end{aligned} \tag{16}$$

   We will use the same idea here to correct the bias terms in the test statistic R_n.

2. De Jong (2002) also recognizes the presence of two biases in nonlinear cointegrating regressions and gives the same expression as ours for the serial correlation bias correction in (11), but takes a different approach to correct the endogeneity bias. Noting that FM regression corrects the endogeneity bias by replacing $\int B_x dB_u$ with $\int B_x dB_{u.x}$, De Jong suggests a direct correction to the regression errors by using $u_t - \hat{\Omega}_{uv}\hat{\Omega}_{vv}^{-1}v_t$ instead of u_t.

The sample quantities and their limits shown in Lemma 4 are closely related to the noncentrality vector ξ defined in Theorem 3. Since the test statistic R_n is a quadratic form involving sample covariances of nonlinear functions, and the noncentrality parameter of its limit distribution correspondingly involves a quadratic form of ξ, we may eliminate the noncentrality by subtracting the sample quantities that converge to ξ from the nonlinear sample covariances. The following theorem explains how to accomplish this modification of the RESET test and remove the noncentrality.

Theorem 5 *Suppose Assumptions A and A1 hold. If $\{X_t, Y_t\}$ are linearly cointegrated, the following modified RESET statistic has a limiting central χ^2 distribution with degrees of freedom k*

$$MR_n = \left\{\hat{u}'F \cdot D_n - E_n' - S_n'\right\}\left(\hat{\Omega}_{uu.v}D_n'\tilde{F}'\tilde{F} \cdot D_n\right)^{-1}\left\{D_n \cdot F'\hat{u} - E_n - S_n\right\} \overset{A}{\sim} \chi^2(k),$$

where $\hat{u}$ is $n \times 1$ vector of residuals from the linear cointegration regression (2), F is an $n \times k$ matrix with the (m,t) element X_t^{m+1}, and $\tilde{F}$ is the regression residual from regressing F on X_t as in Theorem 3. The $k \times k$ normalization matrix D_n and the $(m-1)^{th}$ elements of the two $k \times 1$ correction vectors $E_n = [\ E_n(1), \cdots, E_n(k)]'$ and $S_n = [\ S_n(1), \cdots, S_n(k)]'$ are defined as

$$\begin{aligned}
D_n &= diag(n^{-3/2}, n^{-4/2}, \cdots, n^{-(k+2)/2}) \\
E_n\,(m-1) &= \hat{\Omega}_{uv}\hat{\Omega}_{vv}^{-1} \left[\left\{ \sum_{t=1}^{n} \left(\frac{X_t}{\sqrt{n}}\right)^m \frac{v_t}{\sqrt{n}} - \hat{\Delta}_{vv} \frac{m}{n} \sum_{t=1}^{n} \left(\frac{X_t}{\sqrt{n}}\right)^{m-1} \right\} \right. \\
&\qquad \left. - \left(\frac{1}{n}\sum_{t=1}^{n} X_t v_t - \hat{\Delta}_{vv}\right) \left(\frac{1}{n^2}\sum_{t=1}^{n} X_t^2\right)^{-1} \left(\frac{1}{n}\sum_{t=1}^{n} \left(\frac{X_t}{\sqrt{n}}\right)^{m+1}\right) \right], \\
S_n\,(m-1) &= \hat{\Delta}_{vu} \frac{m}{n} \sum_{t=1}^{n} \left(\frac{X_t}{\sqrt{n}}\right)^{m-1} - \hat{\Lambda}_{vu} \left(\frac{1}{n^2}\sum_{t=1}^{n} X_t^2\right)^{-1} \frac{1}{n}\sum_{t=1}^{n} \left(\frac{X_t}{\sqrt{n}}\right)^{m+1}.
\end{aligned}$$

Remarks

1. Although the RESET test is usually thought of as a general linearity test *without specific alternatives*, it also can be interpreted as an LM test, where the basis functions are treated as possible alternative nonlinear specifications. By construction, the test has highest power against such alternatives. Furthermore, if the test rejects linearity, the estimated nonlinear cointegration relationship provides an alternative nonlinear model, or more specifically a partial approximation to an alternative nonlinear model for the data, at least when the relationship is not spurious.

2. Since this type of test is based on a finite approximation method, power naturally depends on the adequacy of the approximation under the alternative. Note that the goodness of approximation depends on the given nonlinear functional form that is approximated, and the two components that can be controlled – the type and number of basis functions included in the augmented regressors. A good approximation will help in detecting nonlinearity when it is present, but even poor approximations can be effective. This is because the null hypothesis requires that all k coefficients be zero, $\beta_j = 0$ for $j = 1, \cdots, k$, and the test will reject if at least one coefficient deviates enough from zero, i.e. if one basis function is able to catch some "part" of the nonlinearity.

   Therefore, RESET test results should be interpreted conservatively: failure to reject the linearity hypothesis H_0 does not necessarily confirm a linear specification but rather that the relationship does not contain any nonlinearity that can be detected through the basis functions $\{F_j : j = 1, ...k\}$. The relationship between the power of the test and the choice of k is examined in the next section using Monte Carlo simulation.

3. The F-test version of the RESET test asymptotically follows a mixture of doubly noncentral F-distributions as in (13), where both random noncentrality parameters are quadratic forms in ξ. Since our modified test statistic in Theorem 5 implies that $E_n + S_n \Rightarrow \xi$, the bias correction method given above again can be used to construct a modified version of the F-test that has a limiting *central* F-distribution in a similar way.

In practice, the regressor set $\left\{X_t^{j+1}\right\}_{j=1}^{k}$ may suffer from multicollinearity, so that it is a common practice to use their principal components as the regressors.[3] In this case, the bias correction terms need to be adjusted accordingly, and the modified test statistic using principal components can be constructed as in the following corollary.

Corollary 6 *Suppose the conditions in Theorem 5 hold. Let G be the $k\times\tilde{k}$ matrix with $\tilde{k}$ eigenvectors of $F'F$ in its columns, after dividing by the corresponding eigenvalues. The modified test statistic MR_n based on these principal components follows a limiting central $\chi^2(\tilde{k})$ distribution as follows*

$$\left\{\hat{u}'F_n^* - E_n'G - S_n'G\right\}\left(\hat{\Omega}_{uu.v}\tilde{F}_n^{*\prime}\tilde{F}_n^*\right)^{-1}\left\{F_n^{*\prime}\hat{u} - G'E_n - G'S_n\right\} \overset{A}{\sim} \chi^2(\tilde{k}),$$

where $F_n^* = F\cdot D_n\cdot G$ is $n\times\tilde{k}$ normalized matrix with the j^{th} principal component in the j^{th} column. The $\tilde{k}$ eigenvectors are chosen such that the corresponding eigenvalues are the $\tilde{k}$ biggest ones.

3.1 The Modified RESET Test under Alternatives

As discussed earlier, considering nonlinearity together with nonstationarity gives rise to three possible scenarios. Our modified test tests the null hypothesis of linear cointegration against both *nonlinear cointegration* and the *absence of cointegration,* the latter incorporating both the conventional *spurious regression* case and *omitted variable* cases. As shown above, the test has a limiting central χ^2 distribution under linearity, and this subsection examines test power in these alternative scenarios.

3.1.1 The Case of No Cointegration

Kim, Lee, & Newbold (2003) show that many existing linearity tests tend to find spurious nonlinearity when they are applied to two independent $I(1)$ processes. They examine six widely used linearity tests–Ramsey's (1969) RESET test, White's (1989) NN test, the Keenan (1985) test, the McLeod and Li (1983) test, the White (1992) dynamic information matrix test, and Hamilton's (2000) flexible nonlinear test–and find that evidence of spurious nonlinearity increases with the sample size. The following Theorem shows that our modified test statistic also diverges when it is applied to two independent $I(1)$ processes. However, divergence of the test should not be interpreted as evidence of spurious nonlinearity but rather simply a rejection of the linear cointegration specification with two possible alternative cases. For nonstationary time series, a linearity test tests the linear (cointegration) specification against not only nonlinear cointegration models but also absence of cointegration. Therefore, a diverging test statistic in the no-cointegration case *correctly* points out the absence of linear cointegration. A further specification test is needed to determine if the rejection is due to nonlinearity.

[3]While this procedure is often necessary when the test is applied to stationary X_t, multicollinearity seldom arises when X_t is nonstationary. In contrast to mean-reverting stationary time series for which the variation of X_t around zero are dampened by polynomial transformations, integrated X_t spend little time around the origin and their variations are typically magnified by polynomial transformations as n increases.

Theorem 7 *Suppose X_t and Y_t are not cointegrated so that*

$$Y_t = \theta X_t + u_t \qquad t = 1, \cdots, n$$

with the $I(1)$ process u_t satisfying $n^{-1/2} u_{t=[n\cdot]} \Rightarrow^d B_u(\cdot)$. In this case the modified RESET statistic diverges at the rate of n/M, where M is the bandwidth parameter used in kernel estimation of the long-run (co)variances.

This result is of some practical interest. The RESET test was originally developed for testing linearity of the model but, when applied to cointegrating relations, the test has power against lack of cointegrating as well. Thus, the modified RESET test can serve as an *omnibus test* for the null of *linear cointegration* against the alternatives of both *no cointegration* and *nonlinear cointegration*.

A similar idea in the context of detecting unit roots is present in Park(1990)'s unit root test by variable addition. This test uses polynomials of a deterministic process as added variables to detect the presence of leftover stochastic trend(s), the RESET test uses polynomials of the stochastic regressors instead, which have a natural advantage when there is nonlinear cointegration involving these variables.

Since the rate of divergence depends on the relative size of the bandwidth parameter and the number of observations, the choice of M can greatly affect the power of the test against the lack of cointegration. Similar issues arise in other tests that rely on nonparametric estimates, such as the KPSS test for stationarity. We will discuss this issue in the next section together with other practical issues related to applying the modified RESET test.

3.1.2 The Nonlinear Cointegration Case

Among the many types of possible nonlinearities in cointegrated systems, we consider here models that involve transformations belonging to the H-regular and I-regular classes introduced earlier. In particular, we suppose the true cointegrated system has the following nonlinear form

$$Y_t = f(X_t) + u_t \quad , \quad t = 1, \cdots, n \tag{17}$$

where X_t and u_t satisfy Assumptions A and A1 with $f(\cdot)$ belonging either to the H-regular or I-regular nonlinear transformation class.

Theorem 8 *If the true model has the nonlinear form (17) and $\{X_t, u_t\}$ satisfy the conditions of Theorem 5, then the modified test statistic MR_n diverges at the rate $\frac{n}{M}$ in the H-regular nonlinear case, but does not diverge in the I-regular nonlinear case.*

Thus, the power of the modified RESET test depends on the nonlinear functional form. For H-regular nonlinearities, the test statistic diverges at the rate $O_p\left(\frac{n}{M}\right)$, just as in the case of no cointegration. Note that this result includes the case of a threshold model alternative, where the H-regular transformation is based on indicator functions. The asymptotic order in this case is $\kappa = 1$, as in the case of linear cointegration, but the test statistic still diverges in this case at the rate n/M.

Contrary to the H-regular case, the modified test has particularly low power against I-regular type nonlinearity. This is because the variations from the I-regular type nonlinear transformation of X_t that remain in the linear cointegration residuals $\{\hat{u}_t\}$ become negligible relative to the variations of X_t as n increases, while the variations in the basis functions $F_j(X_t)$ remain significant regardless of n.

Since H-regular and I-regular classifications do not exhaust all types of nonlinear transformations, there will be other types of nonlinear transformations that the modified test fails to detect. Whether a certain type of nonlinear cointegration is well detected by the modified test or not is related to the effectiveness of the partial sum approximation reflected in the augmented regressors $\{F_j\}_{j=1}^k$. Again, however, this test does not require a "good" fit to detect nonlinearity. If any polynomial term catches enough of the nonlinearity to make at least one of the fitted β_j coefficients significant, the modified test will have power in that direction to reject the null of a linear cointegration relationship.

3.2 Implications for Nonlinear Regression with Integrated Processes

The two bias terms in the linear cointegration regression (9) are called "second-order" in the sense that they cause bias only in the limit distribution, without affecting the consistency of LS estimator. The same argument applies to nonlinear cointegration case.

As FM regression (16) corrects the two biases using sample moments and sample estimates of the long-run (co)variances, Theorem 5 can be applied to correct the two biases in the LS coefficient estimator in the nonlinear cointegration regression. Suppose we estimate a nonlinear regression of the following form

$$Y_t = \theta f(X_t) + u_t \quad t = 1, \cdots, n$$

where $f(X_t) = X_t^m$. From Lemma 2 and Lemma 4 we can correct the second-order biases

$$\begin{aligned} n^{(m+1)/2}\left(\tilde{\theta}_m - \theta\right) &= \left(\frac{1}{n^{m+1}}\sum X_t^{2m}\right)^{-1}\left\{\frac{1}{n^{(m+1)/2}}\sum X_t^m u_t - E_n(m) - S_n(m)\right\} \\ &\Rightarrow^d \left(\int B_x^{2m}\right)^{-1}\int B_x^m dB_{u\cdot x} \end{aligned}$$

so that the modified estimator has a Gaussian asymptotic distribution around the true value. The two correction terms are defined as follows

$$E_n(m) \equiv \hat{\triangle}_{vu}\frac{m}{n}\sum_{t=1}^n\left(\frac{X_t}{\sqrt{n}}\right)^{m-1}, \quad S_n(m) \equiv \hat{\Omega}_{uv}\hat{\Omega}_{vv}^{-1}\left\{\sum_{t=1}^n\left(\frac{X_t}{\sqrt{n}}\right)^m\frac{v_t}{\sqrt{n}} - \hat{\triangle}_{vv}\frac{m}{n}\sum_{t=1}^n\left(\frac{X_t}{\sqrt{n}}\right)^{m-1}\right\}.$$

When $m = 1$, $\tilde{\theta}_m$ is simply the FM estimator in a typical linear cointegration model and the two correction terms $E_n(m)$ and $S_n(m)$ reduce to the usual form that appear in (16).

4 Simulations

Monte Carlo results are presented in this section to show the size distortion of the RESET test caused by nonstationarity and to investigate how satisfactory the suggested modifications are in achieving the nominal asymptotic size in finite samples. We also report some simulations on the power of the modified RESET test against some specific nonlinear models, choosing the following four models in addition to the linear cointegration model:

$$\begin{aligned}
(1) &: Y_t = 1.1X_t + u_t \\
(2) &: Y_t = \log(|X_t| + 1) + u_t \\
(3) &: Y_t = X_t^2 + u_t \\
(4) &: Y_t = 1.2\exp(-X_t^2) + u_t \\
(5) &: Y_t = 1.1X_t\, I_{\{|n^{-1/2}X_t| \geq 0.6\}} - 0.8X_t\, I_{\{|n^{-1/2}X_t| < 0.6\}} + u_t
\end{aligned}$$

Here, the linear model (1) is used as the reference case, (2) is a monotonically increasing, concave transformation in $\Re_+$ that is symmetric about the origin, (3) is an H-regular type nonlinear transformation that is often used to check the power of a certain test against nonlinear models, (4) is bell-shaped I-regular type nonlinear transformation, and (5) is a threshold model of a type that is commonly used in practical models of economic time series.

The regression error $\{u_t\}_{t=1}^n$ and the integrated regressor X_t are generated from the design

$$\begin{aligned}
\triangle X_t &= v_t = e_{2,t-1} + 0.4e_{2,t-2}, \\
u_t &= \rho u_{t-1} + \frac{1}{\sqrt{2}}\left(e_{1,t} + e_{2,t}\right),
\end{aligned}$$

where $\rho \in [0.2, 0.4, 0.6, 0.8]$ controls the level of serial correlation in the error term, and $\{(e_{1,t}, e_{2,t})\}_{t=1}^n$ are independently and identically distributed as

$$\begin{pmatrix} e_{1,t} \\ e_{2,t} \end{pmatrix} \sim \mathcal{N}\left(0, I_2\right). \tag{18}$$

Note that the innovation processes are constructed in such a way that X_t is predetermined, as specified in Assumption A. Samples of 5 different sizes ($n = 50$, 100, 250, 500, 1000) are drawn with 10,000 replications to examine both small sample properties and rate of convergence to the limit

4.1 Size of the Test

Fig. 1 compares two RESET tests–before and after bias corrections–when X_t and Y_t are linearly cointegrated. The four graphs summarize the test performance under H_0 from Table 1 with (a) varying number of observations for a given level of serial correlation and (b) varying level of serial correlation for a given number of observations. As shown in the upper panels (a), with a moderate level of serial correlation (AR coefficient is 0.6) in the regression error, the RESET test without correction terms shows severe size distortions that become even worse as the sample size increases.[4] For a nominal asymptotic 5% size, the actual probabilities of a type I error are 0.1058 (n=50), 0.1837 (n=100), 0.2817 (n=250), 0.3287 (n=500) and 0.357 (n=1000). This weakness of the RESET test is already well known in the stationary and highly autocorrelated X_t case from work of Porter & Kashyap (1984), and the results here for the case of a cointegrating

[4]The probability of a type I error increases when the regression errors are more serially correlated as shown in (b). In the extreme case of (independent) I(1) errors, the test statistic diverges, as reported in Kim, Lee & Newbold (2003).

relation may be regarded as an extreme version of these earlier findings. While the test without correction terms suffers from increasing type I errors, the modified RESET test in the right panel of Fig. 1 (a) exhibits only a small size distortion, which vanishes as n increases, and at the same time, shows a relatively fast convergence to the limit distribution. The probabilities of the type I errors for the nominal 5% test are 0.1932 (n=50), 0.0926 (n=100), 0.0483 (n=250), 0.0496 (n=500) and 0.0491 (n=1000).

Fig. 1 (b) shows how the bias correction terms work for different ρ values. The left panel confirms the severe size distortions due to the serially correlated errors. For a nominal asymptotic 5% size, the probability of a type I error reaches up to 70% for $\rho = 0.8$, while including two correction terms bring it back to 4.99%. These figures are based on Table 1 which compares two tests for different ρ's and n's with $k = 3$.

4.2 Power of the Test

Table 1 also reports the power of the modified RESET test against some specific nonlinear models. With linear cointegration as the reference case in (1), simulation results show that the modified RESET test is quite sensitive to all the nonlinearities except (4) for a wide range of ρ values. The probabilities of rejecting the linearity null are over 90% in most cases except for (4). As expected, the modified RESET test is most powerful against polynomial type nonlinearity (always higher than 99% in case (3)) but also shows good powers against logarithmic (2) and threshold (5) nonlinearities. Note also that the original RESET test in the second part also shows the similar pattern.

The low power against (4) is due to fact that the regression function is an integrable transform of X_t, which is poorly captured by the polynomial basis terms in the RESET test. In particular, the asymptotic form of the function $e^{-X_t^2}$ when $X_t = O_p\left(\sqrt{t}\right)$ for large t is not captured by the asymptotic form of the polynomial terms $X_t^j = O_p\left(t^{j/2}\right)$ in the RESET basis.

Table 2 shows the probability of rejecting the linear cointegration null hypothesis when the modified test is applied to two I(1) variables that are not cointegrated, i.e.

$$Y_t = 1.1X_t + u_t \quad \text{with } u_t \sim I(1)$$

As discussed in Theorem 7, the modified test statistic diverges at the rate n/M so that the rejection rate is sensitive to the choice of the bandwidth parameter M. We report five cases, corresponding to $M = n^{1/5}$, $M = n^{1/4}$, $M = n^{1/3}$, $M = n^{1/2}$ and the usual data-dependent automatic bandwidth (Andrews, 1991) for a Parzen kernel in Table 2. Two aspects of the results in Table 2 confirm Theorem 7. First, the rejection probability tends to higher for the smaller bandwidth choices for given k and n. Second, the rejection probability increases with n as well as with the number of augmented regressors k in general, espeically for smaller bandwidths. For $M = n^{1/3}$, the effect of increasing k on the rejection probability is not as large as in the case of $M = n^{1/5}$, and even decreases for $M = n^{1/2}$. When an automatic bandwidth rule is employed, increasing k has a more significant effect on power for a given n than increasing n for a given k.

4.3 Limitations and Practical Issues

The limitations of the modified RESET test are related to the approximation method that the test is based on and the nature of the nonlinear cointegration functional forms. As mentioned previously, once the nonlinear cointegrating function is given, the size of the approximation error is determined by the type and number of the basis functions $\{F_j\}_{j=1}^k$. These choices determine how well a linear combination of the basis functions can approximate some nonlinear cointegrating function of X_t. If there exists a set of coefficient $\{\beta_j\}_{j=1}^k$ such that $\sum_{j=1}^k \beta_j F_j(X_t)$ is close to $f(X_t)$ over a wide enough domain (since an $I(1)$ process like X_t visits all points of the space an infinite number of times), then we can expect the test to reject linear cointegration in favor of some form of nonlinear cointegration, corresponding to the non-zero $\{\beta_j\}$ estimates.

Once the basis functions $\{F_j\}_{j=1}^k$ are selected, k needs to be chosen. While larger k may produce an improved approximation to $f(\cdot)$, in a finite sample testing framework there exist some trade-offs. On the one hand, larger values of k will, at least to a certain point[5], generally increase the power of the test by virtue of their improved approximation capability. On the other hand, larger k increases the risk of spurious nonlinearity resulting in a higher probability of a type I error under the null. Moreover, to reject the null hypothesis $H_0 : \beta_1 = \cdots = \beta_k = 0$, at least one significant coefficient will suffice, a condition that is less restrictive than requiring a good fit to $f(X_t)$ by $\sum_{j=1}^k \hat{\beta}_j F_j(X_t)$. Simulations (not reported here) suggests that the use of $k = 2$ or 3 generally produces good size and reasonable power, while increasing k to $k = 3$ or 4 adds power without too much compromise in size.

Although not shown explicitly in the regression equation (2), the choice of bandwidth parameter M for kernel estimation of long-run (co)variance can be another important element that affects the size and the power of the test, especially in small sample. As discussed in Theorem 7 and shown in Table 2, the power against the no-cointegration alternative depends on n/M. The test statistic under the some alternatives diverges faster as M/n becomes smaller, but this makes the test statistic under the null converge to the asymptotic distribution at a slower rate. Therefore, in addition to the choice of k, it is recommended to apply the test with different combinations of k and M to get more a concrete result.

One popular choice for the bandwidth selection is the data dependent method in Andrews (1991). He proposes the automatic bandwidth choices for various kernels, and for the Parzen kernel we use, the automatic bandwidth is

$$M = 2.6614 \left[n \cdot \left(\sum \frac{4\hat{\rho}^2 \hat{\sigma}^2}{(1-\hat{\rho})^8} \right) \Big/ \left(\sum \frac{\hat{\sigma}^4}{(1-\hat{\rho})^4} \right) \right]^{1/5}$$

where $\hat{\rho}$ is the AR(1) coefficient estimate in $\hat{u}_t = \rho \hat{u}_{t-1} + e_t$ and $\hat{\sigma}^2$ is the variance estimate of e_t.

Another important factor that affects the power of the test is the actual nonlinear functional form. Although general approximation methods, including the power series approximations that underlie the RESET test, can provide reasonable approximations to a wide class of nonlinear functions, there are nonlinear transformations that cannot be well approximated by these methods. In particular, certain extensions to polynomial (or rational) approximants are generally needed in order to produce global approximations to functions over the whole real line. Phillips (1983) suggested a class of extended rational approximants that have good global approximant

[5]For k very large, the regressor matrix F can manifest multicollinearity and principal components may be used. In many cases, the first few principal components tend to explain most of variation in F and increasing k then leads to little improvement in the power of the test. Note that increasing k also leads to a decrease in degrees of freedom in the regression.

performance over the whole real line to integrable functions. One has to keep in mind that accepting the null of linear cointegration leaves open the possibility of some undetected nonlinear effects (especially if these are of the 'small' type that would be delivered by integrable transformations). Rejecting the null suggests that there may be nonlinear models that outperform the linear model or that there may be no meaningful cointegrating relation.

Estimated linear combinations of the basis functions can suggest a possible nonlinear alternative if the true relationship is nonlinear. In this case, as discussed in the previous section, the modified RESET test can be interpreted as an LM test which compares a linear cointegration model against an estimated approximation to some unknown nonlinear cointegration model. When the test rejects the null, we can write down an alternative nonlinear model with additional basis functions and re-estimate this model using FM regression. This leaves the remaining issues of choosing a suitable value of k for the regression so that the approximation error is reduced while not attempting to overfit the data. These issues are complex and are beyond the scope of the present paper.[6]

5 Empirical Application to PPP

The introduction of unit root limit theory and cointegration methods have led to a vast number of empirical studies with nonstationary time series, many of them conducted without further attention to specification testing beyond what is implied by unit root and cointegration tests. This section considers the purchasing power parity (PPP) relationship between nominal exchange rates and the foreign-domestic price ratio and applies the modified RESET linearity test to check whether the traditional linear cointegration specification is appropriate in this context.

5.1 PPP Models

PPP is a simple, intuitively appealing empirical proposition dated at least to the sixteenth century in Spain (Dornbusch, 1987). The theory postulates that once converted to a common currency, the price level of traded goods should be equalized across countries due to arbitrage. In this strict sense, the idea is sometimes understood as an extension of the law of one price (LOP). For a nominal exchange rate, S_t, a domestic price of a traded good i at time t, $P_{i,t}$, and the foreign price for the same good, $P^*_{i,t}$, the LOP states that the same good should be sold at the same price in different countries if prices are converted into a common currency (Rogoff, 1996)

$$P_{i,t} = S_t \cdot P^*_{i,t}.$$

Aggregating this relationship over traded goods, PPP states that

$$\sum_i P_{i,t} = S_t \cdot \sum_i P^*_{i,t}.$$

For a variety of reasons, this exact form of PPP, the so-called *absolute* PPP, does not hold and a weaker version of PPP is commonly used to provide a definition of the real exchange rate as

$$q_t = s_t + p^*_t - p_t,$$

[6] Of course, rejecting the null of linear cointegration may be due either to nonlinearity or to lack of cointegration. Developing an approximate nonlinear cointegrated system will be valid only when the rejection is due to nonlinearity.

where q_t and s_t are log transforms of real and nominal exchange rates, and p_t^* and p_t are log transforms of foreign and domestic price levels.

Intuitively accepted as providing a long-run equilibrium relationship among price levels and exchange rates, PPP has been tested in various frameworks, leading to some mixed empirical findings.[7] There have been many attempts to explain, using both economic and statistical arguments, the failure to find concrete empirical evidence for PPP.[8] For example, in the weaker version of PPP, the log of the real exchange rate q_t is usually divided into two parts: a traded goods component and a bilateral difference between the relative price of traded to non-traded goods, viz.,

$$q_t = s_t + p_t^* - p_t = \left\{ s_t + p_t^{T*} - p_t^T \right\} + \left\{ \alpha^*(p_t^{N*} - p_t^{T*}) - \alpha(p_t^N - p_t^T) \right\}$$

where the superscripts 'T' and 'N' stand for 'traded' and 'non-traded' respectively. The price indices are generally assumed to be geometric averages of traded and non-traded goods,

$$p_t = (1-\alpha)p_t^T + \alpha p_t^N \quad \text{and} \quad p_t^* = (1-\alpha^*)p_t^{T*} + \alpha^* p_t^{N*},$$

and, defining

$$P1 = s_t + p_t^{T*} - p_t^T \quad \text{and} \quad P2 = \alpha^*(p_t^{N*} - p_t^{T*}) - \alpha(p_t^N - p_t^T),$$

the real exchange rate is stationary either if $P1$ and $P2$ are stationary, or if $P1$ and $P2$ are nonstationary but cointegrated. Accepting PPP as a long-run equilibrium relationship, $P1$ is stationary and it is not at all surprising that many find the real exchange rate to be nonstationary considering the presence of the possibly nonstationary component $P2$.

Traditional unit root/cointegration approaches have been the most widely used method in PPP empirical studies, but these methods have often failed to find any strong support for PPP. These failures have led to the use of many new methods in searching for evidence of PPP, including longer datasets, panel unit root evaluations, and the use of nonlinear models. Noticing the low power of unit root tests in small samples, researchers have tested PPP using long-horizon data, finding stronger support for PPP (e.g. Lothian & Taylor, 1995) by this method. Many empirical researchers have found that the floating exchange rate system introduced with the Bretton Woods system has led to larger deviations from PPP (e.g. Taylor, 2002). Using cross-country data to improve the power of unit root tests has also tended to produce stronger support for PPP, but these methods have also been criticized by O'Connell (1998) and others for neglecting cross country dependence. While these first two methods have involved the use of different datasets to improve tests of PPP, the last approach takes into account the possibility of different model specifications. Nonlinear specifications are often obtained from market frictions like transaction costs, e.g. Dumas (1992), Sercu, Uppal & van Hulle (1995), and Michael, Nobay & Peel (1997). Because of market frictions, there exists an inactive range around parity in which international arbitrage does not work and adjustments to parity start to occur only when the exchange rate moves out of this range. This nonlinear adjustment to parity can be formulated using variations of the threshold model (e.g. STAR, ESTAR) and some significant empirical

[7]Froot & Rogoff (1995) provide a discussion of the evolution of PPP tests and Rogoff (1996) surveys empirical studies in the area.

[8]See Grilli & Kaminsky (1991), Pedroni (2001) and Ng & Perron (2002) for some statistical arguments and Sercu, Uppal & van Hulle (1995) and Rogoff (1996) for popular economic explanations.

evidence has been found in support of these models, a recent contribution being Saikkonen and Choi (2004) who use smooth nonlinear transitions.

In addition to the nonlinear short-run "adjustment" terms included in long-run linear equilibrium, Basher & Haug (2003) posit a nonlinear PPP relationship and apply a nonlinear cointegration test developed by Breitung (2001), but fail to find any linear and nonlinear cointegration relationship among the G10 countries. We use their model

$$S_t = \alpha + f\left(\frac{P_t^*}{P_t}\right) + u_t \tag{19}$$

and test for linearity in this cointegrating relationship between the nominal exchange rate and the ratio of foreign and domestic prices.

Not having a specific functional form for $f(\cdot)$ offers some advantages. First, even if the threshold model had strong theoretical justification for *one* tradable good, aggregating over all goods and using a general price level inevitably obscures the form of the implied nonlinearity for the aggregate relationship (for instance, because of the manifold threshold points that appear in the aggregation). Setting a regression equation in the general form of (19) allows for a more flexible interpretation. Apart from providing a testable form of PPP, (19) can be thought of as a general model of nominal exchange rate determination in terms of economic fundamentals. Although Meese & Rogoff (1983) find that no existing structural model outperforms a simple random walk model in prediction, the monetary model has been the standard model for exchange rate determination. This model's main implication is that the nominal exchange rate is determined by some economic fundamentals like money (m) and output (y) of the two countries, and the risk premium (ρ). Frankel & Rose (1994) show the following expression, first given by Mussa (1976), can be derived using money market equilibrium, PPP, and uncovered interest parity:

$$s_t = [(m - m^*)_t - \beta(y - y^*)_t + \rho_t] + \alpha \frac{E_t(\triangle s_t)}{dt} + \varepsilon_t$$

Using the price ratio to reflect the economic fundamentals, (19) can be regarded as expressing nominal exchange rates as some unknown function of underlying fundamentals.

In addition to the PPP in levels (or *absolute* PPP), we also test *relative* PPP which can be written as (Rogoff, 1996)

$$\frac{P_t}{P_{t-1}} = \left(\frac{S_t}{S_{t-1}}\right) \cdot \frac{P_t^*}{P_{t-1}^*}$$

Since the price index is the relative value to a base year and we do not know how big the deviation from absolute PPP was at the base year, this relative version of PPP requires the relationship to hold only in terms of changes. In this case, since the logarithms of the price and exchange rate ratios are stationary, we need to interpret empirical results appropriately.[9]

5.2 Data and Empirical Results

We consider three countries (US, Japan and Canada) forming the two pairs: US-Japan and US-Canada. We focus on these two pairs, which represent respectively a relationship between two

[9] Note that if the variables are stationary, our modified test becomes equivalent to the traditional RESET test as the bias and correction terms vanish asymptotically.

big economies and a relationship between countries with fewer trade barriers and transportation costs.

Our US, Japan and Canadian dataset is taken from the IMF's *International Financial Statistics* (IFS) CD-ROM and contains nominal exchange rates, the consumer price index (CPI) and producer price index (PPI) at a monthly frequency. The data spans the period from 1971:1 to 2004:12, yielding 34 years or 408 monthly observations. A monthly average market rate is used for the nominal exchange rate and both the CPI and PPI are used to calculate price ratios. The data are plotted in Figure 2. The left column shows the US-Canada: nominal exchange rate (solid), CPI ratio (dashed) and PPI ratio (dash-dotted), and the right column shows the US-Japan in the same manner. The upper panels plot the nominal exchange rates, CPI ratios and PPI ratios in levels (not in logs). The lower panels plot the same series but in changes calculated by year-to-year ratios, i.e. for the nominal exchange rate (S_t), S_t/S_{t-12} and for the CPI or PPI (P_t), $(P_t/P_{t-12})/(P_t^*/P_{t-12}^*)$.

First, we apply augmented Dickey-Fuller (ADF) tests to determine whether the time series plotted in Figure 2 are integrated processes (the Phillips-Perron test gave similar results). Test results (not reported here) indicate that the nominal exchange rate, CPI, and PPI are all unit root nonstationary in levels (for absolute PPP) and stationary in changes (for relative PPP). Second, we apply ADF and KPSS tests to the regression residuals with varying sample periods, to check whether these tests find any meaningful linear cointegration relationship. As much previous research has reported, conventional (linear) cointegration tests show somewhat mixed results (Table 3).

1. We first consider the whole sample period (1971M1–2004M12: Period 1) and then the post-Volcker period (1983M1–2004M12: Period 2).

2. The ADF test applied to US-Canada and US-Japan does not find evidence of any (linear) cointegration relationship between nominal exchange rate and the ratio of price levels (absolute PPP) with neither CPI nor PPI.

3. The KPSS test finds linear cointegration relations for the whole sample period but some of these are not supported by tests for the different sample period. Although not reported here, it is not hard to find a subsample period where the ADF test finds evidence of a cointegration relation.

4. Depending on the type of cointegration test and the sample period, you may or may not find the cointegration relationship for the same sample.

Since these two popular residual based cointegration tests produce ambiguous findings, we apply our modified RESET test to check whether the relationship is linear. The modified test is used for both absolute PPP and relative PPP with varying bandwidths M and numbers of polynomial terms k. Table 4 summarizes the results from the modified test as well as the original RESET test before modification. While the original RESET test tends to suggest support for a linear relationship much of the time (except for absolute PPP using the CPI), the modified RESET test shows little support for a linear cointegration specification (except in the case of absolute PPP with PPI and $k = 2$ in Japan-US).

These findings corroborate some existing empirical work on real exchange rates reported in Froot & Rogoff (1995). First, according to that work, cointegration is rejected more often in the floating exchange rate system, and our data, which cover the post Bretton Woods period, also show little support for linear cointegration using the modified RESET test. Second, past work with existing tests has found evidence of cointegration with PPI data more often than with CPI data. Correspondingly, our modified RESET test results indicate that (absolute) PPP with PPI data is closer to a linear relationship than the PPP using CPI data. Third, using P_t and P_t^* separately instead of the ratio P_t/P_t^* tends to lead to linear cointegration more often, partly of course because use of P_t/P_t^* is equivalent to restricting one part of cointegration vector and thereby this specification loses some flexibility. On the other hand, restricting this part of cointegration structure may partly be compensated for by allowance for an unspecified nonlinear form $f(\cdot)$.

6 Conclusion

Using some recently developed asymptotic tools in Park & Phillips (1999, 2001), this paper presents a specification test that can be applied to nonstationary time series. Neither conventional cointegration tests nor regression tests of linearity can effectively discriminate linear cointegration from both nonlinear cointegration and lack of cointegration. In contrast, the modified RESET test is a specification test that can be used to assess the adequacy of a linear cointegrating relation against certain forms of nonlinear cointegration and the alternative of no cointegration. We note that the conventional RESET test suffers from severe size distortion when it is applied to unit root nonstationary data and is therefore unsuitable for empirical application. The modifications to the RESET test developed here eliminate the biases that cause these size distortions and lead to a corrected test statistic that has a limiting central χ^2 distribution that is well suited for empirical work and which has good power against both nonlinear cointegration and no cointegration alternatives. When this test is applied to study PPP relationships for the US, Japan and Canada, the test strongly rejects the linear cointegration hypothesis in most PPP relationships, but the test finds that a PPI based PPP relationship is much better approximated by a linear specification than when CPI data is used.

Some related work is in progress. Since the power of the test depends on the choice of basis functions, we are developing a set of linearity tests using different basis functions. This seems particularly appropriate when we want to allow for functions whose behavior is poorly approximated by polynomials, such as integrable functions which attenuate the influence of integrated regressors. At the same time, there is scope for developing a linearity test that is not based directly on an approximating family, so that the power and the size of the test do not depend on so many choices such as the basis functions, the number of basis functions and a bandwidth parameter.

A Appendix

A.1 Additional Assumptions

Assumption A1:
For $\tilde{\zeta}_t = (\eta_{t+1}, \varepsilon_t)'$ *and the sigma field* $F_t = \sigma(\{\zeta_s\}_{-\infty}^t)$,

1. $\sup_{t\geq 1} E(\|\tilde{\zeta}_t\|^r \,|\mathcal{F}_{t-1}) < \infty$ *a.s. for some* $r > 4$
2. $E(\tilde{\zeta}_{i,t}^2 \tilde{\zeta}_{j,t-l}) = 0$ *for all* i, j *and for all* $l \geq 1$
3. ε_t *is iid with* $E|\varepsilon|^r < \infty$ *for some* $r > 8$ *and its distribution is absolutely continuous with respect to Lebesgue measure and has characteristic function* φ *for which* $\varphi(\lambda) = o(\|\lambda\|^{-\delta})$ *as* $\lambda \to \infty$ *for some* $\delta > 0$.

These additional moment conditions for the innovation processes correspond to those in Chang, Park & Phillips (2001) and are used to apply some of their results given in Lemma 10 below. While the first two conditions are fairly general moment conditions, the last condition on the distribution of the innovation process of X_t is somewhat strong than usual. However, it is still satisfied by a wide class of processes, for example all invertible Gaussian ARMA models.

A.2 Lemmas

The following lemma concerns the asymptotic orders of kernel estimators of long-run (co)variance matrices when $\{X_t, Y_t\}$ are not cointegrated and is proved in Lemma 1 of Xiao & Phillips (2002).

Lemma 9 *Suppose* u_t *is I(1) and* X_t *satisfies Assumption A. As* $n \to \infty$, $M \to \infty$, *and* $M/n \to 0$,

$$\frac{1}{M}\hat{\Omega}_{uv} \Rightarrow 2\pi\tilde{w}(0)\left(\int dB_x Q\right) + \Omega_*,$$
$$\frac{1}{M}\hat{\Lambda}_{uv} \Rightarrow 2\pi\tilde{w}_1(0)\left(\int dB_x Q\right) + \Lambda_*,$$
$$\frac{1}{nM}\hat{\Omega}_{uu} \Rightarrow 2\pi\tilde{w}(0)\int Q^2,$$

where M *is bandwidth parameter,* $Q = B_y - \beta^* B_x$ *for* $(\sum X_t^2)^{-1}\sum X_t Y_t \Rightarrow \beta^*$,

$$\tilde{w}(0) = \frac{1}{2\pi}\int_{-1}^{1} K(a)\,da \quad \text{and} \quad \tilde{w}_1(0) = \lim_{M\to\infty}\frac{1}{2\pi M}\sum_{h=0}^{M} K\left(\frac{h}{M}\right) = \frac{1}{2\pi}\int_0^1 K(a)\,da,$$

and

$$\Omega_* = \sum_{h=-\infty}^{\infty} E(v_t q_{t+h}) \quad \text{and} \quad \Lambda_* = \sum_{h=1}^{\infty} E(v_t q_{t+h}),$$

with $q_t = y_t - \hat{\beta}x_t$ *and* $n^{-1/2}\sum_1^{[nr]} q_t \Rightarrow Q(r)$.
□

Lemma 10 below is from Chang, Park & Phillips(2001). It summarizes the limits of the sample moments of nonlinearly transformed time series for I-regular and H-regular nonlinearities.

Lemma 10 *Let $a_i(\cdot)$ be I-regular, $b_i(\cdot)$ be H-regular with asymptotic order κ_i and limit homogeneous function $h_i(\cdot)$ which is piecewise differentiable with locally bounded derivative. Then, as $n \to \infty$,*

$$(1) \quad \frac{1}{n^{1/2}} \sum_{t=1}^{n} a_i(x_t) \Rightarrow^d L(1,0) \int_{-\infty}^{\infty} a_i(s)ds$$

$$(2) \quad \frac{1}{n \cdot \kappa_i} \sum_{t=1}^{n} b_i(x_t) \Rightarrow^d \int_0^1 h_i(B_x(r))dr$$

$$(3) \quad \frac{1}{n^{1/4}} \sum_{t=1}^{n} a_i(x_t)u_t \Rightarrow^d \left(L(1,0) \int_{-\infty}^{\infty} a_i(s)a_i(s)\prime ds \right)^{1/2} W(1)$$

$$(4) \quad \frac{1}{n^{1/2} \cdot \kappa_i} \sum_{t=1}^{n} b_i(x_t)u_t \Rightarrow^d \int_0^1 h_i(V(r))dU(r) \left(=_d \left(\int h_i(B_x)h_i(B_x)\prime \right)^{1/2} W(1) \right)$$

$$(5) \quad \frac{1}{n^{1/2}} \sum_{t=1}^{n} a_i(x_t)a_i(x_t)\prime \Rightarrow^d L(1,0) \int_{-\infty}^{\infty} a_i(s)a_i(s)\prime ds$$

$$(6) \quad \frac{1}{n^{1/2}} \sum_{t=1}^{n} a_i(x_t)a_j(x_t)\prime \Rightarrow^p 0 \qquad \text{for } i \neq j$$

$$(7) \quad \frac{1}{n^{1/2} \cdot \kappa_i} \sum_{t=1}^{n} a_i(x_t)b_j(x_t)\prime = O_p(1)$$

$$(8) \quad \frac{1}{n \cdot \kappa_i \cdot \kappa_j} \sum_{t=1}^{n} b_i(x_t)b_j(x_t)\prime \Rightarrow^d \int_0^1 h_i(B_x(r))h_j(B_x(r))\prime dr$$

where $n^{-1/2}x_t = n^{-1/2}\sum v_j \Rightarrow B_x(r)$, $n^{-1/2}\sum u_t \Rightarrow B_u(r)$, and $\{v_t, u_t\}$ satisfy Assumption A and Assumption A1. $W(r)$ is another Brownian motion independent of $B_x(r)$ and $B_u(r)$ with the same variance as $B_u(r)$. The equality in the parenthesis of (4) holds only if $B_x(r)$ is independent of $B_u(r)$.

□

A.3 Proofs

Proof. [Lemma 2] Applying the Beverage-Nelson (BN) decomposition to u_t as in Phillips and Solo (1992) we can write

$$\sum_{t=1}^{n} \left(\frac{X_t}{\sqrt{n}} \right)^m \frac{u_t}{\sqrt{n}} = \sum_{t=1}^{n} \left(\frac{X_t}{\sqrt{n}} \right)^m \frac{\varepsilon_t}{\sqrt{n}} C(1) - \sum_{t=1}^{n} \left(\frac{X_t}{\sqrt{n}} \right)^m \frac{\triangle \tilde{\varepsilon}_t}{\sqrt{n}}$$

with $\tilde{\varepsilon}_t = \sum_{j=0}^{\infty}\left(\sum_{s=j+1}^{\infty} c_s\right)\varepsilon_{t-j}$ and the difference operator $\triangle$. The first term converges to $\int B_x^m dB_\varepsilon C(1)$ and the second term becomes

$$\begin{aligned}
\sum_{t=1}^{n}\left(\frac{X_t}{\sqrt{n}}\right)^m \frac{\triangle\tilde{\varepsilon}_t}{\sqrt{n}} &= \left(\frac{X_n}{\sqrt{n}}\right)^m \frac{\tilde{\varepsilon}_t}{\sqrt{n}} - \sum_{t=1}^{n}\triangle\left(\frac{X_t}{\sqrt{n}}\right)^m \frac{\tilde{\varepsilon}_{t-1}}{\sqrt{n}} \\
&= o_p(1) - \frac{m}{\sqrt{n}}\sum_{t=1}^{n}\left(\frac{X_t}{\sqrt{n}}\right)^{m-1}\frac{v_t\tilde{\varepsilon}_{t-1}}{\sqrt{n}} \\
&= -\frac{m}{n}\cdot E(w_t)\sum_{t=1}^{n}\left(\frac{X_t}{\sqrt{n}}\right)^{m-1} - \frac{m}{n}\sum_{t=1}^{n}\left(\frac{X_t}{\sqrt{n}}\right)^{m-1}[w_t - E(w_t)] \\
&\Rightarrow -m\cdot E(w_t)\int B_x^{m-1} + o_p(1)
\end{aligned}$$

where $E(w_t) = E(v_t\tilde{\varepsilon}_{t-1}) = \triangle_{vu}$, a one-sided long-run covariance[10], and where we use the binomial expansion of $X_t^{m-1} = (X_{t-1}+v_t)^{m-1}$,

$$X_t^{m-1} = X_{t-1}^{m-1} + \sum_{k=1}^{m-1}\binom{m-1}{k}X_{t-1}^{m-1-k}v_t^k,$$

and Lemma 10.

□

Proof. [Theorem 3] See the proof for Theorem 5. All steps are the same except the fact that the second order bias terms are not corrected but are collected together to form the noncentrality parameter.

□

Proof. [Lemma 4] The serial correlation correction term (14) follows from the sample moment asymptotics for H-regular nonlinear functions in Theorem 3.3 of Park & Phillips (2001). For the endogeneity correction term (15), note that

$$\begin{aligned}
\sum_{t=1}^{n}\left(\frac{X_t}{\sqrt{n}}\right)^m \frac{v_t}{\sqrt{n}} &= \sum_{t=1}^{n}\left(\frac{X_{t-1}}{\sqrt{n}}\right)^m \frac{v_t}{\sqrt{n}} + \sum_{t=1}^{n}\left\{\sum_{k=1}^{m}\binom{m}{k}\left(\frac{X_{t-1}}{\sqrt{n}}\right)^{m-k}\left(\frac{v_t}{\sqrt{n}}\right)^{k+1}\right\} \\
&= P_1 + P_2 \\
&\Rightarrow^d \int B_x^m dB_x + m\triangle_{vv}\int B_x^{m-1}, \qquad (20)
\end{aligned}$$

[10] Note that for u_t and v_t satisfying Assumption A, we have

$$E(v_t\tilde{\varepsilon}_{t-1}) = E\left[\left.\sum_{j=0}^{\infty} d_j\eta_{t-j}\right)\left\{\sum_{k=0}^{\infty}\left.\sum_{s=k+1}^{\infty} c_s\right)\varepsilon_{t-1-k}\right\}\right] = \sigma_{12}\left\{\sum_{j=0}^{\infty} d_j \left.\sum_{s=j+1}^{\infty} c_s\right)\right\}$$

$$\sum_{h=0}^{\infty} E(v_0u_h) = \sum_{h=0}^{\infty} E\left[\left.\sum_{j=0}^{\infty} d_j\eta_{-j}\right)\left.\sum_{i=0}^{\infty} c_i\varepsilon_{h-i}\right)\right] = \sigma_{12}\sum_{h=0}^{\infty}\sum_{j=0}^{\infty} d_j c_{h+j+1} = \sigma_{12}\sum_{j=0}^{\infty} d_j\sum_{s=j+1}^{\infty} c_s$$

and, therefore, $E(v_t\tilde{\varepsilon}_{t-1}) = \triangle_{vu}$.

with

$$\begin{aligned}
P_1 &= \sum_{t=1}^{n}\left(\frac{X_{t-1}}{\sqrt{n}}\right)^m \frac{v_t}{\sqrt{n}} \\
&= \sum_{t=1}^{n}\left(\frac{X_{t-1}}{\sqrt{n}}\right)^m \frac{\eta_t}{\sqrt{n}}D(1) - \sum_{t=1}^{n}\left(\frac{X_{t-1}}{\sqrt{n}}\right)^m \frac{\triangle\tilde{\eta}_t}{\sqrt{n}} \\
&\approx \sum_{t=1}^{n}\left(\frac{X_{t-1}}{\sqrt{n}}\right)^m \frac{\eta_t}{\sqrt{n}}D(1) - \left\{\left(\frac{X_n}{\sqrt{n}}\right)^m \frac{\tilde{\eta}_n}{\sqrt{n}} - \frac{m}{n}\sum_{t=1}^{n}\left(\frac{X_t}{\sqrt{n}}\right)^{m-1} v_t\tilde{\eta}_t\right\} \\
&\Rightarrow^d \int B_x^m dB_x - \left\{o_p(1) - m\cdot E(v_t\tilde{\eta}_t)\int B_x^{m-1}\right\},
\end{aligned}$$

using the Phillips-Solo (1992) device, and

$$\begin{aligned}
P_2 &= \sum_{t=1}^{n}\left\{m\left(\frac{X_{t-1}}{\sqrt{n}}\right)^{m-1}\left(\frac{v_t}{\sqrt{n}}\right)^2 + \frac{m(m-1)}{2}\left(\frac{X_{t-1}}{\sqrt{n}}\right)^{m-2}\left(\frac{v_t}{\sqrt{n}}\right)^3 + \cdots + \left(\frac{v_t}{\sqrt{n}}\right)^{m+1}\right\} \\
&\Rightarrow^d m\cdot E(v_t^2)\int B_x^{m-1} + o_p(1),
\end{aligned}$$

where $E(v_t\tilde{\eta}_t) = \sum_{h=1}^{\infty} E(v_0 v_h) = \Lambda_{vv}$. Use of the consistent estimator $\hat{\triangle}_{vv}$ and (20) now leads to the required result

$$\sum_{t=1}^{n}\left(\frac{X_t}{\sqrt{n}}\right)^m \frac{v_t}{\sqrt{n}} - \hat{\triangle}_{vv}\frac{m}{n}\sum_{t=1}^{n}\left(\frac{X_t}{\sqrt{n}}\right)^{m-1} \Rightarrow^d \int B_x^m dB_x.$$

□

Proof. [Theorem 5]

The test statistic is a quadratic form in $\hat{u}'F\cdot D_n$ and the bias terms, with the weight matrix $(\hat{\Omega}_{uu.v}D_n'\tilde{F}'\tilde{F}\cdot D_n)^{-1}$ as metric in the form. We show that this quadratic form statistic converges to a random variable that follows a $\chi^2(k)$ distribution in two steps. First, conditional on $\mathcal{F}_x = \sigma(B_x(r), 0 \le r \le 1)$, we show that $\hat{u}'F\cdot D_n$ becomes a zero mean Gaussian vector after bias corrections in the limit; and second, that its variance matrix is the limit of the weight matrix.

Start by writing

$$\begin{aligned}
\sum_{t=1}^{n} F_m(X_t)\hat{u}_t &= \sum_{t=1}^{n}\widetilde{F_m(X_t)}u_t \\
&= \sum_{t=1}^{n} F_m(X_t)u_t - \sum_{t=1}^{n} X_t u_t\left(\sum_{t=1}^{n} X_t^2\right)^{-1}\sum_{t=1}^{n} F_m(X_t)X_t,
\end{aligned}$$

where $\widetilde{F_m(X_t)}$ is the regression residual when $F_m(X_t)$ is regressed on X_t, and the first term is examined in Lemma 2 and Lemma 4.

In the RESET test, $F_{m-1}(X_t) = X_t^m$ and using Lemma 2

$$\begin{aligned}\frac{1}{n^{(m+1)/2}}\sum_{t=1}^{n}\widetilde{X_t^m}u_t &= \frac{1}{n^{(m+1)/2}}\sum_{t=1}^{n}X_t^m u_t - \frac{1}{n^{(m+1)/2}}\sum_{t=1}^{n}X_t u_t\left(\sum_{t=1}^{n}X_t^2\right)^{-1}\sum_{t=1}^{n}X_t^{m+1}\\ &\Rightarrow \int B_x^m dB_\varepsilon C(1) + m\triangle_{vu}\int B_x^{m-1} - \left(\int B_x dB_u + \Lambda_{vu}\right)\left(\int B_x^2\right)^{-1}\int B_x^{m+1}\\ &= \int \widetilde{B_x^m}dB_u + m\triangle_{vu}\int B_x^{m-1} - \Lambda_{vu}\left(\int B_x^2\right)^{-1}\int B_x^{m+1},\end{aligned}$$

with $\widetilde{B_x^m} = B_x^m - B_x\left(\int B_x^2\right)^{-1}\int B_x^{m+1}$. The last two terms carry the effects of the serial correlation bias. With consistent estimation of $\triangle_{vu}$ and Λ_{vu}, Lemma 4 gives the correction for the first bias term and the second bias term is a direct application of Lemma 10. So the serial correlation bias correction term for the $(m-1)^{th}$ element is given by

$$S_n(m-1) = \hat{\triangle}_{vu}\frac{m}{n}\sum_{t=1}^{n}\left(\frac{X_t}{\sqrt{n}}\right)^{m-1} - \hat{\Lambda}_{vu}\left(\frac{1}{n^2}\sum_{t=1}^{n}X_t^2\right)^{-1}\frac{1}{n}\sum_{t=1}^{n}\left(\frac{X_t}{\sqrt{n}}\right)^{m+1},$$

and the $(m-1)^{th}$ element of the sample covariance $\hat{u}'F\cdot D_n$ with the serial bias correction terms satisfy

$$\frac{1}{n^{(m+1)/2}}\sum_{t=1}^{n}\widetilde{X_t^m}u_t - S_n(m-1) \Rightarrow^d \int \widetilde{B_x^m}dB_u.$$

Note that this limit can be decomposed into two parts – a zero mean Gaussian mixture with variance matrix $\Omega_{uu.v}\int \widetilde{B_x^m}\widetilde{B_x^m}'$ and the endogeneity bias term –as follows.

$$\int \widetilde{B_x^m}dB_u = \int \widetilde{B_x^m}dB_{u.x} + \Omega_{uv}\Omega_{vv}^{-1}\int \widetilde{B_x^m}dB_x.$$

The endogeneity bias term has the following two components

$$\Omega_{uv}\Omega_{vv}^{-1}\int \widetilde{B_x^m}dB_x = \Omega_{uv}\Omega_{vv}^{-1}\left[\int B_x^m dB_x - \int B_x dB_x\left(\int B_x^2\right)^{-1}\int B_x^{m+1}\right].$$

The second part can be quite easily corrected by

$$\left(\frac{1}{n}\sum_{t=1}^{n}X_t v_t - \hat{\triangle}_{vv}\right)\left(\frac{1}{n^2}\sum_{t=1}^{n}X_t^2\right)^{-1}\left(\frac{1}{n^{(m+3)/2}}\sum_{t=1}^{n}X_t^{m+1}\right) \Rightarrow^d \int B_x dB_x\left(\int B_x^2\right)^{-1}\int B_x^{m+1},$$

and (15) in Lemma 4 gives the correction term for the first part. Therefore, with the endogeneity bias correction term defined as follows

$$\begin{aligned}E_n(m-1) = \hat{\Omega}_{uv}\hat{\Omega}_{vv}^{-1}\Bigg[&\left\{\sum_{t=1}^{n}\left(\frac{X_t}{\sqrt{n}}\right)^m\frac{v_t}{\sqrt{n}} - m\hat{\triangle}_{vv}\left(\frac{1}{n}\sum_{t=1}^{n}\left(\frac{X_t}{\sqrt{n}}\right)^{m-1}\right)\right\}\\ &-\left(\frac{1}{n}\sum_{t=1}^{n}X_t v_t - \hat{\triangle}_{vv}\right)\left(\frac{1}{n^2}\sum_{t=1}^{n}X_t^2\right)^{-1}\left(\frac{1}{n^{(m+3)/2}}\sum_{t=1}^{n}X_t^{m+1}\right)\Bigg],\end{aligned}$$

the $(m-1)^{th}$ element of the sample covariance $\hat{u}'F \cdot D_n$ with two correction terms, becomes

$$\frac{1}{n^{(m+1)/2}} \sum_{t=1}^{n} \widetilde{X_t^m} u_t - E_n(m-1) - S_n(m-1) \Rightarrow^d \int \widetilde{B_x^m} dB_{u.x}, \tag{21}$$

which follows a $\mathcal{N}\left(0, \Omega_{uu.x} \int \widetilde{B_x^m} \widetilde{B_x^m}'\right)$ distribution conditional on $\mathcal{F}_x = \sigma(B_x(r), 0 \leq r \leq 1)$.

Next consider the weight matrix. Using Lemma 10, the (i,j) element of $\left(D_n' \tilde{F}' \tilde{F} \cdot D_n\right)$ has the following limit

$$\frac{1}{n} \frac{1}{n^{(i+j+2)/2}} \sum_{t=1}^{n} \widetilde{X_t^{i+1}} \widetilde{X_t^{j+1}} \Rightarrow^d \int \widetilde{B_x^{i+1}} \widetilde{B_x^{j+1}}'. \tag{22}$$

From (21) and (22), it follows that the modified RESET statistic is a quadratic form with a limiting χ^2 distribution as

$$\begin{aligned} & \left\{\hat{u}'F \cdot D_n - E_n' - S_n'\right\} \left(\hat{\Omega}_{uu.v} D_n' \tilde{F}' \tilde{F} \cdot D_n\right)^{-1} \left\{D_n \cdot F'\hat{u} - E_n - S_n\right\} \\ \Rightarrow^d & \begin{pmatrix} \int \widetilde{B_x^2} dB_{u.x} \\ \vdots \\ \int \widetilde{B_x^{(k+1)}} dB_{u.x} \end{pmatrix}' \left[\Omega_{uu.v} \begin{pmatrix} \int \widetilde{B_x^2} \widetilde{B_x^2}' & \cdots & \int \widetilde{B_x^2} \widetilde{B_x^{k+1}}' \\ \vdots & \ddots & \vdots \\ \int \widetilde{B_x^{k+1}} \widetilde{B_x^2}' & \cdots & \int \widetilde{B_x^{k+1}} \widetilde{B_x^{k+1}}' \end{pmatrix}\right]^{-1} \begin{pmatrix} \int \widetilde{B_x^2} dB_{u.x} \\ \vdots \\ \int \widetilde{B_x^{(k+1)}} dB_{u.x} \end{pmatrix} \\ \overset{A}{\sim} & \ \chi^2(k), \end{aligned}$$

using Muirhead (1982).[11] Note that the test statistic follows a central $\chi^2(k)$ unconditionally.

□

Proof. **[Corollary 6]** Remember that G is the $k \times \tilde{k}$ matrix with $\tilde{k}$ eigenvectors of $F'F$ in its columns, after divided by corresponding eigenvalues. Using G, the principal components of the normalized regressors $F \cdot D_n$ can be written as $F_n^* = F \cdot D_n \cdot G$ (Theil, 1971). Note that (21) implies that conditional on $\mathcal{F}_x = \sigma(B_x(r), 0 \leq r \leq 1)$, the sample covariance of nonlinearly transformed X_t with two correction terms follows mean zero Gaussian distribution. Multiplying this by G gives

$$\left\{\hat{u}'F \cdot D_n \cdot G - E_n'G - S_n'G\right\} = \left\{\hat{u}'F_n^* - E_n'G - S_n'G\right\}$$

and this converges to

$$\mathcal{N}\left(0, \Omega_{uu.v}\left[\lim_{n\to\infty} G'D_n\tilde{F}'\tilde{F}D_nG\right]\right) = \mathcal{N}\left(0, \Omega_{uu.v}\left[\lim_{n\to\infty} \tilde{F}_n^{*\prime}\tilde{F}_n^*\right]\right)$$

conditional on $\mathcal{F}_x$. With the consistent estimator of $\Omega_{uu.v}$, $\hat{\Omega}_{uu.v}\tilde{F}_n^{*\prime}\tilde{F}_n^*$ converges to the variance, and the modified test statistic with principal components follows $\chi^2(\tilde{k})$ asymptotically.

[11] Denoting Z as the limit of weight matrix and Σ as the variance of the limit of $\hat{u}'F \cdot D_n - E_n - S_n$, the necessary and sufficient condition for the limit of the test statistic to be χ^2 is that either one of the following holds

$$(1)\ Z\Sigma Z\Sigma = Z\Sigma \quad \text{or} \quad (2)\ Z\Sigma Z = Z,$$

which is equivalent to the idempotency of $Z\Sigma$.

□

Proof. **[Theorem 7]** Note that the sample covariance of nonlinearly transformed X_t and the integrated process u_t diverges at the rate of n, as

$$\frac{1}{n}\left[\sum_{t=1}^{n}\left(\frac{X_t}{\sqrt{n}}\right)^m \frac{u_t}{\sqrt{n}}\right] \Rightarrow^d \int B_x^m B_u.$$

The two correction terms diverge at the rate of M because

$$\begin{aligned}\frac{1}{M}E_n(m-1) &= \frac{1}{M}\hat{\Omega}_{uv}\hat{\Omega}_{vv}^{-1}\left[\left\{\sum_{t=1}^{n}\left(\frac{X_t}{\sqrt{n}}\right)^m \frac{v_t}{\sqrt{n}} - \hat{\triangle}_{vv}\frac{m}{n}\sum_{t=1}^{n}\left(\frac{X_t}{\sqrt{n}}\right)^{m-1}\right\}\right. \\ &\quad \left. - \left(\frac{1}{n}\sum_{t=1}^{n}X_t v_t - \hat{\triangle}_{vv}\right)\left(\frac{1}{n^2}\sum_{t=1}^{n}X_t^2\right)^{-1}\left(\frac{1}{n}\sum_{t=1}^{n}\left(\frac{X_t}{\sqrt{n}}\right)^{m+1}\right)\right] \\ &\Rightarrow^d \left[\tilde{w}(0)\int dB_x Q + \Omega_*\right]\Omega_{vv}^{-1}\left[\int B_x^m dB_x - \int B_x\, dB_x\left(\int B_x^2\right)^{-1}\int B_x^{m+1}\right],\end{aligned}$$

and

$$\begin{aligned}\frac{1}{M}S_n(m-1) &= \frac{1}{M}\left[\hat{\triangle}_{vu}\frac{m}{n}\sum_{t=1}^{n}\left(\frac{X_t}{\sqrt{n}}\right)^{m-1} - \hat{\Lambda}_{vu}\left(\frac{1}{n^2}\sum_{t=1}^{n}X_t^2\right)^{-1}\frac{1}{n}\sum_{t=1}^{n}\left(\frac{X_t}{\sqrt{n}}\right)^{m+1}\right] \\ &\Rightarrow^d \left[\tilde{w}_1(0)\int dB_x Q + \triangle_*\right] m\int B_x^{m-1} - \left[\tilde{w}_2(0)\int dB_x Q + \Lambda_*\right]\left(\int B_x^2\right)^{-1}\int B_x^{m+1},\end{aligned}$$

using Lemma 9 in the Appendix. Therefore, the sample covariance augmented by the correction terms diverges at the rate of n, so that

$$\hat{u}'F \cdot D_n - E_n' - S_n' = O_p(n) - O_p(M) - O_p(M).$$

The variance matrix term in the statistic diverges at the rate of nM so that

$$\hat{\Omega}_{uu.v}D_n\tilde{F}'\tilde{F}D_n = O_p(nM)$$

using Lemma 9. Thus, the modified RESET test statistic quadratic form diverges at the rate of n/M.

□

Proof. **[Theorem 8]** Under the alternative specification, the modified test statistic changes only through $\hat{u}$, i.e. through $\hat{u}'F \cdot D_n$, $\hat{\triangle}_{vu}$, $\hat{\Omega}_{uv}$ and $\hat{\Omega}_{uu.v}$. We will examine the changes in the statistic by checking the orders of each of these terms.

First, suppose we estimate the following misspecified linear regression is fitted by least squares

$$Y_t = \theta X_t + e_t.$$

Using Lemma 10, we can show that the coefficient estimate $\hat{\theta}$ becomes either convergent to zero in the I-regular case (assuming $x^k f(x)$ is integrable) or of the order of $n^{-1/2}\kappa_n$ for the H-regular case. Using affixes I and H to designate these cases, we have

$$\hat{\theta}^{(I)} \approx \frac{1}{n}\left(\int B_x^2\right)^{-1}\left(o_p(1) + \int B_x dB_u + \Lambda_{vu}\right) \equiv \theta^{(I)},$$
$$\hat{\theta}^{(H)} \approx \frac{\kappa_n}{\sqrt{n}}\left(\int B_x^2\right)^{-1}\left(\int H(B_x)B_x + \frac{1}{\kappa_n\sqrt{n}}\int B_x\, dB_u + \frac{\Lambda_{vu}}{\kappa_n\sqrt{n}}\right) \equiv \theta^{(H)},$$

where $\kappa_n = \kappa(\sqrt{n})$ is the asymptotic order and $H(\cdot)$ is the asymptotic homogenous function of $f(\cdot)$. Note that when the true model is linear cointegration we have $H(B_x) = \theta B_x$ and then $\hat{\theta}^{(H)} - \theta = O_p(n^{-1})$, as usual.

We first consider the H-regular case. The $(m-1)^{th}$ element of the normalized nonlinear sample covariance $\hat{u}'F \cdot D_n$ becomes

$$\frac{1}{\kappa_n\sqrt{n}}\left\{\sum_{t=1}^{n}\left(\frac{X_t}{\sqrt{n}}\right)^m \frac{\hat{u}_t}{\sqrt{n}}\right\} = \frac{1}{\kappa_n\sqrt{n}}\left\{\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(\frac{X_t}{\sqrt{n}}\right)^m\left[f(X_t) + u_t - \hat{\theta}^{(H)}X_t\right]\right\}$$
$$\approx \int B_x^m\, h(B_x) + O_p(n^{-1/2}\kappa_n^{-1}) - \theta^{(H)}\int B_x^{m+1} \qquad (23)$$

so that the maximum order of $\hat{u}'F \cdot D_n$ is $\kappa_n\sqrt{n}$. Again, note that for the linear cointegration case, the first and third terms of (23) cancel and we are left with $n^{-1}\sum_{t=1}^{n} X_t u_t = O_p(1)$, as usual for linear cointegration.

The orders of the two bias correction terms S_n and E_n in Theorem 5 depend on the asymptotic order of the kernel estimators $\hat{\triangle}_{vu}$, $\hat{\Lambda}_{vu}$ and $\hat{\Omega}_{uv}$. Letting $K(j/M)$ be the lag kernel, we may decompose each of these estimates as follows

$$\hat{\Omega}_{uv} = \sum_{j=-M}^{M} K\left(\frac{j}{M}\right)\left\{\frac{1}{n}\sum_t \hat{u}_{t+j}v_t\right\}$$
$$= \sum_{j=-M}^{M} K\left(\frac{j}{M}\right)\left\{\frac{1}{n}\sum u_{t+j}v_t + \frac{1}{n}\sum f(X_{t+j})v_t - \hat{\theta}^{(H)}\frac{1}{n}\sum X_{t+j}v_t\right\}. \qquad (24)$$

The first term in braces in (24) is $O_p(1)$ and the other two are $O_p(n^{-1/2}\kappa_n)$, so that the overall maximum asymptotic orders of $\hat{\triangle}_{vu}$, $\hat{\Lambda}_{vu}$ and $\hat{\Omega}_{uv}$ are all $n^{-1/2}M\kappa_n$. Again, when linear cointegration holds the leading term dominates and is $O_p(1)$. Thus, combining (24) and (23) we find that

$$\hat{u}'F \cdot D_n - E_n - S_n = O_p(n^{1/2}\kappa_n) - O_p(n^{-1/2}M\kappa_n) - O_p(n^{-1/2}M\kappa_n)$$

has order $O_p(n^{1/2}\kappa_n)$ since $M/n \to 0$.

For the variance term $\hat{\Omega}_{uu.v}D_n\tilde{F}'\tilde{F}D_n$, the order now depends on the order of $\hat{\Omega}_{uu.v} = \hat{\Omega}_{uu} - \hat{\Omega}_{uv}\hat{\Omega}_{vv}^{-1}\hat{\Omega}_{vu}$, since the remaining factor is of order $O_p(1)$ under both the null and alternative hypotheses. The kernel estimator $\hat{\Omega}_{uu}$ can be shown to be of the maximum order of $M\kappa_n^2$ (but

$O_p(1)$ as usual under the null hypothesis). In particular,

$$
\begin{aligned}
\hat{\Omega}_{uu} &= \sum_{j=-M}^{M} K\left(\frac{j}{M}\right)\left[\frac{1}{n}\sum_t \hat{u}_t\hat{u}_{t+j}\right] \\
&= \sum_{j=-M}^{M} K\left(\frac{j}{M}\right)\left[\frac{1}{n}\sum_t f(X_t)f(X_{t+j}) + \frac{1}{n}\sum_t f(X_t)u_{t+j} - \hat{\theta}^{(H)}\frac{1}{n}\sum_t f(X_t)X_{t+j}\right. \\
&\qquad + \frac{1}{n}\sum_t f(X_{t+j})u_t + \frac{1}{n}\sum_t u_t u_{t+j} - \hat{\theta}^{(H)}\frac{1}{n}\sum_t X_{t+j}u_t \\
&\qquad \left. -\hat{\theta}^{(H)}\frac{1}{n}\sum_t f(X_{t+j})X_t - \hat{\theta}^{(H)}\frac{1}{n}\sum_t X_t u_{t+j} + \hat{\theta}^{(H)2}\frac{1}{n}\sum_t X_t X_{t+j}\right] \\
&= O_p(M\kappa_n^2),
\end{aligned}
$$

since, using Lemma 10, the maximum order of each term in the square bracket can be determined as follows.

1. The order of $n^{-1}\sum_t f(X_t)f(X_{t+j})$ is at most of order κ_n^2 by virtue of the Cauchy inequality

$$
\left(\frac{1}{n}\sum_t f(X_t)f(X_{t+j})\right)^2 \leq \left(\frac{1}{n}\sum_t f(X_t)^2\right)\left(\frac{1}{n}\sum_t f(X_{t+j})^2\right) = O_p(\kappa_n^2)\cdot O_p(\kappa_n^2).
$$

2. The orders of $n^{-1}\sum_t f(X_t)u_{t+j}$ and $n^{-1}\sum_t f(X_{t+j})u_t$ are $n^{-1/2}\kappa_n$.

3. The maximum orders of $\hat{\theta}^{(H)}\frac{1}{n}\sum_t f(X_t)X_{t+j}$ and $\hat{\theta}^{(H)}\frac{1}{n}\sum_t f(X_{t+j})X_t$ are κ_n^2 since

$$
\left(\frac{1}{n}\sum_t f(X_t)X_{t+j}\right)^2 \leq \left(\frac{1}{n}\sum_t f(X_t)^2\right)\left(\frac{1}{n}\sum_t X_{t+j}^2\right) = O_p(\kappa_n^2)\cdot O_p(n).
$$

   from Cauchy inequality and $\hat{\theta}^{(H)} = O_p(n^{-1/2}\kappa_n)$.

4. $n^{-1}\sum_t u_t u_{t+j} = O_p(1)$.

5. $\hat{\theta}^{(H)}\frac{1}{n}\sum_t X_{t+j}u_t$ and $\hat{\theta}^{(H)}\frac{1}{n}\sum_t X_t u_{t+j}$ are of the order of $n^{-1/2}\kappa_n$ since

$$
\begin{aligned}
&\hat{\theta}^{(H)}\left[\frac{1}{n}\sum_t X_{t+j}u_t + \frac{1}{n}\sum_t X_t u_{t+j}\right] \\
&= \hat{\theta}^{(H)}\left[\frac{1}{n}\sum_t X_t u_t + \frac{1}{n}\sum_t X_{t+j}u_{t+j} + \frac{1}{n}\sum_t (v_{t+1}+\cdots+v_{t+j})\,u_t\right. \\
&\qquad \left. -\frac{1}{n}\sum_t (v_{t+1}+\cdots+v_{t+j})\,u_{t+j}\right] \\
&\approx O_p(n^{-1/2}\kappa_n)\left[2\left(\int B_x\,dB_u + \Lambda_{vu}\right) + \sum_{h=1}^{j}\Gamma_{uv}(h) - \sum_{h=1}^{j}\Gamma_{uv}(-j)\right] = O_p(n^{-1/2}\kappa_n).
\end{aligned}
$$

6. $\hat{\theta}^{(H)^2}\frac{1}{n}\sum_t X_t X_{t+j} = O_p(n^{-1}\kappa_n^2) \cdot O_p(n) = O_p(\kappa_n^2)$.

Putting these together, the modified RESET test statistic is a quadratic form in a vector of $O_p(n^{1/2}\kappa_n)$ elements with matrix in the form whose elements are of order $O_p(M^{-1}\kappa_n^{-2})$, so that the overall order of the test statistic is at most $O_p(n/M)$.

For the I-regular case, we can show that the sample covariance does not diverge

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(\frac{X_t}{\sqrt{n}}\right)^m\left[f\left(X_t\right)+u_t-\hat{\theta}^{(I)}X_t\right]\approx O_p\left(1\right)+\int B_x^m dB_u-\theta^{(I)}\int B_x^{m+1}=O_p\left(1\right)$$

using Lemma 10. The kernel estimator of long-run (co)variance is

$$\begin{aligned}
\hat{\Omega}_{uv} &= \sum_{j=-M}^{M} K\left(\frac{j}{M}\right)\left[\frac{1}{n}\sum \hat{u}_{t+j}v_t\right] \\
&= \sum_{j=-M}^{M} K\left(\frac{j}{M}\right)\left[\frac{1}{n}\sum u_{t+j}v_t+\frac{1}{n}\sum f(X_{t+j})v_t-\hat{\theta}^{(I)}\frac{1}{n}\sum X_{t+j}v_t\right] \\
&\approx \Omega_{uv}+O_p(M/n^{3/4})-O_p(M/n)=O_p(\max\{1,M/n^{3/4}\})
\end{aligned}$$

so that the two correction terms E_n and S_n do not diverge either as long as $M/n^{3/4}\to 0$. Therefore,

$$\hat{u}'F\cdot D_n-E_n-S_n=O_p(\max\{1,M/n^{3/4}\})$$

Now consider the variance term. As in the H-regular case, we can show that

$$\hat{\Omega}_{uu}\approx\Omega_{uu}+O_p(M/n^{3/4})+O_p(M/n)=O_p(\max\{1,M/n^{3/4}\})$$

so that with $\hat{\Omega}_{uu.v}=\hat{\Omega}_{uu}-\hat{\Omega}_{uv}\hat{\Omega}_{vv}^{-1}\hat{\Omega}_{vu}$ the test statistic becomes $O_p(1)$ variable.

□

References

[1] Andrews, D.A.K. (1991), "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica* 59:3, pp. 817-858

[2] Baghestani, H. (1991), "Application of the RESET Test to the Original Andersen-Jordan Equation," *Journal of Macroeconomics* vol.13, No.1, pp. 157-169

[3] Barahona, M. and C-S. Pooh (1996), "Detection of Nonlinear Dynamics in Short, Noisy Time Series," *Nature* vol.381·16, pp. 215-217

[4] Basher, S. and A.A. Haug (2003), "Nonlinear Cointegration and Purchasing Power Parity: Results from Rank Test," *mimeo*

[5] Berben, R. and D. Dijk (1999), "Unit Root Tests and Asymmetric Adjustment: A Reassessment," *Econometric Institute Research Report*, ET-9902/A

[6] Blake, A.P. and G. Kapetanios (2000), "A Radial Basis Function Artificial Neural Network Test for Neglected Nonlinearity," *National Institute of Economics and Social Research*

[7] Breitung, J. (2001), "Rank Tests for Nonlinear Cointegration," *Journal of Business & Economic Statistics* vol.19, No.3, pp. 331-340

[8] Campbell, J.Y., A.W. Lo and A.C. MacKinlay (1997) *The Econometrics of Financial Markets.* Princeton University Press

[9] Chang, Y., J. Park and P.C.B. Phillip (2001), "Nonlinear Econometric Models with Cointegrated and Deterministically Trending Regressors," *Econometrics Journal* 4, pp. 1-36

[10] DeBenedictis, L.F. and D.E.A. Giles (1998), "Diagnostic Testing in Econometrics: Variable Addition, RESET and Fourier Approximations," in Ullah, A. and Giles, D.E.A. eds., *Handbook of Applied Economic Statistics*, Marcel Dekker, New York, pp. 383-417

[11] DeBenedictis, L.F. and D.E.A. Giles (1999), "Robust Specification Testing in Regression: the FRESET Test and Autocorrelated Errors," *Journal of Quantitative Economics* 15, pp. 67-75

[12] De Jong, R.M. (2002), "Nonlinear Estimators with Integrated Regressors but without Exogeneity." *mimeo*

[13] Dornbusch, R. (1987), "Purchasing Power Parity," in *The New Palgrave: A Dictionary of Economics.* Eds.: Eastwell, J. Milgate, M., and Newman, P. London: MacMillan; New York: Stockton Press, pp. 1075-85

[14] Dumas, B. (1992), "Dynamic Equilibrium and the Real Exchange Rate in a Spatially Separated World," *Review of Financial Studies* 5(2), pp. 153-180

[15] Ermini, L. and C.W.J. Granger (1993), "Some Generalizations on the Algebra of I(1) Processes," *Journal of Econometrics* 58, pp. 369-384

[16] Frankel, J.A. and A.K. Rose (1994), "A Survey of Empirical Research on Nominal Exchange Rates," NBER Working Paper No. 4865

[17] Franses, P.H. and D. van Dijk (2002), "Simple Test for PPP Among Traded Goods," *Econometric Institute Report EI 2002-2*

[18] Froot, K. and K. Rogoff (1995), "Perspectives on PPP and Long-run Real Exchange Rates," in Grossman, G. and Rogoff, K. eds., *Handbook of International Economics*, North Holland, pp. 1647-1688

[19] Granger, C.W.J. (1995), "Modelling Nonlinear Relationships Between Extended-Memory Variables," *Econometrica* 63:2, pp. 265-79

[20] Granger, C.W.J. and J. Hallman (1989), "Nonlinear Transformations of Integrated Time Series," *Journal of Time Series Analysis*, vol. 12, pp. 207-224

[21] Gilli, V. and G. Kaminsky (1991), "Nominal Exchange Rate Regimes and the Real Exchange Rate," *Journal of Monetary Economics* 27, pp. 191-212

[22] Hamilton, J.D. (2000), "A Parametric Approach to Flexible Nonlinear Inference," *Econometrica* 69, pp. 801-812.

[23] Ibragimov, R. and P. C. B. Phillips (2004). "Regression Asymptotics using Martingale Convergence Methods" Cowles Foundation Discussion Paper, #1473.

[24] Johnson, N.L. and S. Kotz (1970) *Distributions in Statistics: Continuous Univariate Distributions*, vol.2, Houghton Mifflin, New York

[25] Kapetanios, G. (2003), "Bootstrap Neural Network Cointegration Tests Against Nonlinear Alternative Hypotheses," *Studies in Nonlinear Dynamics & Econometrics* vol.7, No.2,.

[26] Keenan, D.M. (1985) "A Tukey Nonadditivity Type Test for Time Series Nonlinearity." *Biometrika* 72, pp. 39-44

[27] Kim, T-H., Y-S. Lee and P. Newbold (2003), "Spurious Nonlinear Regression in Econometrics," *mimeo*

[28] Lee, T-H, White, H., and C.W.J. Granger (1993), "Testing for Neglected Nonlinearity in Time Series Models: A Comparison of Neural Network Methods and Alternative Tests," *Journal of Econometrics*, 56, pp. 269-290

[29] Leung, S.F. and S. Yu (2001), "The Sensitivity of the RESET Tests to Disturbance Autocorrelation in Regression Analysis," *Empirical Economics* 26, pp. 721-726

[30] Lo, M.C. and E. Zivot (2001), "Threshold Cointegration and Nonlinear Adjustment to the Law of One Price," *Macroeconomic Dynamics* 5, pp. 533-576

[31] Lothian, J.R. and M.P. Taylor (1996), "Real Exchange Rate Behavior: The Recent Float from the Perspective of the Past Two Centuries," *Journal of Political Economy* 104:3, pp. 488-509

[32] Meese, R.A. and K. Rogoff (1983), "Empirical Exchange Rate Models of the Seventies," *Journal of International Economics* 14, pp. 3-24

[33] Michael, P., A.R. Nobay and D.A. Peel (1997), "Transactions Costs and Nonlinear Adjustment in Real Exchange Rates: An Empirical Investigation," *The Journal of Political Economy* vol. 105, No. 4, pp. 862-879

[34] Muirhead, R.J. (1982), *Aspects of Multivariate Statistical Theory*, John Wiley & Sons, Inc.

[35] Moshiri, S., N.E. Cameron and D. Scuse (1999), "Static, Dynamic, and Hybrid Neural Networks in Forecasting Inflation," *Computational Economics* 14, pp. 219-235

[36] Mussa, M. (1976), "The Exchange Rate, the Balance of Payments, and Monetary and Fiscal Policy Under a Regime of Controlled Floating," *Scandinavian Journal of Economics* 78, pp. 229-248

[37] Ng, S. and P. Perron (2002), "PPP May Not Hold Afterall: A Further Investigation," *Annals of Economics and Finance* 3,pp. 43-64

[38] O'Connell, P.G.J. (1998), "The Overvaluation of Purchasing Power Parity," *Journal of International Economics* 44:1, 1-19

[39] Park, J. (1990), "Testing for Unit Roots and Cointegration by Variable Addition," in Fomby, T.B. and Rodes, G.F. (eds.) *Advances in Econometrics*, vol.8, JAI Press, pp. 107-133

[40] Park, J. and P.C.B. Phillips (1999) "Asymptotics for Nonlinear Transformations of Integrated Time Series," *Econometric Theory*, vol. 15, pp. 269-298

[41] Park, J. and P.C.B. Phillips (2000), "Nonstationary Binary Choice." *Econometrica* 68, pp. 1249-1280

[42] Park, J. and P.C.B. Phillips (2001), "Nonlinear Regression with Integrated Time Series." *Econometrica* 69, pp. 117-161

[43] Pedroni, P. (2001), "Purchasing Power Parity Tests in Cointegrated Panels," *Review of Economics and Statistics* 83, pp.727-731

[44] Peters, S. (2000), "On the Use of the RESET Test in Microeconometric Models," *Applied Economics Letters* 7, pp. 361-365

[45] Phillips, P. C. B. (1983) "ERA's: A New Approach to Small Sample Theory," *Econometrica* 51:5, pp. 1505–1525.

[46] Phillips, P.C.B. (1995), "Fully Modified Least Squares and Vector Autoregression," *Econometrica*, 63:5, pp. 1023-1078

[47] Phillips, P.C.B. (1998), "Econometric Analysis of Fisher's Equation." *Cowles Foundation Discussion Paper*, No. 1180

[48] Phillips, P.C.B. (2001), "Descriptive Econometrics for Nonstationary Time Series with Empirical Illustrations." *Journal of Applied Econometrics*, 16, pp. 389-413.

[49] Phillips, P.C.B. and B.E. Hansen (1990), "Statistical Inference in Instrumental Variables Regression with $I(1)$ Processes," *Review of Economic Studies*, vol.57, pp. 99-125

[50] Phillips, P.C.B. and J. Park (1998) "Nonstationary Density Estimation and Kernel Autoregression." *Cowles Foundation Discussion Paper*, No. 1182

[51] Phillips, P.C.B. and V. Solo (1992), "Asymptotics for Linear Processes," *Annals of Statistics* 20, pp. 971-1001

[52] Porter, R. and A. Kashyap (1984), "Autocorrelation and the Sensitivity of RESET," *Economics Letters* 14, 229-33

[53] Ramsey, J.B. (1969), "Tests for Specification Errors in Classical Linear Least-squares Regression Analysis," *Journal of the Royal Statistical Society, Ser. B*, 31:2, pp. 350-371

[54] Ramsey, J.B. and P. Schmidt (1976), "Some Further Results on the Use of OLS and BLUS Residuals in Specification Error Tests," *Journal of the American Statistical Association*, vol.71, No.354, pp. 389-390

[55] Rao, C. R. and S. K. Mitra (1971). *Generalized Inverse of Matrices and its Applications.* Wiley.

[56] Rogoff, K. (1996), "The Purchasing Power Parity Puzzle," *Journal of Economic Literature*, pp. 647-668

[57] Saikkonen, P. and I. Choi (2004), "Cointegrating Smooth Transition Regressions," *Econometric Theory* 20, pp. 301-340

[58] Sercu, P. Uppal, R. and C. van Hulle (1995), "The Exchange Rate in the Presence of Transactions Costs: Implication for Tests of Purchasing Power Parity," *Journal of Finance* 50, pp. 1309-1319

[59] Swanson, N.R. (1999), "Finite Sample Properties of a Simple LM Test for Neglected Nonlinearity in Error-Correcting Regression Equations," *Statistica Neerlandica* vol.53, nr.1, pp. 76-95

[60] Taylor, A.M. (2002), "A Century of Purchasing-Power Parity," *The Review of Economics and Statistics* 84(1), pp. 139-150

[61] Teräsvirta, T. and A. Eliasson (2001), "Nonlinear Error Correction and the UK Demand for Broad Money, 1878-1993," *Journal of Applied Econometrics* 16, pp. 277-288

[62] Theil, H. (1971), *Principles of Econometrics*, Wiley John & Sons Inc., New York

[63] Tsay, R.S. (1986), "Nonlinearity Tests for Time Series." *Biometrika* 73, pp. 461-466

[64] Thursby, J. (1979), "Alternative Specification Error Tests: A Comparative Study," *Journal of the American Statistical Association* vol.74, No.365, pp. 222-225

[65] Thursby, J. (1981), "A Test Strategy for Discriminating Between Autocorrelation and Misspecification in Regression Analysis," *Review of Economics and Statistics* 63, pp. 117-123

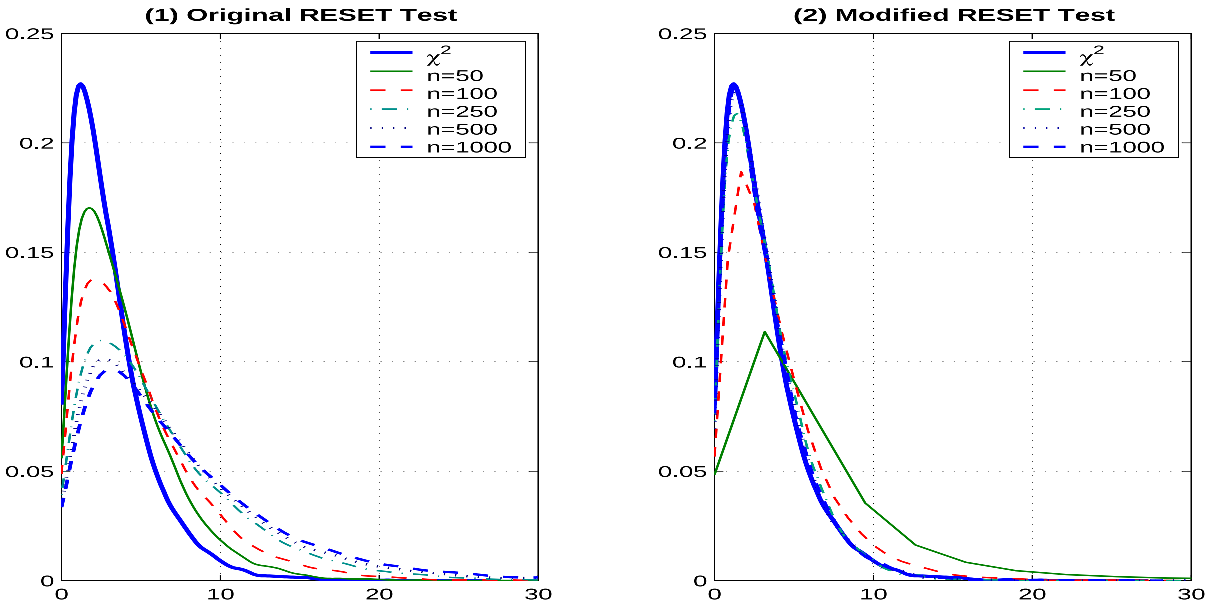
[66] Thursby, J. and P. Schmidt (1977), "Some Properties of Tests for Specification Error in a Linear Regression Model," *Journal of the American Statistical Association* vol.72, No.359, pp. 635-641

[67] Vitaliano, D.F. (1987), "On the Estimation of Hospital Cost Functions," *Journal of Health Economics* 6, pp. 305-18

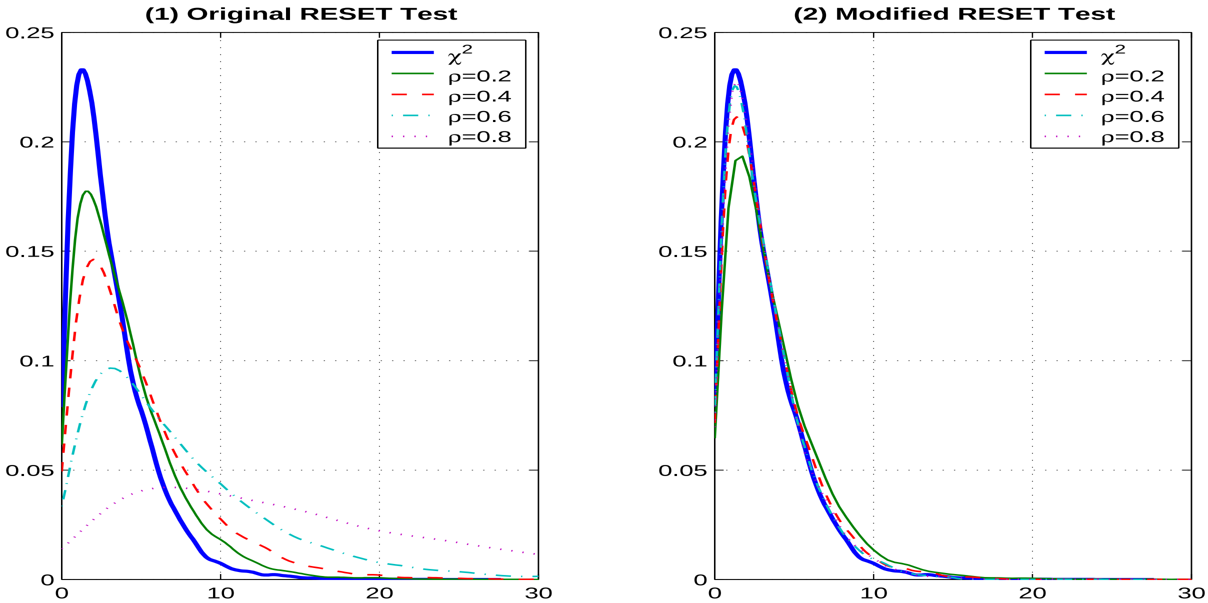
[68] White, H. (1992), *Estimation, Inference and Specification Analysis*, Cambridge University Press, New York

[69] Xiao, Z. and P.C.B. Phillips (2002), "A CUSUM Test for Cointegration Using Regression Residuals"), *Journal of Econometrics*, 108, pp. 43 - 61.

Figure 1: The RESET Test Statistics Before and After Modification under H_0



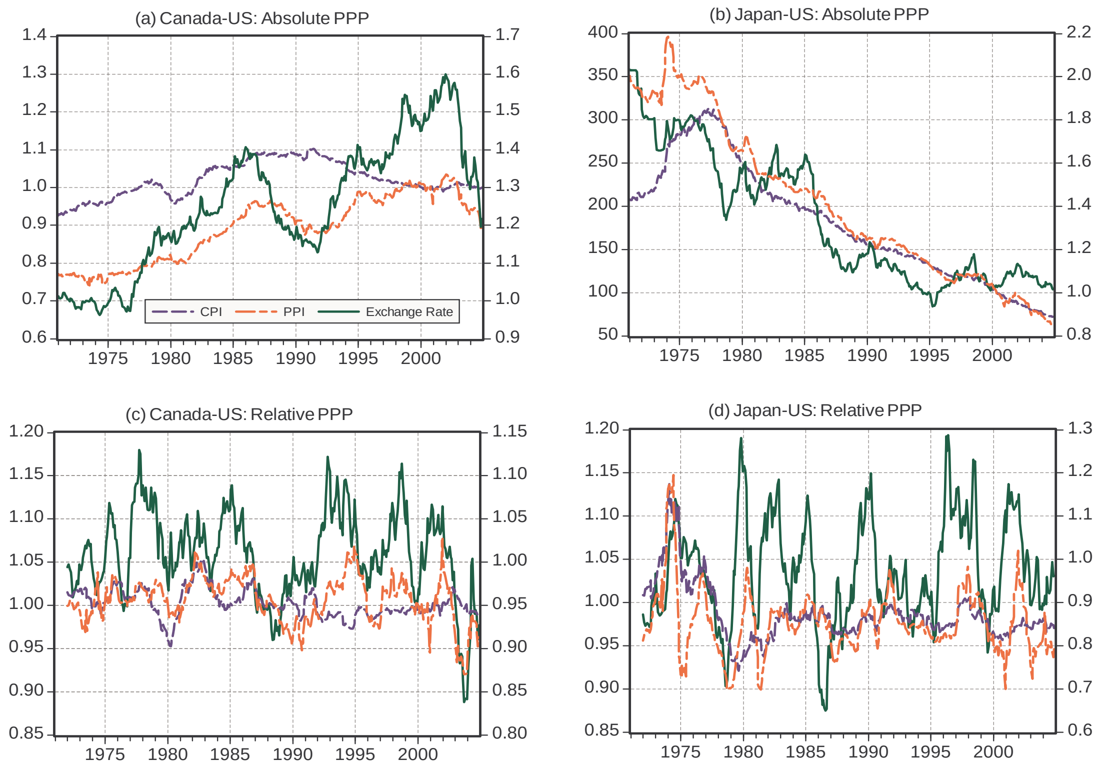
(a) Varying Number of Observations with $\rho = 0.6$ and $k = 3$



(b) Varying Serial Correlation with $n = 1000$ and $k = 3$

[Note] Empirical distributions of the test statistic shown above are from 10,000 simulated samples with $k = 3$. The bandwidth for the kernel estimator of long-run (co)variance is chosen automatically following Andrews (1991). χ^2-distribution in a thick solid line represents the limit distribution of test statistic from a central $\chi^2(k)$ distribution.

Figure 2: Nominal Exchange Rates and Price Ratios:US-Canada and US-Japan



[Note] The sample spans from 1971:1 to 2004:12. Upper figures plot dataset in levels and lower figures plot in the changes, S_t/S_{t-12} and $(P_t/P_{t-12})/(P_t^*/P_{t-12}^*)$ where the nominal exchange rate is plotted in the solid line, CPI ratio in the dashed line and PPI ratio in the dash-dotted line.

Table 1: Probability of Rejecting H_0 of Linear Cointegration

Function	Type	Modified RESET Test (1)	(2)	(3)	(4)	(5)	Original RESET Test (1)	(2)	(3)	(4)	(5)
ρ=0.2:											
	n=50	24.08	50.89	99.72	41.12	94.10	6.51	48.17	100.00	33.20	94.97
	n=100	15.30	71.77	99.83	39.02	96.19	7.87	76.85	100.00	42.45	96.72
	n=250	10.64	94.38	99.75	41.47	97.70	9.54	96.68	100.00	49.52	98.05
	n=500	8.39	99.16	99.76	42.64	98.33	10.19	99.81	100.00	51.87	98.64
	n=1000	8.27	99.98	99.69	45.35	98.63	10.04	99.99	100.00	54.05	98.89
ρ=0.4:											
	n=50	19.14	41.37	99.83	33.38	93.74	7.64	44.80	100.00	31.50	95.08
	n=100	11.27	59.64	99.68	28.19	96.03	11.38	75.84	100.00	41.36	96.75
	n=250	7.06	88.61	99.73	27.67	97.25	15.50	95.24	100.00	49.68	97.73
	n=500	6.39	98.36	99.77	28.85	98.09	17.09	99.59	100.00	51.70	98.44
	n=1000	6.10	99.91	99.68	31.13	98.60	17.96	99.99	100.00	54.66	98.94
ρ=0.6:											
	n=50	19.32	34.50	99.80	28.04	92.79	10.58	41.27	100.00	30.58	94.94
	n=100	9.26	39.57	99.76	18.45	95.72	18.37	71.01	100.00	41.60	96.81
	n=250	4.83	70.03	99.72	14.44	97.23	28.17	92.72	100.00	51.63	98.13
	n=500	4.96	90.88	99.82	14.46	98.06	32.87	98.72	100.00	55.59	98.71
	n=1000	4.91	98.69	99.85	15.83	98.79	35.70	99.93	100.00	58.22	99.16
ρ=0.8:											
	n=50	25.82	33.89	99.55	29.09	89.83	17.57	39.14	100.00	31.07	94.77
	n=100	10.69	23.99	99.68	13.18	94.86	34.15	65.00	100.00	46.04	97.05
	n=250	4.31	30.10	99.66	6.71	97.19	53.39	88.12	100.00	61.15	98.74
	n=500	4.07	53.04	99.76	6.19	97.83	63.72	95.72	100.00	69.85	99.06
	n=1000	4.99	80.01	99.67	7.22	98.46	70.18	99.02	100.00	74.64	99.57

[Note] (1)-(5) denote the functional forms defined in the beginning of simulation. The probabilities are calculated from 10,000 simulated samples with k=3 and the bandwidth is chosen automatically following Andrews (1991).

Table 2: Probability of Rejecting Linearity/Cointegration When X_t and Y_t are not cointegrated

	Number of Basis Functions (k)				
Bandwidth	1	2	3	4	5
$M = n^{1/5}$					
n=50	15.94	13.02	10.64	8.92	6.92
100	25.62	25.06	22.68	19.88	16.88
500	54.00	64.96	67.58	67.92	67.26
1000	64.40	78.88	82.46	84.12	84.54
$M = n^{1/4}$					
n=50	13.84	10.14	8.10	6.52	4.98
100	22.16	19.94	17.24	14.46	11.96
500	48.72	57.90	59.38	58.00	56.76
1000	59.40	73.08	76.26	76.72	76.98
$M = n^{1/3}$					
n=50	10.10	5.91	4.35	3.99	3.96
100	14.52	10.50	8.07	5.73	4.13
500	37.53	42.76	41.06	39.05	36.54
1000	49.16	58.31	59.39	58.85	57.05
$M = n^{1/2}$					
n=50	6.09	3.81	3.41	5.33	6.31
100	6.73	3.05	2.10	1.54	1.41
500	16.59	12.98	9.84	7.68	5.25
1000	23.68	21.70	18.14	15.13	11.93
Automatic					
n=50	26.02	36.87	42.42	54.43	57.14
100	23.08	30.36	33.89	42.31	45.01
500	17.77	21.40	23.60	27.23	28.63
1000	17.30	19.30	20.65	22.90	24.10

[Note] The rejection probabilities are calculated from 10,000 replications for the nominal 5% test. The automatic data-determined bandwidth choice in the bottom panel is based on Andrews (1991).

Table 3: Linear Cointegration Tests of Absolute PPP: ADF and KPSS

		ADF Test				KPSS Test
Number of Lags		1	2	3	4	
Period 1:						
		PPP with Consumer Price Index				
US-Canada	(1)	-0.9850	-0.8228	-0.9342	-1.0758	0.1295
	(2)	-0.9468	-0.7698	-0.8726	-1.0081	
US-Japan	(1)	-2.4526	-2.2983	-2.4672	-2.6947	0.3389***
	(2)	-2.4486	-2.2951	-2.4646	-2.6927	
		PPP with Producer Price Index				
US-Canada	(1)	-2.3966	-2.2435	-2.1414	-2.1075	0.1895
	(2)	-2.3910	-2.2365	-2.1335	-2.0994	
US-Japan	(1)	-3.0016	-2.9066	-3.0173	-3.2448	0.1853
	(2)	-3.0020	-2.9119	-3.0809	-3.2589	
Period 2:						
		PPP with Consumer Price Index				
US-Canada	(1)	-1.7996	-1.3970	-1.5988	-1.7630	0.0732
	(2)	-1.7400	-1.3175	-1.5032	-1.6614	
US-Japan	(1)	-2.5091	-2.1480	-1.9787	-1.8106	0.3227***
	(2)	-2.5051	-2.1459	-1.9786	-1.8106	
		PPP with Producer Price Index				
US-Canada	(1)	-2.3776	-2.2044	-2.1072	-1.9968	0.2354***
	(2)	-2.3778	-2.2051	-2.1098	-2.0010	
US-Japan	(1)	-2.3365	-2.2941	-2.5456	-2.5605	0.2610***
	(2)	-2.3271	-2.2834	-2.5344	-2.5496	

[Note] The cointegration regression is estimated for Period 1 (1971M1–2004M12) and Period 2 (1983M1–2004M12) with a constant and a linear trend. The number of lags in the column shows the number of lagged terms in the Dickey-Fuller regression for the regression residuals. The ADF test statistics with a constant term are reported in (1) and statistics with both a constant and a linear time trend are tabulated in (2). *'s show the null hypothesis rejected. One asterisk means rejection at a 10% significance level, two and three asterisks imply 5% and 1% respectively.

Table 4: P-values of the Modified and Unmodified RESET Tests

Choice of k		Modified RESET Test 2	3	4	Original RESET Test 2	3	4
		PPP with Consumer Price Index					
[A] Absolute PPP	Bandwidth						
US-Canada	$M = n^{1/3}$	0.000	0.000	0.000	0.000	0.000	0.000
	$M = n^{2/3}$	0.000	0.000	0.000	0.107	0.165	0.278
	Auto	0.000	0.000	0.000	0.000	0.000	0.001
US-Japan	$M = n^{1/3}$	0.000	0.000	0.000	0.000	0.000	0.000
	$M = n^{2/3}$	0.000	0.000	0.000	0.033	0.058	0.039
	Auto	0.150	0.000	0.000	0.003	0.005	0.002
[B] Relative PPP							
US-Canada	$M = n^{1/3}$	0.000	0.000	0.000	0.473	0.680	0.591
	$M = n^{2/3}$	0.000	0.000	0.000	0.766	0.910	0.910
	Auto	0.000	0.000	0.000	0.597	0.792	0.749
US-Japan	$M = n^{1/3}$	0.000	0.000	0.000	0.127	0.228	0.344
	$M = n^{2/3}$	0.000	0.000	0.000	0.307	0.479	0.632
	Auto	0.000	0.000	0.000	0.104	0.190	0.294
		PPP with Producer Price Index					
[A] Absolute PPP	Bandwidth						
US-Canada	$M = n^{1/3}$	0.000	0.000	0.000	0.000	0.000	0.000
	$M = n^{2/3}$	0.000	0.000	0.000	0.110	0.215	0.273
	Auto	0.000	0.000	0.000	0.078	0.162	0.205
US-Japan	$M = n^{1/3}$	0.002	0.005	0.000	0.001	0.002	0.004
	$M = n^{2/3}$	0.447	0.015	0.000	0.151	0.253	0.392
	Auto	0.000	0.000	0.000	0.048	0.088	0.160
[B] Relative PPP							
US-Canada	$M = n^{1/3}$	0.000	0.000	0.000	0.190	0.198	0.311
	$M = n^{2/3}$	0.000	0.000	0.000	0.300	0.337	0.483
	Auto	0.000	0.000	0.000	0.226	0.244	0.370
US-Japan	$M = n^{1/3}$	0.005	0.000	0.000	0.717	0.866	1.000
	$M = n^{2/3}$	0.000	0.000	0.000	0.832	0.939	1.000
	Auto	0.000	0.000	0.000	0.756	0.893	1.000

[Note] The modified RESET test results with bandwidths $M = n^{1/3}$ and $M = n^{2/3}$ and automatic bandwidth are reported. The p-values from the original RESET test without bias corrections are reported in the right panel for comparison.