



# Searching for evidence of long-run PPP from a post-Bretton Woods panel: separating the wheat from the chaff

Chi-Young Choi\*

*Department of Economics, University of New Hampshire, Durham, NH 03824, USA*

---

## Abstract

This paper reexamines the issue of long-run PPP using multiple panel tests in the framework of confirmatory analysis. Application of six panel tests under competing null hypotheses to the real exchange rates of 21 industrial countries yields seemingly contradictory evidence on the parity during the post-Bretton Woods period. Regardless of numeraire currency, four I(1) panel tests unanimously reject the null hypothesis in favor of long-run PPP, whereas two I(0) panel tests lend little support to the parity at conventional significance levels. Confirmatory analysis suggests that this puzzling result can be explained either by nonlinear dynamics of the real exchange rates or by a mixture of I(0) and I(1) series in the panel. Monte Carlo experiments indicate that potential mix of I(0) and I(1) series is more relevant to the empirical finding. The use of a sequential classification method sorts out six real exchange rates which exhibit most persistent deviations from long-run equilibrium. Systematic behavior of these series can be characterized better by country specific factors than by observable macroeconomic variables.

© 2004 Elsevier Ltd. All rights reserved.

*JEL classification:* C32; F31; F47

*Keywords:* PPP; Panel unit-root tests; Confirmatory analysis; Mixed panel; Sequential classification method

---

---

\* Tel.: +1-603-862-3372; fax: +1-603-862-3383.

*E-mail address:* [cy.choi@unh.edu](mailto:cy.choi@unh.edu) (C.-Y. Choi).

## 1. Introduction

This paper reexamines the issue of long-run Purchasing Power Parity (PPP) by applying panel data techniques to a popular data set. As a key building block for many models of exchange rate determination, PPP has been one of the most heavily studied subjects in international macroeconomics. Despite extensive research, however, the empirical evidence on PPP remains inconclusive, largely because of econometric challenges involved in determining its validity.<sup>1</sup> Indeed the empirical evidence on long-run PPP has experienced a rollercoaster of sorts with the evolution of relevant time series econometric tools. Earlier studies based on conventional univariate unit-root or cointegration tests found little evidence of PPP over the post-Bretton Woods period when nominal exchange rates were allowed to float, whereas studies using longer-horizon data sets or panel methods during the floating era tended to generate favorable evidence for parity (Table 1).<sup>2</sup>

During the past decade, tests for unit-roots in panel data have been widely employed in the study of long-run PPP as they are believed to circumvent the low power problem of univariate tests. However, the usefulness of panel methods has been questioned on a couple of grounds. First, as pointed out by O'Connell (1998a), the failure to account for cross sectional dependence across individual series results in serious size distortions for panel unit-root tests which are constructed under the restrictive assumption of cross sectional independence. Because the size distortion problem may outweigh the potential benefits of panel tests from increased power, some authors (e.g., Banerjee et al., 2001; Lyhagen, 2000) argue against the use of panel tests. Given that cross sectional dependence is present almost by construction in panels of real exchange rates through the base currency, the problem not only places serious doubt on the results of earlier panel tests, but poses a major limitation to the applicability of popular panel tests for the study of PPP.

The second problem associated with extant panel techniques is that it is not clear what we learn from a rejection of the null hypothesis. Many empirical studies using popular panel unit-root tests interpret rejections of the unit-root (hereafter I(1)) null hypothesis as support for long-run PPP, even though a rejection can be caused by the presence of as few as one stationary (hereafter I(0)) series.<sup>3</sup> Recently a group of dynamic panel tests has been developed under the null hypothesis of stationarity which is perhaps more natural to the study of PPP. Unfortunately, flipping the null hypothesis to stationarity does not necessarily facilitate interpretation

---

<sup>1</sup> It is generally agreed that most real exchange rates show very slow convergence which makes estimating long-run relationships difficult with existing statistical tools. Caner and Kilian (2001) claim that the roots of the AR(1) model range between 0.944 and 0.981 for quarterly real exchange rates. See Rogoff (1996) and Lothian (1998) for further discussion.

<sup>2</sup> MacDonald (1996), Wu (1996), and Papell (1997), among others, report supportive evidence of PPP based on the panel unit-root tests proposed by Levin et al. (2002) and Im et al. (2003).

<sup>3</sup> See Taylor and Sarno (1998) and Breuer et al. (2001). Throughout the paper, the I(1) (or I(0)) null hypothesis means the null hypothesis that all series in the panel are I(1) (or I(0)).

Table 1  
Summary of previous panel studies on PPP

	Data [numeraire]	CSD <sup>a</sup>	Methodology	Results
Abuaf and Jorion (1990)	Monthly (1973.1–1987.12), 10 industrial countries [USD]	YES	Restricted GLS	No support PPP
Bai and Ng (2001)	Quarterly (1974.Q1–1997.Q4), 21 industrial countries [USD]	YES (Factor Model)	PANIC	Mixed Support PPP
Henin et al. (2001)	Monthly (1973.1–1998.12), 17 OECD countries [USD]	YES (Bootstrap)	Iterated IPS test	Mixed Support PPP
Higgins and Zakrajsek (1999)	Quarterly (and Monthly), (1973.Q1–1994.Q3) 25 OECD countries [USD]	YES	LLC test and IPS test	Support PPP
Jorion and Sweeney (1996)	Monthly (1973.7–1993.12), G-10 countries [USD]	YES	SUR	Support PPP
Koedijk et al. (1998)	Annual (1949–1996), 24 industrial countries [USD,DM]	YES	NH test	Support PPP
Kuo and Mikkola (2001)	Annual (1949–1996), 24 industrial countries [USD,DM]	YES	NH test	Support PPP
MacDonald (1996)	Annual (1973–1992), 23 OECD countries [USD]	NO	LLC test	Support PPP
O'Connell (1998a)	Quarterly (1973.Q2–1995.Q4), 64 countries	YES	FGLS	No support PPP
Papell and Theodoridis (2001)	Quarterly (1973.Q1–1996.Q4), 21 industrial countries [USD]	NO YES	LLC test FGLS	Support PPP
Taylor and Sarno (1998)	Quarterly (1973.Q1–1996.Q2), G-5 countries [USD]	NO	Johansen's Likelihood ratio test	Support PPP
J. Wu and S. Wu (2001)	Quarterly (and monthly) (1973.Q2–1997.Q4), 20 industrial countries [USD,DM]	YES (non-parametric bootstrap)	IPS test MW test	Support PPP
J. Wu (1996)	Quarterly (and Annual, Monthly) (74.Q1–93.Q1), 19 industrial countries [USD]	YES (Common Time Effect)	LLC test	Support PPP

<sup>a</sup> CSD, Cross Sectional Dependence.

because a rejection of the null can still be triggered with just one I(1) series. As a strategy to overcome this problem, Choi (2001b) suggests the use of *confirmatory analysis* which involves combining panel tests under competing null hypotheses. According to Choi, such a strategy can lead to improved inference compared to the standard practice of employing only the I(1) or I(0) panel tests, especially when the outcomes corroborate each other. For example, we will be more confident on stationarity of the panel under study if a rejection by a panel test under the I(1) null is reinforced by a nonrejection of another panel test under the I(0) null. However, it is still uninformative in the case when both the I(1) and I(0) nulls are rejected.

The primary purpose of this paper is to reconsider the issue of long-run PPP by tackling these two issues. Although more recent panel studies (e.g., Bai and Ng, 2001; Chang, 2002) in the PPP literature address the dependence problem in panel tests, the inability of panel tests to deliver informative rejections of the null has received less attention. Toward this end, this paper adopts the strategy of Choi (2001b) and employs four commonly used panel tests under the I(1) null together with two panel tests under the I(0) null in the framework of confirmatory analysis. The use of multiple panel unit-root tests is readily justified by the fact that no single test dominates the others in heterogeneous panels. However, some panel tests considered here are formulated under the assumption of cross sectional independence and hence may yield biased results when applied to the panel of real exchange rates. For these tests, a bootstrap technique is utilized following Maddala and Wu (1999), Chang (2004), Mark and Sul (2001) and Wu and Wu (2001). By drawing inference from the bootstrapped distribution instead of the asymptotic one, we can alleviate the size distortion problem while retaining good power properties of panel tests.

Joint application of the six panel tests to a sample of 21 industrial countries over the current float period reveals an interesting result. Regardless of the choice of numeraire currency, the four I(1) panel tests consistently reject the null hypothesis at conventional significance levels, thereby mimicking the results found elsewhere. As for the two I(0) panel tests, however, little evidence of PPP is provided as they reject the I(0) null hypothesis for the same sample. The overall evidence on long-run PPP is therefore inconclusive. Confirmatory analysis suggests a couple of explanations for these seemingly contradictory findings. First, some or all of the real exchange rates may involve nonlinear dynamics. Since most panel tests employed here are built under the maintained hypothesis of linear dynamics, they would reject the nulls if the true underlying model is in fact nonlinear stationary. Second, the panel of real exchange rates may be comprised of I(0) as well as I(1) series. Given that panel tests are built under the null hypothesis that all series are either I(0) or I(1), the nulls are subject to reject if a panel is in fact mixed with I(0) series and I(1) series (hereafter, mixed panel).

A Monte Carlo study conducted here indicates that potential mix of I(0) and I(1) series can explain the contradictory result better, but ambiguity still remains as to which particular individual series are I(0) and I(1). To identify which series are I(0) or I(1), I propose a *sequential classification method*. By carrying out confirmatory analysis on various sub-samples of the mixed panel in sequential manner, I

can classify the panel of 20 real exchange rates into two groups: a group of fourteen  $I(0)$  series and another group of six  $I(1)$  series. This implies that long-run PPP holds in the majority of industrial countries although it does not so in all of them. Interestingly, the conclusion is robust to the choice of numeraire currency. I then explore underlying factors that might account for the systematic difference in dynamic patterns of real exchange rates between the two groups.

## 2. Econometric methodology

### 2.1. Confirmatory analysis

Empirical studies based on  $I(1)$  panel tests tend to provide favorable evidence to PPP mainly because they interpret rejections of the  $I(1)$  null hypothesis as convincing evidence of  $I(0)$  in real exchange rates. This interpretation, however, has been questioned on the grounds that a rejection of the null hypothesis can arise by the presence of as few as one  $I(0)$  series in the panel. The inability of panel tests to deliver informative rejection does not necessarily improve by considering tests of  $I(0)$  null hypothesis even though the null seems more natural to the study of PPP. A rejection of  $I(0)$  null is still uninformative as it is consistent both with a  $I(1)$  panel and with a mixture of  $I(0)$  and  $I(1)$  series. In other words, given the structure of hypotheses in panel tests, we are not able to make a correct inference if the panel under study is actually comprised of some  $I(0)$  series and some  $I(1)$  series. Sad to say, mixed panel structure is quite plausible in reality considering that substantial cross-country differences are easily observed in popular panel data sets.

As a strategy to tackle this problem, Choi (2001b) suggests performing *confirmatory analysis* which involves comparing the outcomes of two panel tests under competing null hypotheses. The basic idea of this strategy is that inferences made about  $I(0)$  or  $I(1)$  of the panel can be strengthened when the outcomes of two tests reinforce each other. For example, when a test under the  $I(0)$  null hypothesis rejects while another test under the  $I(1)$  null does not reject, we have confirmation on  $I(1)$  of the time series in the panel.<sup>4</sup> Via Monte Carlo experiments, Choi (2001b) shows that joint inference based on two panel tests under competing null hypotheses can improve the reliability of test inference over using either test alone.

The following chart lists four possible outcomes when a panel test is paired with another panel test under the competing null hypothesis. Two agreement outcomes

---

<sup>4</sup> This kind of joint inference is valid only when the two tests engaged are independent. Otherwise, inference needs to be made based on critical values from joint confirmation hypothesis as shown in Charemza and Syczewska (1998). Throughout the paper, however, inference is drawn from respective marginal critical values of two tests instead of joint confirmation hypothesis, not merely because it is technically challenging to form joint confirmation hypothesis for every possible combination of two panel tests but because the two approaches may produce little difference in inference. Using a combination of two popular univariate unit-root tests, Carrion-i-Silvestre et al. (2001) show that marginal critical values perform better than critical values based on joint confirmation hypothesis when the true process is stationary.

help confirm conclusions from respective single testing, whereas two disagreement outcomes produce contradictions. It is ideal to have agreement outcomes, but disagreement outcomes are often attainable in practice. Fortunately the probability of joint nonrejection (type-I disagreement) is quite low in panel approach mainly due to the power improvement of panel tests. By contrast, the probability of joint rejection (type-II disagreement) is still nonnegligible but it is interpretable under certain maintained assumptions. Under a linear dynamic model specification, for example, the joint rejection indicates a potential mix of linear  $I(0)$  and  $I(1)$  series in the panel.<sup>5</sup>

		I(1) Null	
		Fail to Reject	Reject
I(0) Null	Fail to Reject	DISAGREEMENT-I Contradiction: Inference is impossible. Data cannot distinguish between the competing models.	AGREEMENT-I Confirmation on Stationarity. All series in the panel are $I(0)$ .
	Reject	AGREEMENT-II Confirmation on unit-root. All series in the panel are $I(1)$ .	DISAGREEMENT-II Contradiction: Conclusion is unclear but interpretable.

*Note:* For the sake of the economy of exposition, I abuse the terminology a little bit throughout the paper by representing the outcomes of DISAGREEMENT-I as A-A, AGREEMENT-I as A-R, DISAGREEMENT-II as R-A, and DISAGREEMENT-II as R-R.

Since finite sample properties of confirmatory analysis vary over combination of panel tests, the present study employs six panel tests for confirmatory analysis: two tests under the  $I(0)$  null hypothesis are matched with four tests under the  $I(1)$  null. The following section briefly overviews these tests. The readers interested in further details are referred to the original papers for their work.

## 2.2. Panel tests under the $I(0)$ null

### 2.2.1. The panel G-test

The panel G-test (PG) is proposed by Choi (2001a) as a panel extension of the univariate G-test originally developed by Park and Choi (1988) and Park (1990). It is a *variable addition test* based on the regression of a given time series onto time

<sup>5</sup> Under more general model specifications, the joint rejection can be also caused by nonlinear dynamics or fractional integration.

polynomials including superfluous time polynomial terms. Specifically, a time series is regressed on a time polynomial with order dictated by the null hypothesis and then some superfluous higher-order time polynomial terms are added. By testing the significance of these superfluous time polynomial terms with the standard tests (such as the Wald test), we attempt to tell whether the series is I(0) (around a deterministic trend) or I(1). The superfluous regressors will be insignificant if the time series is I(0), whereas they will be significant if the series contains a unit-root component.

To test whether all real exchanges in a panel,  $\{q_{it}\}_{t=1,\dots,T}^{i=1,\dots,N}$ , are level stationary against the alternative that at least one of them is I(1), the panel G-statistic is articulated as

$$G^*(0, 2) = \frac{NT(\hat{\sigma}^2 - \tilde{\sigma}^2)}{\frac{1}{N} \sum_{i=1}^N \hat{\omega}_i^2} \xrightarrow{d} \chi_{2N}^2, \tag{1}$$

where

$$\hat{\sigma}^2 = \frac{1}{N} \frac{1}{T} \sum_{i=1}^N \sum_{t=1}^T \hat{e}_{it}^2, \quad \hat{e}_{it} = q_{it} - \hat{\alpha}_{0i},$$

$$\tilde{\sigma}^2 = \frac{1}{N} \frac{1}{T} \sum_{i=1}^N \sum_{t=1}^T \tilde{e}_{it}^2, \quad \tilde{e}_{it} = q_{it} - \tilde{\alpha}_{0i} - \tilde{\alpha}_{1i}t - \tilde{\alpha}_{2i}t^2,$$

and  $\hat{\omega}_i^2$  is the long run variance of  $\hat{e}_{it}$ ,  $\alpha_1$  and  $\alpha_2$  are the coefficients for the superfluous time polynomial terms.<sup>6</sup>

The null hypothesis will be rejected in favor of the alternative if the test statistic is larger than certain critical values. In this paper inferences are drawn from the p-values based on a nonparametric bootstrap method described in the Appendix A instead of the asymptotic distribution in order to control for cross sectional dependence.

### 2.2.2. The Nyblom and Harvey (NH) test

Nyblom and Harvey (2000, hereafter, NH) propose a multivariate generalization of a univariate test developed by Kwiatkowski et al. (1992), now widely referred to as KPSS. Consider the following model with an N-vector time series

$$y_t = \mu_t + \beta t + \varepsilon_t, \quad \varepsilon_t \sim N\left(0, \sum_{\varepsilon}\right), \tag{2}$$

$$\mu_t = \mu_{t-1} + \eta_t, \quad \eta_t \sim NID\left(0, \sum_{\eta}\right), \quad t = 1, \dots, T, \tag{3}$$

where  $y_t = (y_{1t}, y_{2t}, \dots, y_{Nt})'$ ,  $\mu_t$  is a vector random-walk with  $\mu_t = (\mu_{1t}, \mu_{2t}, \dots, \mu_{Nt})'$ ,

<sup>6</sup> For the long run variance estimation, Newey and West's (1987) Bartlett kernel is used with the fixed bandwidth of  $l = \text{integer}[8(T/100)^{1/4}]$  (for quarterly data).

$\beta$  is an  $N \times 1$  vector of zeros in the study of PPP. Under the null hypothesis of no random-walk component in the system ( $H_0: \sum \eta_i = 0$ ), against the alternative that at least one series is a random walk, the test statistic is formulated as

$$\xi_N = tr \left[ \left( \hat{\Gamma}_0 + \sum_{\tau=1}^m \frac{(m+1-\tau)}{m+1} [\hat{\Gamma}_\tau + \hat{\Gamma}'_\tau] \right)^{-1} \left( T^{-2} \sum_{j=1}^T \left[ \sum_{t=1}^j \hat{\epsilon}_t \right] \left[ \sum_{t=1}^j \hat{\epsilon}_t \right]' \right) \right], \tag{4}$$

where  $\hat{\Gamma}_\tau = T^{-1} \sum_{t=\tau+1}^T \hat{\epsilon}_t \hat{\epsilon}'_t$ ,  $\hat{\epsilon}_t = y_t - \bar{y}$  is the OLS residual under the null, and  $\sum_{t=1}^j \hat{\epsilon}_t$  is the partial sum of  $\hat{\epsilon}_t$ . The limiting distribution of the NH test statistic has a functional form of a standard vector Brownian bridge such that

$$\xi_N \xrightarrow{d} \int_0^1 B(r)' B(r) dr = \sum_{k=1}^{\infty} (\pi k)^{-2} u'_k u_k, \tag{5}$$

where  $u_k$  is an  $N \times 1$  vector of  $u_k \sim NID(0, I_N)$  and  $u'_k u_k \sim \chi^2(N)$ .

The NH test has been recently applied by Kuo and Mikkola (2001) to the real exchange rates of 24 OECD countries. Kuo and Mikkola drew inferences based on the simulated finite sample distribution to mitigate size distortion stemming from serial correlation. Their modified version of the NH test is employed in this paper.

### 2.3. Panel tests under the I(1) null

#### 2.3.1. The Levin, Lin and Chu (LLC) test

LLC (2002) propose several panel unit-root tests based on the following Augmented Dickey Fuller (ADF) model<sup>7</sup>

$$\Delta y_{it} = \alpha_i + \beta y_{i,t-1} + \sum_{j=1}^{k_i} \phi_{ij} \Delta y_{i,t-j} + \varepsilon_{it}, \quad i = 1, 2, \dots, N; \quad t = 1, 2, \dots, T. \tag{6}$$

Under the null hypothesis that all series in the panel are I(1) ( $H_0: \beta_1 = \dots = \beta_N = \beta = 0$ ) against the alternative that all series are I(0) ( $H_A: \beta_1 = \dots = \beta_N = \beta < 0$ ), the “adjusted t-statistic” ( $\tau^*$ ) obtained from pooled regression packages has a limiting distribution of standard normal,

$$\tau^* \xrightarrow{d} N(0, 1). \tag{7}$$

Note that a homogeneity restriction is imposed in the implicit alternative hypothesis such that all series, rather than at least one of them, are I(0). Despite this

<sup>7</sup> Throughout the paper, the lag length ( $k_i$ ) in the ADF regression equation is chosen by Hall’s (1994) general-to-specific method based on the recursive t-test.



restriction, rejection of the null hypothesis can occur when a small number of I(0) series are present in the panel.<sup>8</sup> For this reason, rejection of the null can be interpreted as at least one I(0) series exists in the panel.

### 2.3.2. The Im, Pesaran, and Shin (IPS) test

IPS (2003) develop a group mean panel unit-root test which allows for heterogeneities in intercept and serial correlation as well as convergence rate of real exchange rates across countries.<sup>9</sup> In the following regression equation,

$$\Delta y_{it} = \alpha_i + \rho_i y_{i,t-1} + \sum_{j=1}^{k_i} \phi_{i,j} \Delta y_{i,t-j} + \varepsilon_{it}, \quad i = 1, 2, \dots, N; \quad t = 1, 2, \dots, T, \quad (8)$$

IPS apply the univariate ADF test to each individual series and analyze the sample mean of the resulting t-statistics ( $\tau_i$ ),  $\bar{\tau}_N = (1/N) \sum_{i=1}^N \tau_i$ . Under the null hypothesis that all series in the panel are I(1) ( $H_0: \rho_i = 0$  for all  $i$ ) against the alternative that at least one of them is I(0) ( $H_A: \rho_i < 0$  for at least one  $i$ ), the IPS test statistic has a standard normal limiting distribution,

$$\frac{\sqrt{N}(\bar{\tau}_N - E(\bar{\tau}_N))}{\sqrt{\text{Var}(\bar{\tau}_N)}} \xrightarrow{d} N(0, 1). \quad (9)$$

### 2.3.3. The Maddala and Wu (MW) test

MW (1999) develop a panel unit-root test combining the p-values from individual ADF test across cross section units in the panel.<sup>10</sup> Under the null hypothesis that all series in the panel are I(1) against the alternative that at least one series is I(0), their test statistic is

$$\lambda = -2 \sum_{i=1}^N \log p_i \xrightarrow{d} \chi_{2N}^2, \quad (10)$$

where  $p_i$  denotes the p-value of the ADF statistic for the  $i_{th}$  unit in the panel.

### 2.3.4. The panel nonlinear IV unit-root test

Chang (2002) has developed another group mean panel test based on nonlinear Instrumental Variable (IV) estimation method. In the augmented autoregression of Eq. (8) with cross-sectional dependence across error terms,  $\rho_i$  is estimated using a nonlinear IV method to deal with the cross-sectional dependence, and then its corresponding t-ratio ( $Z_i$ ) is constructed. In testing the null hypothesis that all series are I(1) ( $H_0: \rho_i = 0$  for all  $i = 1, \dots, N$ ) against the alternative that at least one of

<sup>8</sup> See Bowman (1997), Breuer et al. (2001), and Mark (2001, in page 44).

<sup>9</sup> IPS propose another panel test (LM-bar test) based on the average of individual Lagrange Multiplier test, but the ADF based  $\bar{t}$ -test is considered here.

<sup>10</sup> A similar panel test has been developed by Choi (2001c).

them is  $I(0)$  ( $H_A$ :  $\rho_i < 0$  for at least one  $i$ ), the test statistic ( $S_N$ ) is built on averaging individual t-statistics such that

$$S_N = \frac{1}{\sqrt{N}} \sum_{i=1}^N Z_i, \quad (11)$$

which has a standard normal limiting distribution,  $S_N \xrightarrow{d} N(0, 1)$ .

#### 2.4. Cross sectional dependence and Bootstrap method

In his influential paper, O'Connell (1998a) shows that panel unit-root tests constructed under the assumption of cross sectional independence suffer from substantial size distortion if the assumption is violated. Through Monte Carlo experiments, he demonstrates that the significance level of the LLC test rises to as high as 50 percent for the nominal 5 percent in the presence of cross sectional dependence and consequently overrejects the true null.<sup>11</sup> Similar results have been reported by Maddala and Wu (1999) and Chang (2004). Given that economic shocks in general are common to more than one country and that real exchange rates are defined using common base currency, cross sectional dependence exists almost by construction in the study of PPP. It seems therefore unwarranted, if not erroneous, to draw quick conclusions on PPP without accounting for cross sectional dependence in panel study.

Except for the NH test and the IV test which are designed to accommodate cross sectional dependence, the rest of the panel tests considered here are built under the assumption of cross sectional independence. To deal with this problem, a *nonparametric residual-based bootstrap* method is employed here following Maddala and Wu (1999), Mark and Sul (2001) and Wu and Wu (2001).<sup>12</sup> By drawing inference from the bootstrapped distribution instead of the asymptotic one, I attempt to mitigate the size distortion problem stemming from cross sectional dependence. As claimed by Chang (2004), bootstrap method is also capable of circumventing some nuisance parameter problems in nonstationary panels with cross sectional dependence. A detailed description of the bootstrap procedure is presented in the Appendix A.

<sup>11</sup> O'Connell's finding also applies to other panel test procedures which are built on the same restrictive assumption. In the IPS test, for example, since the individual t-statistics would be correlated in the existence of cross sectional dependence, the application of Central Limit Theorem will be no longer valid.

<sup>12</sup> An anonymous referee of this journal suggests a useful alternative method to control for cross sectional dependence which involves adjusting test statistics directly using a modified GLS approach. Though appealing in relatively small  $N$ , the benefit of this strategy is easily outweighed by the computation cost associated with the parameter proliferation problem which entails constructing  $N \times N$  covariance matrix. Recently some other interesting strategies have been proposed by Bai and Ng (2001) and Phillips and Sul (2003).

### 3. Empirical results

#### 3.1. The data

The (log) real exchange rate for country  $i$  at time  $t$  is defined as

$$q_{it} \equiv s_{it} + p_t^* - p_{it}, \quad (12)$$

where  $p_{it}$  and  $p_t^*$  respectively denote the logarithms of the consumer price indices in country  $i$  and in the base country, and  $s_{it}$  represents the logarithm of nominal exchange rate against the base currency. Long-run PPP is said to hold if the  $\{q_{it}\}$  sequence is  $I(0)$ .

The data used here are quarterly nominal exchange rates and consumer price indices over 1973:1–1998:4, resulting in 104 observations respectively. They were retrieved from the International Monetary Fund's *International Financial Statistics (IFS)* for 21 industrial countries: Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Japan, the Netherlands, New Zealand, Norway, Portugal, Spain, Sweden, Switzerland, the United Kingdom, and the United States. The selection of countries is determined by a requirement of floating exchange rates during the post-Bretton Woods period and by the comparison purpose with earlier studies. Nominal exchange rates are end-of-quarter observations (IFS line code AE) and CPIs are quarterly averages (IFS line code 64).

#### 3.2. Test results

Table 2 presents the results of the six panel tests for the full sample of 20 real exchange rates based on two different numeraire currencies: columns 3–4 report the results using the US dollar (USD) as the base currency and the next two columns show the results using the deutsche mark (DM) as the numeraire.

Table 2  
Application of panel tests to 20 real exchange rates

Tests	Null Hypothesis	USD		DM	
		Statistic	(p-value)	Statistic	(p-value)
PG	I(0)	148.7935	(0.0134)	137.5613	(0.0760)
NH	I(0)	–	(0.0792)	–	(0.0944)
IPS	I(1)	–2.0795	(0.0266)	–2.0978	(0.0300)
LLC	I(1)	–8.7042	(0.0032)	–8.8705	(0.0064)
MW	I(1)	58.8374	(0.0804)	60.2609	(0.0690)
IV	I(1)	–4.687***	–	–3.241***	–

Note: Figures in the parentheses are nonparametric bootstrap p-values from 5,000 iterations. I(0) and I(1) represent that all series in the panel are I(0) and I(1), respectively. The statistics of the NH test is not reported because inference is drawn from p-values based on simulations. For the IV-test, only the test statistic is reported although p-values can be easily computed from its asymptotic distribution of standard normal.

\*\*\* Denotes the case where the null hypothesis cannot be rejected at one percent significance level.

Table 3  
Summary of Confirmatory Analysis

	I(0) Tests		I(1) Tests			
	PG	NH	IPS	LLC	MW	IV
$H_0$	All series are I(0)		All series are I(1)			
Results	Reject $H_0$		Reject $H_0$			
Inference	At least one series is I(1)		At least one series is I(0)			
Evidence of PPP	NO		YES			

When the USD serves as the base currency, the four I(1) panel tests unanimously reject the unit root null hypothesis at the 10% significance level, confirming the results of earlier studies in which the rejection was often interpreted as the supportive evidence of long-run PPP. On the other hand, the two I(0) panel tests reject the null for the same sample at the same significance level, which is in stark contrast with the result from the four I(1) tests. Interestingly, this finding is unaffected by numeraire choice. With the DM as the numeraire, the I(0) tests and the I(1) tests consistently reject their respective null hypotheses at conventional significance levels. According to O'Connell (1998a), so far as cross-sectional dependence is controlled for under GLS or FGLS estimation, the choice of numeraire currency becomes irrelevant to panel tests of PPP either if error terms are serially uncorrelated or if serial correlation has identical structures across individual series. Papell and Theodoridis (2001), however, show that the real exchange rates during the post Bretton Woods period exhibit considerably heterogeneous serial correlations and hence evidence of PPP gets stronger with the DM than with the USD as base currency.<sup>13</sup> The empirical result found in the current study stands somewhere between these two cases. Although the evidence of PPP gets marginally stronger with the DM in the I(0) tests, it is not substantial enough to lend a full support to the argument that PPP holds with the DM but not with the USD.

Table 3 reformulates the empirical results of Table 2 in the framework of confirmatory analysis. Irrespective of combination of joint tests, two paired tests consistently produce joint rejections, leading to inconclusive overall evidence on long-run PPP in the 21 industrial countries.

#### 4. Interpreting the empirical results

How to interpret this seemingly contradictory result? Two possible explanations are considered here: (1) the true underlying Data Generating Process (DGP) of the real exchange rates is nonlinear stationary process; (2) the panel of real exchange

<sup>13</sup> Papell and Theodoridis considered 21 OECD countries as numeraire and found that evidence of PPP is stronger for European than for non-European base currencies.

rates is comprised of both I(0) series and I(1) series.<sup>14</sup> Considering that most panel tests are built under the maintained hypothesis of linear dynamics and that test procedures are formulated under the null hypothesis that all series are either I(0) or I(1), the joint rejection can arise either when the true underlying model of real exchange rates is *not* linear (regardless of the stationarity) or when the panel under study consists of both I(0) and I(1) series.

#### 4.1. *Nonlinear dynamics of real exchange rates*

In the theoretical models of real exchange rate determination, goods market frictions such as transportation costs and trade barriers often implies a nonlinear adjustment process toward PPP (see Dumas, 1992; Sercu et al., 1995; Obstfeld and Taylor, 1997; O'Connell, 1998b).<sup>15</sup> For example, in the presence of transportation costs in international goods market arbitrage, small deviations of real exchange rates from long-run equilibrium will be uncorrected until they are large enough to cover the cost of tradings. Given that variations in the real exchange rate represent deviations from PPP, this suggests nonlinear behavior of real exchange rates—more like unit root processes in the vicinity of long-run equilibrium and more mean reverting the further they are away from equilibrium.<sup>16</sup> Recently a number of authors have applied various nonlinear dynamic adjustment models to characterize the behavior of real exchange rates. They generally find evidence of nonlinear mean reversion with faster convergence speeds in some selected bilateral real exchange rates (e.g., Michael et al., 1997; Baum et al., 2001; Parsley and Popper, 2001; Taylor, 2001; Taylor et al., 2001).

Most panel tests employed here are constructed under the maintained hypothesis of linear dynamics. If the true underlying process of real exchange rate is in fact nonlinear stationary, panel tests based on linear model may lead to rejection of the null hypotheses. In this sense, the seemingly contradictory empirical result obtained in the preceding section may be attributable to the nonlinear behavior of real exchange rates. Before investigating this issue further, it should be noted that the purpose of this section is not to evaluate the validity of nonlinear mean reverting models in characterizing real exchange rate behavior but to assess the relevance of nonlinear adjustment behavior of real exchange rate to my empirical result.

I designed the following Monte Carlo simulation experiments. Data are generated from the following three popular nonlinear stationary models. The first model, called DGP1, is due to Parsley and Popper (2001) in which nonlinearity is

---

<sup>14</sup> Engel and Kim (1999) find that the transitory component of the monthly U.S./U.K. real exchange rate shifts among three different states during 1885–1995. This possibility, however, is not considered here because no explicit exchange regime shift has been observed during the post Bretton Woods era.

<sup>15</sup> The author is grateful to an anonymous referee for bringing this feature to the author's attention.

<sup>16</sup> Linear characterization of real exchange rate movement treats the speed of adjustment constant irrespective of the degree of the deviation from long-run equilibrium. As pointed out by Abuaf and Jorion (1990), linearity restricts the dynamics of real exchange rates to only three possibilities: an explosive process, a random walk, or a monotonic adjustment to a constant value.

incorporated by multiplying the lagged deviation to its own absolute value. The other two DGPs utilize the exponential smooth transition autoregressive (ESTAR) model. DGP2 is a modified version of the best-fitting ESTAR model of the UK–US annual real exchange rate in Michael et al. (1997),<sup>17</sup> and DGP3 is extracted from the ESTAR formulation estimated by Taylor et al. (2001) using four monthly bilateral real exchange rates.

$$\begin{aligned} \text{DGP1 : } \Delta y_{i,t} &= \alpha_i + (1 - \rho_i)y_{i,t-1} - 0.1y_{i,t-1}|y_{i,t-1}| + \varepsilon_{i,t}, \quad \varepsilon_{it} = \lambda_i\varepsilon_{i,t-1} + u_{it}, \\ \text{DGP2 : } \Delta y_{i,t} &= 0.40\Delta y_{i,t-1} + (-y_{i,t-1} + 0.1\Delta y_{i,t-1} + 0.59\Delta y_{i,t-2} + 0.05\Delta y_{i,t-3} \\ &\quad + 0.57\Delta y_{i,t-4} - 0.017)\left\{1 - \exp\left[-\gamma_i(y_{i,t-1} - 0.038)^2\right]\right\} + \varepsilon_{i,t} \\ \text{DGP3 : } \Delta y_{i,t} &= -[y_{i,t-1} + \mu_i]\left\{1 - \exp\left[-\delta_i(y_{i,t-1} - \mu_i)^2\right]\right\} + \varepsilon_{i,t} \end{aligned}$$

where  $t = 1, \dots, T$ ,  $i = 1, \dots, N$ ,  $0.1 < \alpha_i < 0.5$ ,  $0.75 < \rho_i < 0.95$ ,  $0.2 < \lambda_i < 0.4$ , and  $-0.5 < \mu_i < 0$ .  $1 < \gamma_i < 500$  and  $0.2 < \delta_i < 0.5$  are the parameters for the speed of transition,  $u_{it}$  and  $\varepsilon_{it}$  are generated with  $N(0,1)$  after allowing for cross sectional dependence as discussed in the Appendix A. Panel sizes of  $(N, T) = (10, 50)$  and  $(20, 100)$  are considered here to reflect the actual data set. Two types of panel structures are designed. One is that all series in the panel are nonlinear stationary and the other is that a half series are nonlinear stationary and a half contain unit-roots. Then the six panel tests under consideration are applied to the generated pseudo data to record the average test statistics and the corresponding bootstrap-based p-values. Each simulation run is carried out with 5,000 replications. The rejection rates are then computed by the fraction of times when the p-value is less than 5% and 10% nominal levels.

Table 4 presents the simulation results. When the panel is comprised only of nonlinear stationary processes, the average p-values of the two I(0) tests are far in excess of 0.1 in all cases considered, indicating that they correctly do not reject the I(0) null. By contrast, the I(1) tests except for the IV test do not reject their false null hypothesis at conventional significance levels, suggesting that they are not able to distinguish nonlinear stationary process from I(1) process. Only the IV-test consistently provides correct inference by rejecting the null when all series in the panel are nonlinear stationary.

When the panel consists of a half nonlinear stationary series along with the other half of I(1) series, the picture changes slightly with the I(0) tests but not with the I(1) tests. The panel G-test correctly rejects the null for the mixed panel, whereas the NH test does not. Again the I(1) tests other than the IV-test continue to fail to reject the false null. Taken together, only the panel G-test and the IV test have discriminatory power between nonlinear stationary processes against nonstationary series. This argument is readily supported by the rejection rates reported in the second panel of Table 4. The rejection rates of the two tests reach more than 90 percent when the null hypothesis is not true, whilst that of the IPS test is very low.

<sup>17</sup> In Michael et al.,  $\gamma_i = 532.4$  and the coefficients for  $\Delta y_{i,t-1}$  and  $\Delta y_{i,t-3}$  are set to zeros.

Table 4  
Application of panel tests to nonlinear mean reverting processes

Panel Size	Tests	$H_0$	Nonlinear stationary panel			Mixed panel		
			DGP1	DGP2	DGP3	DGP1	DGP2	DGP3
<i>Test statistics and P-values</i>								
N = 10, T = 50	PG	I(0)	0.5836	0.7548	0.6088	0.0794	0.1608	0.0854
	NH	I(0)	0.5586	0.1648	0.6074	0.6774	0.6166	0.6720
	IPS	I(1)	0.4502	0.6222	0.5562	0.3668	0.5306	0.4480
	LLC	I(1)	0.3044	0.6016	0.5308	0.3320	0.4254	0.3176
N = 20, T = 100	MW	I(1)	0.5040	0.4962	0.5648	0.5194	0.4888	0.5246
	IV	I(1)	-3.5109***	-2.9845***	-3.1354***	-1.5898**	-1.4251**	-1.4755**
	PG	I(0)	0.4642	0.7848	0.4814	0.0250	0.0956	0.0266
	NH	I(0)	0.4350	0.1524	0.6078	0.5398	0.5018	0.5688
N = 10, T = 50	IPS	I(1)	0.1964	0.5032	0.2810	0.2266	0.4380	0.2760
	LLC	I(1)	0.1238	0.3848	0.2790	0.2298	0.2696	0.1726
	MW	I(1)	0.1572	0.2234	0.0854	0.2956	0.3488	0.2098
	IV	I(1)	-8.9730***	-9.0201***	-7.8036***	-4.2960***	-4.4714***	-3.8984***
<i>Rejection Rates</i>								
N = 10, T = 50	PG	5%	0.0230	0.0040	0.0300	0.7004	0.6548	0.6950
		10%	0.0560	0.0140	0.0558	0.7660	0.7010	0.7430
	IPS	5%	0.0056	0.0040	0.0034	0.0330	0.0080	0.0120
		10%	0.0140	0.0098	0.0312	0.0680	0.0310	0.0756
N = 20, T = 100	IV	5%	0.9764	0.8990	0.9370	0.4766	0.4038	0.4318
		10%	0.9900	0.9540	0.9720	0.6216	0.5514	0.5742
	PG	5%	0.0542	0.0022	0.0454	0.9200	0.8300	0.9210
		10%	0.1020	0.0086	0.0910	0.9364	0.8540	0.9344
N = 10, T = 50	IPS	5%	0.2824	0.0050	0.0812	0.2672	0.0310	0.0654
		10%	0.4286	0.0230	0.2336	0.3890	0.0850	0.2240
	IV	5%	1.0000	1.0000	1.0000	0.9980	0.9990	0.9850
		10%	1.0000	1.0000	1.0000	0.9998	1.0000	0.9936

Note: Except for the IV test, entries represent p-values from nonparametric bootstrap based on 5,000 replications. I(0) panel represents that all series in the panel are nonlinear stationary. Mixed panel depicts that half series in the panel are nonlinear stationary while the rest half series are unit root nonstationary. Rejection rate refers to the fraction of times when the p-values are less than the corresponding nominal significance levels.

\*\* Denote the case where the null hypothesis cannot be rejected at five percent significance level.

\*\*\* Denotes the case where the null hypothesis cannot be rejected at one percent significance level.

It now seems fair to claim that the contradictory empirical result in the preceding section is not consistent with the nonlinear dynamic models considered here.<sup>18</sup> Three I(1) panel tests (LLC, IPS, and MW) rejected the I(1) null for the sample of real exchange rates although the simulation results suggest that they can hardly distinguish I(1) process from nonlinear stationary process. Moreover the panel G-test, which is shown to have a fairly good discriminatory power between nonlinear stationary series and I(1) processes, rejected the I(0) null for the sample of 20 real exchange rates.

#### 4.2. Mixed structure of panel

The second possibility responsible for the joint rejection outcome is a potential mix of linear I(0) and I(1) series in the panel. Given that panel tests are built under the null hypothesis that all series in the panel are either I(1) or I(0), the nulls are subject to reject if the panel consists of both I(0) and I(1) series. In addition the Monte Carlo simulation results reported in Table 4 suggests that joint rejection by the combination of the panel G-test and the IV test is a good indicator of mixed panels. Recalling that the two tests have rejected their respective nulls for the sample of 20 industrial real exchange rates, mixed structure of panel looks more relevant to the puzzling empirical result. Then which real exchange rates are I(0) and which ones are I(1)? I now address this question using a sequential classification method.

##### 4.2.1. Sequential classification method

In principle the constituents of a mixed panel can be classified into a group of I(0) series and a group of I(1) series. An immediate but naive strategy to classify them may be to implement univariate unit root tests to individual series. As is well known, however, the performance of extant univariate tests is not reliable enough for this purpose. An alternative strategy exercised here is *sequential classification method* which involves applying confirmatory analysis to sub-samples of the mixed panel in sequential manner. The basic idea of this strategy is that if a panel of  $N$  cross section dimension is composed of  $N_0$  I(0) series and  $N_1$  I(1) series, where  $N = N_0 + N_1$ , confirmatory analysis will keep producing joint rejection outcomes until the original panel is successfully classified into two homogeneous sub-panels, a panel with only I(0) series and the other with only I(1) series.

There are two approaches to classifying a mixed panel, *bottom-up* (specific to general) and *top-down* (general to specific). In the bottom-up approach, we begin with a subset of series which are more homogeneous than the others judging from a prior information or a pre-testing procedure.<sup>19</sup> The selected subset could be either I(0) or I(1) depending on the interest of researcher. Then add a series from the remainders to the subset and perform confirmatory analysis. If the added series

<sup>18</sup> Of course, this does not necessarily mean that all nonlinear models are virtually untenable. Rather, it implies that the nonlinear models considered here are not compatible with the empirical results.

<sup>19</sup> For instance, confirmatory analysis based on univariate tests can be used for pre-testing purpose.



has the same dynamics with the series in the subset, two panel tests under competing null hypotheses will reach an agreement, whereas they will reject their respective nulls if it has different dynamics. The added series will be saved in the subset if an agreement occurs, while it will be replaced by another series among the remainders if joint rejection is obtained. This procedure will continue until confirmatory analysis produces agreement outcomes both on the subset and on its complement. In the top-down approach, on the other hand, we start with taking out a series from the original mixed panel and implement confirmatory analysis on the remainder. If the removal yields an agreement outcome, the classification process will be over. Otherwise, further classification will be followed by replacing it with another series, ensuing  $N$  different combinations each with  $N - 1$  series. If two tests do not reach agreement on any of ' $N$ ' combination, we proceed to remove two series from the original panel, resulting in ' $N \times (N - 1)$ ' combinations of panel each with  $N - 2$  units, and perform confirmatory analysis. The process will continue until the panel tests agree on two classified sub-samples.

#### 4.2.2. *Classifying the mixed panel*

Because of the relative computational convenience, the bottom-up approach is adopted here to classify the panel of 20 real exchange rates. I begin with a sample of five real exchange rates (vis-a-vis the U.S. dollar), British pound, Finnish markka, Dutch guilder, New Zealand dollar, and Spanish peseta, as a homogeneous  $I(0)$  sub-group formed by exploiting the results of recent empirical studies in the PPP literature.<sup>20</sup> The stationarity of these five series, denoted as group-1, is also reinforced by confirmatory analysis. The panel G-test and the NH test concurrently do not reject the  $I(0)$  null for group-1 at the 10% significance level, whereas the four  $I(1)$  tests consistently reject the  $I(1)$  null even at the 5%, providing a strong confirmatory evidence on the stationarity of the sample. It is worth noting that the result is robust to the choice of base currency, albeit the p-value of the panel G-test is somewhat larger with the DM compared to the USD. For the sample of the remaining 15 series, denoted as group-2, we cannot draw a confirmatory inference. Although the two  $I(0)$  tests consistently reject the null at the 10%, the  $I(1)$  tests produce mixed results on the sample. Specifically, the IPS test with the DM as numeraire and the MW test cannot reject the  $I(1)$  null, while the other two  $I(1)$  tests can reject the null at the 10%. Combining together, it is reasonable to conjecture that group-2 is a mixed panel which necessitates a further classification.

Table 5 presents two homogeneous sub-groups obtained from the bottom-up sequential classification approach: group-3 consists of fourteen  $I(0)$  real exchange rates and group-4 embraces the other six real exchange rates of Australian dollar,

---

<sup>20</sup> Using the KPSS test and the DF-GLS test due to Elliott et al. (1996), Caner and Kilian (2001, Table 2 in page 650) report that the two univariate tests under conflicting null hypotheses are in agreement over all 20 quarterly real exchange rates considered - fifteen of them are  $I(0)$  and the rest are  $I(1)$ . The five series selected here are based on the strength of stationarity evidence reported in Culver and Papell (1999, Table 2 in page 757), Murray and Papell (2002, Table 8 in page 16), and Wu and Wu (2001, Table 1 in page 808).

Table 5  
Sub-groups and sequential classification process

Panel Test	Null Hypothesis	Base Currency	starting sample		after classification	
			Group-1	Group-2	Group-3	Group-4
PG	I(0)	USD	0.1658	0.0006	0.2314	0.0000
		DM	0.5368	0.0770	0.1658	0.0000
IPS	I(1)	USD	0.0006	0.0698	0.0130	0.2888
		DM	0.0102	0.1270	0.0392	0.5316
LLC	I(1)	USD	0.0006	0.0286	0.0006	0.1612
MW	I(1)	USD	0.0012	0.1918	0.0470	0.3976
NH	I(0)	USD	0.3962	0.0156	0.8638	0.0962
IV	I(1)	USD	-2.9742***	-3.6950***	-4.9213***	-1.0400

Notes: Entries represent p-values except for the IV-test where test statistics are reported. Group-1 includes 5 countries (Finland, the Netherlands, New Zealand, Spain, and UK) and Group-2 consists of the rest 15 countries. Group-3 contains 14 countries (Austria, Belgium, Denmark, Finland, France, Germany, Greece, Italy, the Netherlands, New Zealand, Norway, Spain, Sweden, and U.K.) and Group-4 is comprised of 6 countries (Australia, Canada, Ireland, Japan, Portugal, and Switzerland) classified as I(1) countries.

\*\*\* Denotes the cases where the null hypothesis can be rejected at the one percent significance level.

Canadian dollar, Irish pound, Japanese yen, Portuguese escudo, and the Swiss franc (relative to the U.S. dollar) categorized as I(1) processes. As shown in Table 5, confirmatory analysis provides a strong agreement on the stationarity for group-3 as the two I(0) panel tests fail to reject the null at the 10% and the four I(1) panel tests consistently reject the unit root null even at the 5%. For group-4, confirmatory analysis also produces a solid accord on the I(1) as the I(0) tests reject the null while the I(1) tests fail to reject the null at the 10% significance level. Interestingly, the results do not appear to be sensitive to numeraire currency. Table 6 summarizes the countries belong to each group.

#### 4.3. The robustness of the econometric techniques

The direct implication of the sequential classification result is that although long-run PPP does not hold in all industrial countries under study, it does so in the

Table 6  
Countries by Group

Group	Countries
I(0)	Austria, Belgium, Denmark, Finland, France, Germany, Greece, Italy, the Netherlands, New Zealand, Norway, Spain, Sweden, U.K.
I(1)	Australia, Canada, Ireland, Japan, Portugal, Switzerland

Note: The U.S. dollar as the base currency.

majority of them (14 out of 20). This finding sheds some light on the inconclusive evidence on long-run PPP, but it challenges the standard findings in the empirical PPP literature and thus poses a question of how reliable the finding is compared to the extant evidence of PPP. Given that the data set used here is similar to those used in other studies, the inconsistency is more likely due to the difference in the employed econometric tools.

The panel techniques used in this study are not entirely new to the PPP literature. Most of them have been adopted by numerous authors to study the mean reversion of real exchange rates during the floating era. The current study largely replicates their findings, particularly using  $I(1)$  panel tests which unanimously reject the null hypothesis at conventional significance levels.<sup>21</sup> However unlike earlier studies in which the rejection is interpreted as a convincing evidence of PPP, this study attributes it to the mixed structure of panel because the interpretation is not confirmed by  $I(0)$  panel tests.<sup>22</sup>

However, since not much is known about the finite sample performances of the panel G-test and confirmatory analysis, it is instructive to ensure that the result in this paper is not driven by their poor finite sample performances. To this end, this section reports some Monte Carlo simulation experiments to evaluate the overall finite sample performances of the panel G-test and confirmatory analysis when it is paired with the IV-test. Simulation experiments are designed to consider three different panel structures,  $I(0)$  panel,  $I(1)$  panel, and mixed panel, with panel sizes of  $(N, T) = (10, 100)$  and  $(20, 100)$  comparable to the actual data set. To remind, non-parametric bootstrap method is used for the panel G-test throughout the simulations. Mixed structure of panel is artificially devised such that the first  $N/2$  elements are  $I(1)$  and the next  $N/2$  components are  $I(0)$  which are generated from the maintained DGP shown in the Appendix A. Note that error terms are designed to be cross sectionally dependent across individual series.

Table 7 presents the average test statistics and the corresponding p-values of the panel G-test and the IV-test out of 5,000 replications. Rejection rates are calculated by the number of times out of 5,000 simulations that p-values of the panel G-test are in excess of 0.01, 0.05 and 0.1, or that the IV-test statistics are greater than the critical values of the 1%, 5% and 10% significance levels. The two tests appear to have fairly good small sample properties and their overall performances

---

<sup>21</sup> Using the NH test, Kuo and Mikkola (2001) cannot reject the stationarity null for the 24 industrial countries, whereas the null is rejected here for the 21 industrial countries with the same test. The discrepancy between the two studies primarily comes from the difference in the data span. Their covered time period (1949–1996) spans across the fixed exchange rate regime as well as over the floating period while mine is confined to the post Bretton Woods period.

<sup>22</sup> This conclusion is in accord with more recent panel studies based on diverse panel techniques. For instance, using a modified Johansen-type test proposed by Taylor and Sarno (1998), Cheung and Lai (2000) were unable to find evidence of PPP for the G-7 and European countries. In addition, by testing common and idiosyncratic components of real exchange rates separately, Bai and Ng (2001) report that seven series (Australia, Austria, Germany, Ireland, Japan, Sweden, and Switzerland) are nonstationary. Based on a sequential panel approach, Henin et al. (2001) also find little evidence of stationarity for the real exchange rates of eight out of seventeen industrial countries.

Table 7

Finite sample performance of the panel G-test and the IV-test (T = 100)

		I(0) Panel		Mixed Panel		I(1) Panel	
		PG test	IV test	PG test	IV test	PG test	IV test
N = 10	Statistics	37.8503	-5.4779***	86.4407	-2.6661***	93.7792	0.2032
	P-value	0.3598		0.0678		0.0116	
	1%	0.0346	1.0000	0.7396	0.6510	0.9240	0.0080
	5%	0.1136	1.0000	0.8056	0.8616	0.9580	0.0440
	10%	0.1760	1.0000	0.8406	0.9236	0.9696	0.0876
N = 20	Statistics	76.0784	-7.7623***	178.0654	-3.7278***	190.4516	0.2972
	P-value	0.3350		0.0390		0.0060	
	1%	0.0388	1.0000	0.8466	0.9300	0.9798	0.0060
	5%	0.1262	1.0000	0.8960	0.9840	0.9864	0.0340
	10%	0.1914	1.0000	0.9160	0.9960	0.9888	0.0680

Note: Entries are based on 5000 replications. In I(0) (or I(1)) panel, all series are I(0) (or I(1)). In the mixed panel, half series are I(0) and the other half are I(1). In the DGP, the AR(1) coefficient ( $\rho$ 's) for I(0) series are randomly generated on U[0.80,0.95]. 1%, 5% and 10% respectively represent rejection rates which are calculated by the number of times out of 5,000 simulations that p-values of the panel G-test are in excess of 0.01, 0.05 and 0.1, or that the IV-test statistics are greater than the critical values of the 1%, 5% and 10% significance levels.

\*\*\* Denotes significance at the one percent level.

improve as  $N$  increases from 10 to 20. The frequencies of rejecting true null hypothesis (or the size of test), represented by the rejection rates of the panel G-test for I(0) panel and those of the IV-test for I(1) panel, are not far from the nominal sizes.<sup>23</sup>

Despite the original intuition that the need to account for cross sectional dependence through bootstrap method may reduce its power, the discriminatory power of the panel G-test at the 10% significance level is approximately 0.9 for mixed panel and it is close to one for I(1) panel. At the same significance level, the rejection rates of the IV-test are also close to 1 for mixed panel, implying that the combination of the two tests can serve as a reliable tool for detecting panel with mixed structures. This is mirrored in Table 8 which exhibits the frequencies of four possible outcomes of confirmatory analysis when the panel G-test is combined with the IV-test. For mixed panel, the combination yields a high frequency of joint rejection (R-R) outcomes, more than 90% of the time for  $N = 20$ . Curiously, the frequency of joint nonrejection (A-A) outcomes is close to nil, reflecting the improved power performance of panel tests. When panel is comprised of only I(0) or I(1) series, the combination produces a high frequency of correct inference (A-R or R-A). Table 8 also reports evidence that panel analysis results in a substantial improvement in precision over univariate counterpart in confirmatory analysis. For example, the

<sup>23</sup> Although the IV test mildly underrejects while the panel G-test tends to overreject, it should be noted that the size distortion is not as serious as the case reported in O'Connell (1998a).

Table 8  
Finite sample performance of confirmatory analysis (10% significance level)

test combinations	DGP	A-A	A-R	R-A	R-R
PG test and IV test	I(0) panel	0.0	<b>81.9</b>	0.0	18.1
	mixed panel	0.0	8.6	0.6	<b>90.8</b>
	I(1) panel	0.7	0.4	<b>92.5</b>	6.4
KPSS test and DF-GLS test	I(0) series	11.6	<b>62.1</b>	9.9	16.4
	I(1) series	16.6	19.4	<b>45.3</b>	18.7

*Note:* Entries are based on 5000 replications with the panel size of  $N = 20$ ,  $T = 100$ . See the Appendix A for the details of data generation. A-A denotes the fraction of times when both tests under the conflicting null hypotheses fail to reject their respective nulls. A-R denotes the fraction of times when the first test fails to reject the null while the other test rejects the null. R-A denotes the fraction of times when the first test rejects the null while the other test fails to reject the null. R-R denotes the fraction of times when both tests reject their respective nulls. Bold face indicates the portion of correct inference.

combination of the KPSS test and the DF-GLS test due to Elliott et al. (1996), which is adopted in Caner and Kilian (2001), identifies I(0) series correctly just 6 times out of 10 while the accuracy even drops below 50% for I(1) series. It is thus risky to rely on univariate tests to sort out mixed panel.

Overall, the simulation results thereby lend credibility to the econometric techniques adopted in the current study.

## 5. Underlying factors of the departure from PPP

So far we have seen that the panel of 20 real exchange rates can be classified into two sub-samples, I(0) group and I(1) group. What then accounts for the different dynamic patterns of real exchange rates between the two groups? Put differently, what elements do the countries in the I(1) group share that set their real exchange rates apart from the long run equilibrium? This section explores the potential underlying factors by linking the observed persistence in PPP deviations to the observable characteristics of country groups.

There is a large empirical literature that studies the causes of deviations from PPP, which include, but not limited to, productivity growth differential (the Balassa–Samuelson effect), inflation, government spending, geographical proximity, and pricing to market (e.g., Froot and Rogoff, 1994; Cheung and Lai, 2000). Table 9 presents the mean values of these variables for the two groups. The average volatility of real exchange rates is far greater when the countries in I(1) group instead of I(0) group serve as the base country, confirming the finding by Wei and Parsley (1995) that the deviations from PPP are positively related to exchange rate volatility. The volatility difference between the two groups is more pronounced in terms of the real effective exchange rate (REER). The productivity growth, measured by the average growth rates of per capita real GDP, is also higher in I(1) group, consistent with the Balassa–Samuelson model which predicts that systematic differential in productivity growth leads to a permanent deviation from PPP. This pattern is more obvious when the differentials in productivity growth are normal-

Table 9  
Empirical determinants of the departures from PPP

	I(1) Group	I(0) Group
Volatility	0.54 {0.062}	0.30 {0.045}
Volatility of REER	117.43	49.23
Productivity Growth	3.12 (1.00) [0.92]	2.64 (0.62) [0.43]
Inflation	7.86 {5.40}	7.30 {6.16}
Government Expenditure	0.31	0.36
Openness	0.59	0.64

*Note:* The entries denote the average of variables across countries in each group. Volatility represents the average volatility of log real exchange rates with countries in each group as numeraire. Volatility in the curved bracket are the volatility of the first difference of real exchange rates. REER represents the CPI-based real effective exchange rates which are obtained from the IFS CD-Rom (line code REC). The growth rate of per capita real GDP during 1960–2000 is used as proxy for productivity growth rate. Productivity growth in the parenthesis and the squared bracket represent normalized productivity growths relative to the U.S. and Germany, respectively. Annual inflation rates are measured by CPI. Inflation rates in the curved bracket excludes Portugal from I(1) group and Greece from I(0) group. Government expenditure denotes the ratio of government expenditure (IFS line code 82) to GDP. Openness is measured as the ratio of the sum of imports and exports to the size of GDP. Figures in the parenthesis and the square brackets respectively represent the normalized differentials from U.S. and Germany.

ized with respect to U.S. (or Germany). The normalized productivity growth of the I(1) group is almost twice as high as that of I(0) group. The effect of inflation, however, appears to be somewhat puzzling. The parity is known to hold better for high inflation countries, but the countries in I(1) group have experienced a higher inflation on average during the post Bretton Woods era.<sup>24</sup> The effects of government spending and geographical proximity are rather counter-intuitive or at least irrelevant. Since government spending tends to fall more heavily on nontraded goods, it usually induces a real appreciation (Rogoff, 1996). However, as can be seen in Table 9, the average ratio of government expenditure to GDP is slightly lower in the countries belong to I(1) group. Geographical proximity is also inapplicable in the sense that geographically neighboring countries such as Australia and New Zealand, U.K. and Ireland, Spain and Portugal exhibit quite different dynamics of real exchange rates. The openness of economy turns out to be inconsequential as well since the two groups exhibit little difference in the ratio of the sum of imports and exports to the size of GDP. Overall, the general bond between selected observable macroeconomic variables and the dynamics of real exchange rates in industrial countries is not that strong.

Attention is now redirected toward the factors specific to the countries. Indeed Bai and Ng (2001) observe that idiosyncratic components are the main underlying source of nonstationarity in real exchange rates. As is widely known, the Japanese

<sup>24</sup> This result is reversed when two relatively high inflation countries, Greece and Portugal, are dropped.

Table 10  
Testing stationarity for relative prices of traded and non-traded

Countries	Traded		Non-traded	
	ADF	KPSS	ADF	KPSS
Canada	-1.8715	0.1164	-0.9303	<b>0.5912</b>
France	-1.4579	0.2050	-1.4600	0.1262
Italy	-1.3943	0.4374	-1.4078	0.3791
Japan	-0.9806	<b>0.6139</b>	-1.4741	<b>0.5109</b>
U.K.	-2.3612	0.3032	-2.2710	0.2850

*Note:* Real exchange rates are constructed from IFS CD-ROM data on end-of-period nominal exchange rates relative to US dollar and from JB Kim on CPI of traded and non-traded goods during 1994.21–1998.24. For the lag length selection, the sequential t-test is used in the ADF test by setting the maximum lag length at 8 and the fixed lag length rule of  $l = \text{integer} [12(T/100)^{1/4}]$  is used in the KPSS test. Bold face represents a rejection of the null hypothesis at the five percent significance level.

yen has steadily appreciated against the U.S. dollar during the post Bretton Woods period and thus more likely to violate the PPP hypothesis. The Canadian dollar is also reported to have fragile evidence of PPP (e.g., Flores et al., 1999). This argument is supported by Table 10 which presents the results of univariate unit root tests for sectorally decomposed real exchange rates (relative to the U.S. dollar). According to the commodity-arbitrage view of PPP, the law of one price holds only in traded goods and the departures from PPP is primarily attributed to the large weight placed on nontraded goods in the CPI. Among the five countries considered,<sup>25</sup> evidence of PPP can be observed in neither sector for the Japanese and only for the traded good sector for the Canadian dollar not contradictory to my classification results.<sup>26</sup>

For the other I(1) group countries, country-specific factors are less concrete. Nonetheless Papell and Theodoridis (2001) find a much weaker evidence of PPP for Australia compared to its neighbor, New Zealand. According to Papell and Theodoridis (2001), Australia maintained a fixed trade-weighted nominal exchange rate regime until the early 1980s and floated thereafter. The Irish pound was pegged to the British pound until 1979 when it joined EMS. Though Ireland has effectively linked its nominal exchange rate to the Deutsche mark afterwards, a large inflation differential with its main trading partner, the U.K., is believed to impart an unstable pattern in real exchange rates during the 1980s. The Swiss economy has been characterized by relatively low rates of inflation and a trend appreciation in the real effective exchange rate until the mid 1990s. Portugal has experienced two turning points around its accession to the European Union in 1986 and the ERM

<sup>25</sup> Due to the data availability, only five bilateral exchange rates are considered here. I thank Jaebeom Kim for sharing the data.

<sup>26</sup> Bai and Ng (2001) also find that the real exchange rate variations of Canada are dominated by non-stationary idiosyncratic components.

crisis in 1992 which might have contributed to the departure of real exchange rate from long-run equilibrium level.

## 6. Concluding remarks

This paper has reexamined the issue of long-run PPP using multiple panel tests in the framework of confirmatory analysis. Application of six panel tests under competing null hypotheses to the real exchange rates of 21 industrial countries produces seemingly contradictory evidence on long-run PPP during the post-Bretton Woods period. Irrespective of numeraire currency, four I(1) panel tests unanimously reject the null hypothesis in favor of long-run PPP, whereas two I(0) panel tests lend little support to the parity by consistently rejecting the null at conventional significance levels.

This puzzling result can be generated either when the real exchange rates involve nonlinear dynamics or when the panel is comprised of both I(0) and I(1) series. A Monte Carlo study indicates that a potential mixture of I(0) and I(1) series is more relevant to the empirical finding. The use of a sequential classification method as a strategy to sort out the mixed panel reveals that long-run PPP holds in the majority of currencies with the possible exception of six currencies which display most persistent deviations from long-run equilibrium levels. Systematic behavior of these series can be characterized better by country specific factors than by observable macroeconomic variables.

The central message from this study is straightforward. In view of growing evidence on the substantial cross-country differences in the behavior of real exchange rates, it is crucial for empirical researchers to exercise caution in implementing panel tests whose power gains mainly come from the assumption that PPP holds equally well for every country. We need to separate the wheat from the chaff before making flour. The current study moves a step toward this direction.

## Acknowledgements

I thank seminar participants at the University of New Hampshire, the Ohio State University, University of Oregon, and the 11th Annual Meeting of the Midwest Econometrics Conference at Kansas City. Special thanks go to Nelson C. Mark, Mike Goldberg, Donggyu Sul, Robert Mohr, and an anonymous referee for valuable comments and suggestions and to Jaebeom Kim for sharing data. Financial support through Summer Research Fund from the University of New Hampshire is also gratefully acknowledged. Any remaining errors are the author's.

## Appendix A

### *A.1. Nonparametric bootstrap procedures*

The residual-based nonparametric bootstrap method is used as follows.



First, by using the iterative seemingly unrelated regression (SUR) method, fit the following equation to get an estimator of the parameters ( $\hat{\alpha}_i$ ,  $\hat{\rho}_i$ , and  $\hat{\gamma}_{ij}$ ) together with  $\hat{\varepsilon}_{it}$ , the fitted residuals of  $\varepsilon_{it}$ ,

$$q_{it} = \alpha_i + \rho_i q_{i,t-1} + \sum_{j=1}^{k_i} \gamma_{ij} \Delta q_{i,t-j} + \varepsilon_{it}, \quad (13)$$

where  $k_i$  is chosen from data using Hall's (1994) method.

Second, to account for cross sectional dependence, estimate the variance and covariance of  $\varepsilon_{it}$ ,  $\Sigma$ , by  $\hat{\Sigma} = (1/T) \sum_{t=1}^T \hat{\varepsilon}_t \hat{\varepsilon}_t'$  where  $\hat{\varepsilon}_t = (\hat{\varepsilon}_{1t}, \dots, \hat{\varepsilon}_{Nt})$  is the vector of residuals using the iterative SUR method.

Third, resample the estimated residuals with a cross section index fixed in order to preserve cross sectional dependence among individual series. This is nonparametric bootstrap since error terms are drawn using the moving block method without further assumption on the error term distribution. Then, generate the pseudo-observations ( $q_{it}^*$ ) for  $q_{it}$  following Eq. (14). Initial values of  $q_{it}^*$  are obtained from block resampling as described in Berkowitz and Kilian (2000) by dividing  $q_{it}$  into  $T - k$  overlapping blocks with length  $k + 1$  and choose a block randomly with replacement for  $q_{it}^*$ .

$$q_{it}^* = \hat{\alpha}_i + \hat{\rho}_i q_{i,t-1} + \sum_{j=1}^{k_i} \hat{\gamma}_{ij} \Delta q_{i,t-j} + \hat{\varepsilon}_{it}^*, \quad (14)$$

where  $\hat{\alpha}_i$ ,  $\hat{\gamma}_i$  and  $\hat{\rho}_i$  are the SUR estimators obtained from the first step and  $\hat{\varepsilon}_{it}^*$  is a pseudo-innovation drawn from the resampling.

Finally, run the panel tests on the pseudo-data,  $q_{it}^*$ , to derive empirical distribution of the test statistics and the corresponding p-values. The number of replications used in each experiment is 5,000.

Despite the computational ease to accommodate an arbitrary pattern of cross-sectional dependence, however, one problem associated with the use of the SUR method is that it may not be appropriate if the panel under study has a bigger dimension in cross-section ( $N$ ) than in time series ( $T$ ) because SUR estimators are known to have low accuracy unless  $T$  is appreciably higher than  $N$ . Fortunately, this is not the case for the data employed here as  $T = 104$  is far exceeding  $N = 20$ .

## A.2. Monte Carlo simulation design

The following maintained DGP is used for simulations in the paper unless specified otherwise.

$$y_{it} = (1 - \rho_i) \alpha_i + \rho_i y_{i,t-1} + u_{it}, \quad (15)$$

where  $i = 1, 2, \dots, N$ ,  $t = 1, 2, \dots, T$  and  $y_{i0}$  is randomly selected.  $y_{it}$  is I(1) process if  $\rho_i = 1$ , whereas  $y_{it}$  will be I(0) process when  $|\rho_i| < 1$ .  $\rho_i$  and  $\alpha_i$  are randomly generated on U[0.8,0.95] and N(0,1), respectively, and they are fixed at their realized values after the draw. The error term  $u_{it}$  is set to follow an AR(1) process,  $u_{it} = \theta_i u_{i,t-1} + \varepsilon_{it}$ ,  $\theta_i$  is randomly generated on U[0.2,0.4] where U denotes the

uniform distribution and they are different for each  $i$  while fixed for each model after selected. Cross sectional dependence is incorporated across the error terms following Chang (2002). Specifically, the innovations in the error term,  $\varepsilon_t = (\varepsilon_{1t}, \dots, \varepsilon_{Nt})'$  are drawn from an  $N$ -dimensional multivariate normal distribution with mean zero and covariance matrix  $\Sigma$  which is generated following the steps outlined in Chang (2002). First, generate an  $(N \times N)$  matrix  $M$  from  $U[0,1]$  to construct an orthogonal matrix  $H = M(M'M)^{-1/2}$ . Second, generate a set of  $N$  eigenvalues,  $\lambda_1, \dots, \lambda_N$  where  $\lambda_1 = r > 0$ ,  $\lambda_N = 1$  and draw  $\lambda_2, \dots, \lambda_{N-1}$  from  $U[r,1]$  to ensure symmetry and nonsingularity of  $\Sigma$ .  $\tau$  is set at 0.1. Next, form a diagonal matrix  $\Lambda$  with  $(\lambda_1, \dots, \lambda_N)$  on the diagonal. Then, construct the covariance matrix  $\Sigma$  as a spectral representation  $\Sigma = H\Lambda H'$ . Cross-sectional heterogeneity is allowed in the DGP by the random generation of  $\theta_i$ ,  $\rho_i$  and  $\alpha_i$ . The first 100 observations of  $y_{it}$  are discarded after generating extra  $y$ 's.

Mixed panels are contrived by setting the portion of  $I(1)$  series in the panel to be a half. That is, for the panel of  $N = 10$ ,  $\rho_i = 1$  for 5 series and  $0.8 < \rho_i < 0.95$  for the rest five series in Eq. (15).

## References

- Abuaf, N., Jorion, P., 1990. Purchasing Power Parity in the Long Run. *Journal of Finance* 45 (1), 157–174.
- Bai, J., Ng, S., 2001. A New Look at Panel Testing of Stationarity and the PPP Hypothesis. Unpublished manuscript.
- Banerjee, A., Marcellino, M., Osbat, C., 2001. Testing for PPP: Should We Use Panel Methods? Unpublished manuscript.
- Baum, C.F., Barkoulas, J.T., Caglayan, M., 2001. Nonlinear Adjustment to Purchasing Power Parity in the Post-Bretton Woods Era. *Journal of International Money and Finance* 20 (3), 379–399.
- Berkowitz, J., Kilian, L., 2000. Recent Development in Bootstrapping Time Series. *Econometric Reviews* 19 (1), 1–48.
- Bowman, D., 1997. Efficient Tests for Autogressive Unit Roots in Panel Data. Unpublished manuscript.
- Breuer, J.B., McNown, R., Wallace, M.S., 2001. Misleading Inferences from Panel Unit Root Tests with an Illustration from PPP. *Review of International Economics* 9 (3), 482–493.
- Caner, M., Kilian, L., 2001. Size Distortions of Tests of the Null Hypothesis of Stationarity: Evidence and Implications for the PPP Debate. *Journal of International Money and Finance* 20 (5), 639–657.
- Carrion-i-Silvestre, J.L., Sansó-i-Rosselló, A., Ortuno, M.A., 2001. Unit Root And Stationarity Tests' Wedding. *Economics Letters* 70 (1), 1–8.
- Chang, Y., 2004. Bootstrap Unit Root Tests in Panels with Cross-sectional dependency. *Journal of Econometrics* 120 (2), 263–293.
- Chang, Y., 2002. Nonlinear IV Unit Root Tests in Panels with Cross-sectional Dependency. *Journal of Econometrics* 110 (2), 261–292.
- Charemza, W.W., Syczewska, E.M., 1998. Joint Application of the Dickey-Fuller and KPSS Tests. *Economics Letters* 61 (1), 17–21.
- Cheung, Y.W., Lai, K.S., 2000. On Cross-Country Differences in the Persistence of Real Exchange Rates. *Journal of International Economics* 50 (2), 375–397.
- Choi, C.Y., 2001a. Panel Unit Root Tests under the Null Hypothesis of Stationarity. Unpublished manuscript.
- Choi, C.Y., 2001b. Confirmatory Analysis in Panel Unit Root Tests. Unpublished manuscript.
- Choi, I., 2001c. Unit Root Tests for Panel Data. *Journal of International Money and Finance* 20 (2), 249–272.

- Culver, S.E., Papell, D.H., 1999. Long-Run Purchasing Power Parity with Short-Run Data: Evidence with a Null Hypothesis of Stationarity. *Journal of International Money and Finance* 18 (5), 751–768.
- Dumas, B., 1992. Dynamic Equilibrium and the Real Exchange Rate in A Spatially Separated World. *Review of Financial Studies* 5 (2), 153–180.
- Elliott, G., Rothenberg, T.J., Stock, J.H., 1996. Efficient Tests for an Autoregressive Unit Root. *Econometrica* 64 (4), 813–836.
- Engel, C., Kim, C.J., 1999. The Long-Run U.S./U.K. Real Exchange Rate. *Journal of Money, Credit and Banking* 31 (3), 335–356.
- Flores, R., Jorion, P., Preumont, P., Szafarz, A., 1999. Multivariate Unit Root Tests of the PPP Hypothesis. *Journal of Empirical Finance* 6 (4), 335–353.
- Froot, K.A., Rogoff, K., 1994. Perspectives on PPP and Long-run Real Exchange Rates. NBER Working Paper No. 4952.
- Hall, A., 1994. Testing for A Unit Root in Time Series with Pretest Data Based Model Selection. *Journal of Business and Economic Statistics* 12 (4), 461–470.
- Henin, P., Jolivaldt, P., Nguyen, A., 2001. Testing for Unit Roots on Heterogeneous Panels: A Sequential Approach. Unpublished manuscript.
- Higgins, M., Zakrajsek, E., 1999. Purchasing Power Parity: Three Stakes through the Heart of the Unit Root Null. Unpublished manuscript.
- Im, K.S., Pesaran, M.H., Shin, Y., 2003. Testing for Unit Roots in Heterogeneous Panels. *Journal of Econometrics* 115 (1), 53–74.
- Jorion, P., Sweeney, R.J., 1996. Mean Reversion in Real Exchange Rates: Evidence and Implications for Forecasting. *Journal of International Money and Finance* 15 (4), 535–550.
- Koedijk, K.G., Schotman, P.C., van Dijk, M.A., 1998. The Re-Emergence of PPP in the 1990s. *Journal of International Money and Finance* 17 (1), 51–61.
- Kuo, B.S., Mikkola, A., 2001. How Sure Are We About PPP? Panel Evidence with the Null of Stationary Real Exchange Rates. *Journal of Money, Credit and Banking* 33 (3), 767–789.
- Kwiatkowski, D., Phillips, P.C.B., Schmidt, P., Shin, Y., 1992. Testing the Null Hypothesis of Stationary against the Alternative of a unit Root. *Journal of Econometrics* 54 (1–3), 159–178.
- Levin, A., Lin, C.F., Chu, C.J., 2002. Unit Root Tests in Panel Data: Asymptotic and Finite-Sample Properties. *Journal of Econometrics* 108 (1), 1–24.
- Lothian, J.R., 1998. Some New Stylized Facts of Floating Exchange Rates. *Journal of International Money and Finance* 17 (1), 29–39.
- Lyhagen, J., 2000. Why Not Use Standard Panel Unit Root Test for Testing PPP. Unpublished manuscript.
- MacDonald, R., 1996. Panel Unit Root Tests and Real Exchange Rates. *Economics Letters* 50 (1), 7–11.
- Maddala, G.S., Wu, S., 1999. A Comparative Study of Unit Root Tests with Panel Data and a New Simple Test. *Oxford Bulletin of Economics and Statistics* 61, 631–652.
- Mark, N.C., 2001. *International Macroeconomics and Finance: Theory and Empirical Methods*. Blackwell Publishing Company.
- Mark, N.C., Sul, D., 2001. Nominal Exchange Rates and Monetary Fundamentals: Evidence from A Small Post-Bretton Woods Panel. *Journal of International Economics* 53 (1), 29–52.
- Michael, P., Nobay, A.R., Peel, D.A., 1997. Transaction Costs and Nonlinear Adjustment in Real Exchange Rates: An Empirical Investigation. *Journal of Political Economy* 105 (4), 862–879.
- Murray, C.J., Papell, D.H., 2002. The Purchasing Power Parity Persistence Paradigm. *Journal of International Economics* 56 (1), 1–19.
- Newey, W., West, K., 1987. A Simple Positive-Definitive, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica* 53 (3), 703–708.
- Nyblom, J., Harvey, A., 2000. Testing for Common Stochastic Trends. *Econometric Theory* 16 (2), 176–199.
- Obstfeld, M., Taylor, A.M., 1997. Nonlinear aspects of goods-market arbitrage and adjustment: Heckscher's commodity points revisited. *Journal of the Japanese and International Economics* 11 (4), 441–479.

- O'Connell, P.G.J., 1998a. The Overvaluation of Purchasing Power Parity. *Journal of International Economics* 44 (1), 1–19.
- O'Connell, P.G.J., 1998b. Market Frictions And Real Exchange Rates. *Journal of International Money and Finance* 17 (1), 71–95.
- Papell, D.H., 1997. Searching for Stationarity: Purchasing Power Parity under the Current Float. *Journal of International Economics* 43 (3–4), 313–332.
- Papell, D.H., Theodoridis, H., 2001. The Choice of Numeraire Currency in Panel Tests of Purchasing Power Parity. *Journal for Money, Credit and Banking* 33 (3), 790–803.
- Park, J.Y., 1990. Testing for Unit Roots and Cointegration by Variable Addition. *Advances in Econometrics* 8, 107–133.
- Park, J.Y., Choi, B., 1988. A New Approach to Testing for a Unit Root. Unpublished manuscript.
- Parsley, D.C., Popper, H.A., 2001. Official Exchange Rate Arrangements and Real Exchange Rate Behavior. *Journal of Money, Credit and Banking* 33 (4), 976–993.
- Phillips, P.C.B., Sul, D., 2003. Dynamic Panel Estimation and Homogeneity Testing Under Cross Section Dependence. *Econometrics Journal* 6, 217–260.
- Rogoff, K., 1996. The Purchasing Power Parity Puzzle. *Journal of Economic Literature* 34 (2), 647–668.
- Sercu, P., Uppal, R., Hulle, C.V., 1995. The Exchange Rate in the Presence of Transaction Costs: Implications for Tests of Purchasing Power Parity. *Journal of Finance* 50 (4), 1309–1319.
- Taylor, A.M., 2001. Potential Pitfalls for the PPP Puzzle? Sampling and Specification Biases in Mean-Reversion Tests of the Law of One Price. *Econometrica* 69 (2), 473–498.
- Taylor, M.P., Sarno, L., 1998. The Behavior of Real Exchange Rates During the Post-Bretton Woods Period. *Journal of International Economics* 46 (2), 281–312.
- Taylor, M.P., Peel, D.A., Sarno, L., 2001. Nonlinear Mean-Reversion in Real Exchange Rates: Toward A Solution to the Purchasing Power Parity Puzzles. *International Economics Review* 42 (4), 1015–1042.
- Wei, S.J., Parsley, D.C., 1995. Purchasing Power Parity during the Floating Rate Period: Exchange Rate Volatility, Trade Barriers and Other Culprits. NBER Working Paper No. 5032.
- Wu, J.L., Wu, S., 2001. Is Purchasing Power Parity Overvalued? *Journal of Money, Credit and Banking* 33 (3), 804–812.
- Wu, Y., 1996. Are Real Exchange Rates Non-stationary? Evidence from a Panel-Data Test. *Journal of Money, Credit and Banking* 28 (1), 54–63.