



Contents lists available at ScienceDirect

Journal of International Money and Finance

journal homepage: www.elsevier.com/locate/jimf



Why panel tests of purchasing power parity should allow for heterogeneous mean reversion

Kees G. Koedijk^{a,*}, Ben Tims^b, Mathijs A. van Dijk^b

^a Tilburg University and CEPR, PO Box 90153, 5000 LE Tilburg, The Netherlands

^b Rotterdam School of Management, Erasmus University, PO Box 1738, 3000 DR Rotterdam, The Netherlands

A B S T R A C T

JEL classification:

F31
F33
F47

Keywords:

PPP
Real exchange rates
Panel models
Unit root tests
Heterogeneity
Panel tests

Recent studies of purchasing power parity (PPP) use panel tests that fail to take into account heterogeneity in the speed of mean reversion across real exchange rates. In contrast to several other severe restrictions of panel models and tests of PPP, the assumption of homogeneous mean reversion is still widely used and its consequences are virtually unexplored. This paper analyzes the properties of homogeneous and heterogeneous panel unit root testing methodologies. Using Monte Carlo simulation, we uncover important adverse properties of the panel approach that relies on homogeneous estimation and testing. More specifically, power functions are low and assume irregular shapes. Furthermore, homogeneous estimates of the mean reversion parameters exhibit potentially large biases. These properties can lead to misleading inferences on the validity of PPP. Our findings highlight the importance of allowing for heterogeneous estimation when testing for a unit root in panels of real exchange rates.

© 2010 Elsevier Ltd. All rights reserved.

1. Introduction

Since the early 1990s, researchers interested in testing the hypothesis of purchasing power parity (PPP) have turned to panel tests to increase the statistical power.¹ Initial applications of panel models to real exchange rates imposed severe restrictions on the structure of the model. Two of these restrictions have

* Corresponding author. Tel.: +31 10 40 82300; fax: +31 10 40 89017.

E-mail addresses: c.koedijk@uvt.nl (K.G. Koedijk), btims@rsm.nl (B. Tims), madijk@rsm.nl (M.A. van Dijk).

¹ Examples of early panel PPP studies include Abuaf and Jorion (1990), Frankel and Rose (1996), and Jorion and Sweeney (1996). We refer to Rogoff (1996) and Taylor and Taylor (2004) for reviews of the literature.

been successfully challenged. First, O'Connell (1998) questions the common assumption in panel PPP studies that the real exchange rates are cross-sectionally independent. He shows that spurious rejections of the unit root null can occur when cross-sectional dependence is neglected. In response to this critique, nearly all subsequent researchers relax this restriction and take cross-sectional dependence into account. Second, Papell and Theodoridis (2001) and Wu and Wu (2001) criticize the common restriction that the serial correlation properties of all real exchange rates in the panel are the same. Both papers show that assuming a restrictive homogeneous serial correlation structure weakens the evidence against the unit root null. In line with these findings, recent panel studies on PPP abandon this second restriction as well.

Yet, the use of a third important restriction on the structure of panel models of real exchange rates is still widespread in the academic literature on PPP. The vast majority of recent panel studies assume a common mean reversion coefficient across all real exchange rates.² From an economic perspective, the justification for the assumption that PPP holds equally well for all country pairs is weak. The speed of mean reversion of a real exchange rate between two countries should depend on, for example, their relative proximity, their mutual trade regulations, and the openness of their economies. The econometric consequences of imposing homogeneous mean reversion for the properties of panel PPP tests have not been thoroughly investigated to date. Exploratory research on the statistical properties of panel data models (notably Pesaran and Smith, 1995; Robertson and Symons, 1992) suggests that the homogeneity assumption in dynamic panel models may have serious consequences. Pooling heterogeneous panel data can lead to biases in the parameter estimates, as a result of which estimated half-lives are potentially misleading. An important question is to what extent this affects the inferences drawn from panel studies on the PPP hypothesis.

This paper compares the properties of three different panel estimation and testing methodologies for investigating PPP. The first involves homogeneous estimation of the mean reversion parameters and a unit root test on the validity of PPP for the full panel of real exchange rates. This methodology, or a variation thereof, is applied by a large number of recent empirical papers on PPP.² The second involves estimating the model heterogeneously, but still testing the PPP hypothesis jointly for all series in the panel. This means that while any inferences about PPP still concern the panel as a whole, the test allows for differences in mean reversion across countries. This approach is followed by, among others, Im et al. (2003), Taylor and Sarno (1998), and Wu and Wu (2001). In addition, we propose a third methodology in which both estimation and testing are heterogeneous.

We employ Monte Carlo simulation to examine the performance of the three panel methodologies. Our Monte Carlo experiments are based on a sample of the real exchange rates between five of the world's largest economies (Canada, the Euro area, Japan, the U.K., and the U.S.) over the period 1978:12–2003:12. Our simulations include "mixed" panels (consisting of stationary and non-stationary series) as well as "heterogeneous" panels (consisting of stationary series with different mean reversion parameters).

We show that when the mean reversion is heterogeneous across real exchange rates, the methodology with homogeneous estimation and testing suffers from important adverse properties. First, the homogeneous estimates of the mean reversion parameter exhibit serious biases. Second, large estimation uncertainties arise as a result of the homogeneity restriction. This implies that the statistical power of the homogeneous test against the unit root null is generally limited. Third, the power function is not monotonically increasing when the mean reversion parameters generated under the alternative hypothesis move away from the unit root null. These properties are observed in both mixed and heterogeneous panels and are robust to the number and the composition of real exchange rates in the panel.

We are among the first to carry out a systematic examination of the estimation biases and the power properties of panel tests of PPP under homogeneous mean reversion.³ Our finding of non-monotonic

² Recent examples include Choi (2004), Fleissig and Strauss (2000), Frankel and Rose (1996), Jorion and Sweeney (1996), Levin et al. (2002), Lopez and Papell (2007), O'Connell (1998), Papell (1997, 2002, 2006), and Papell and Theodoridis (2001). Lothian (1997) also reports the results of panel tests with homogeneous mean reversion, but the focus of his study is on cross-country regressions of real exchange rate changes on nominal exchange rate changes.

³ Breuer et al. (2001, 2002) also study the properties of panel unit root tests to assess the PPP hypotheses. However, they do not present power functions and they mainly focus on comparing heterogeneous panel unit root tests with univariate unit root tests.

power functions of unit root tests in particular (which we refer to as the “boomerang” effect) is a novel and remarkable result. Although the existence of non-monotonic power is not new,⁴ the surprising result is that this phenomenon occurs in simple linear models and that it can be quite severe. The adverse properties of tests based on homogeneous estimation described above can be attributed to two main underlying causes. First, the most persistent real exchange rate dominates the estimation of the homogeneous mean reversion parameter. As a result, a potentially large upward bias arises in the estimated homogeneous mean reversion parameter, which leads to relatively infrequent rejections of the unit root null. Second, we present evidence that suggests that the serial correlation structure of the model drives the boomerang effect. Using a number of procedures most commonly employed in the literature to determine the lag length, we illustrate under which circumstances the boomerang effect is found.

Our paper adds to the literature on the consequences of imposing parameter restrictions in panel models of real exchange rates. We focus on the restriction of homogeneous mean reversion – arguably the strongest assumption from an economic perspective. Our results indicate that a homogeneous estimation methodology can lead to misleading inferences about the validity of the PPP hypothesis. Panel testing methodologies for PPP should therefore not only take cross-sectional dependence and heterogeneous serial correlation into account, but also heterogeneous mean reversion.

2. Methodology

The three panel estimation and testing methodologies analyzed in this paper are all based on the following linear regression:

$$R_{i,t} = \alpha_i + \beta_i R_{i,t-1} + \sum_{k=1}^{l_i} \gamma_{i,k} \Delta R_{i,t-k} + u_{i,t}, \quad (1)$$

where $R_{i,t}$ is the logarithm of the real exchange rate for each country (currency) i ($i = 1, \dots, N$) at time t ($t = 1, \dots, T$) expressed relative to some numeraire currency 0, β_i are the mean reversion parameters, α_i are the intercepts, $\gamma_{i,k}$ are the serial correlation coefficients, l_i denotes the number of lags for currency i , and $u_{i,t}$ is a stationary error term: $u_t \sim N(0, \Sigma)$. The null-hypothesis of a unit root is expressed by $H_0: \beta_i = 1$ for all currencies i . The three methodologies differ in the estimation of the parameters in equation (1) as well as in the way the test of the unit root hypothesis is performed. For all three methodologies, we incorporate the suggestions made by O’Connell (1998) to allow for contemporaneous correlations between the error terms $u_{i,t}$ and by Papell and Theodoridis (2001) and Wu and Wu (2001) to allow the serial correlation parameters $\gamma_{i,k}$ to vary across exchange rates.

In the first methodology, estimation of equation (1) and testing the unit root hypothesis are carried out homogeneously. That is, the restriction $\beta_i = \beta$ is imposed for all i and the null-hypothesis and alternative hypothesis can thus be expressed as $H_0: \beta = 1$ and $H_A: \beta < 1$, respectively. This implies that the PPP hypothesis is only evaluated for the panel as a whole and not for individual country pairs. A large number of recent empirical papers adopt (a variation of) this methodology.² For example, O’Connell (1998) applies this methodology to study the restriction that exchange rates are cross-sectionally independent. Accounting for cross-sectional dependence, he finds no empirical evidence for PPP in a panel of 64 countries over the period 1973:I–1995:IV. Lopez and Papell (2007) investigate convergence toward PPP within the euro zone and between the euro zone and other countries. Using data over the period 1973:I–2001:IV, they present evidence of convergence toward PPP within the euro zone, but not for the real exchange rates of the euro versus other currencies.

The second methodology we investigate involves heterogeneous estimation of the mean reversion parameters β_i in equation (1), but still testing the unit root hypothesis in a homogeneous way. This means that while inferences about PPP concern the entire panel and no statements can be made about individual country pairs, the speed of mean reversion is allowed to differ across countries. The null-

⁴ Nelson and Savin (1990) demonstrate that the power function of a Wald test in a non-linear (exponential) regression model can be non-monotonic.

hypothesis can be expressed as $H_0: \beta_i = 1$ for all i . There are several possibilities for the alternative, for example, $H_A: \beta_i < 1$ for some i or $H_A: \beta_i < 1$ for all i . The interpretation of the first alternative is that at least one of the exchange rates is stationary, while the second states that all exchange rates are mean-reverting. This implies that rejection of the unit root null does not necessarily provide information on how many and which real exchange rates in the panel are stationary. An example of a study using this methodology is Taylor and Sarno (1998). They estimate an equivalent panel model as in equation (1) and perform a joint unit root test on all N equations by constructing a standard Wald test statistic, which they refer to as the Panel Augmented Dickey–Fuller (MADF) test. The alternative hypothesis for this test is $H_A: \beta_i < 1$ for some i . Critical values for this test are obtained by Monte Carlo simulation. The MADF test rejects the unit root null in a panel containing the UK, France, Germany, Japan, and the U.S. (with the U.S. as numeraire country) over the period 1973:I–1996:II. Taylor and Sarno (1998) suggest an alternative testing methodology, based on the Johansen likelihood ratio (JLR) test for cointegration, which tests the unit root null-hypothesis versus the alternative that $H_A: \beta_i < 1$ for all i . The JLR test also rejects the null-hypothesis of non-stationarity for their panel of exchange rates. Other papers that use a methodology in which estimation is performed heterogeneously, but the unit root null-hypothesis is defined for all real exchange rates in the panel simultaneously, are Im et al. (2003) and Wu and Wu (2001). Im et al. (2003) develop a unit root test with the same null-hypothesis as described above, but with alternative hypothesis $H_A: \beta_i < 1, i = 1, \dots, N_1, \beta_i = 1, i = N_1 + 1, \dots, N$. Hence, a rejection of the null-hypothesis in this model does not necessarily mean that all exchange rates in the panel are mean-reverting. Wu and Wu (2001) apply the tests developed in the paper of Maddala and Wu (1999) and in an earlier version of the paper of Im et al. (2003). Both tests are based on univariate Augmented Dickey–Fuller (ADF) regressions and have $\beta_i < 1$ for all i as alternative hypothesis.⁵ Wu and Wu (2001) take heterogeneous serial correlation and cross-sectional dependence into account and document substantial evidence for PPP in a panel of 20 industrial countries over the period 1973:II–1997:IV.

The third methodology entails both heterogeneous estimation and heterogeneous testing of equation (1). The unit root null-hypothesis and alternative hypothesis for every real exchange rate i ($i = 1, \dots, N$) are $H_0: \beta_i = 1$ and $H_A: \beta_i < 1$, respectively. In this framework, the PPP tenet is evaluated for each individual real exchange rate in the panel. Flôres et al. (1999) apply such a methodology and report evidence in favor of long-run PPP for the G10 currencies versus the U.S. dollar in the period 1973:01–1994:12. Koedijk et al. (2004) enhance this approach by allowing for heterogeneous serial correlation. They investigate a panel of ten countries within the Euro area over the period 1973:02–2003:03 as well as a panel of real exchange rates of the euro versus other major currencies over the period 1978:12–2003:03. They conclude that the mean reversion properties differ importantly across real exchange rates. Engel et al. (1997) estimate a panel of real exchange rates for eight cities in four countries and two continents and allow for different speeds of mean reversion for within-country, within-continent, and cross-country city pairs. They fail to reject the unit root hypothesis based on any of these mean reversion parameters.

To avoid confusion, we introduce the following notational shorthand. At the peril of having the paper labeled the “Santa Claus paper,” we refer to the first methodology as HoHo (homogeneous estimation and testing), to the second as HeHo (heterogeneous estimation and homogeneous testing), and to the final as HeHe (heterogeneous estimation and testing).

All three methodologies rely upon Seemingly Unrelated Regression (SUR) estimation of equation (1). We use the univariate Augmented Dickey–Fuller (ADF) τ -statistic to test for a unit root in the HoHo and HeHe methodologies and the MADF (Wald) test statistic of Taylor and Sarno (1998) in HeHo. Appendix A provides details about the estimation and testing procedures as well as the Monte Carlo simulations. To analyze the finite sample properties of the three methodologies, we use equation (1) to simulate real exchange rate series for several data generating processes (DGP's) with certain mean reversion parameters β_i . We follow Taylor and Sarno (1998) and use the following set of values for β_i : {0.99, 0.975, 0.95, 0.925, 0.9}. We use the simulated series to construct power functions. We also track the bias in the estimate of β that can arise because of the homogeneity restriction. More details are in Appendix A.

⁵ We note that this alternative hypothesis is different from the one used in the published version of the Im et al. (2003) paper.

Table 1

Summary statistics and univariate unit root tests. This table presents summary statistics and the results of the univariate ADF unit root test of the real exchange rates of Canada, the Euro area, Japan, and the U.K. versus the U.S. over the period 1981:01–2003:12. We estimate the following equation:

$$R_{i,t} = \alpha_i + \beta_i R_{i,t-1} + \sum_{k=1}^{l_i} \gamma_{i,k} \Delta R_{i,t-k} + u_{i,t},$$

where $R_{i,t}$ is the log of the real exchange rate and the value of l_i is determined by the recursive t -statistic procedure of Campbell and Perron (1991). The critical values for the Augmented Dickey–Fuller (ADF) τ -statistic to test for a unit root are obtained from Dickey and Fuller (1979). * and ** denote the significance at the 10% and 5% level, respectively.

Country	Mean	St. dev.	Skewness	Kurtosis	Correlation coefficient		
Canada	0.040	0.117	0.137	2.116	Canada	Euro area	Japan
Euro area	0.077	0.157	0.477	2.323	0.392		
Japan	−0.108	0.209	0.406	2.541	−0.011	0.755	
U.K.	0.252	0.123	0.875	4.430	0.144	0.737	0.536
	α_i	s.e.	β_i	s.e.	half-life	l_i	τ_i
Canada	0.000	0.001	0.983	0.008	4.8	16	−2.10
Euro area	0.002	0.002	0.966	0.013	2.5	13	−2.70*
Japan	−0.002	0.002	0.982	0.011	3.7	22	−1.69
U.K.	0.013	0.005	0.946	0.018	2.1	21	−3.01**
10% critical value							−2.57
5% critical value							−2.88

3. Data

To construct the real exchange rates we collect monthly consumer price index (CPI) and nominal exchange rate data for Canada, the Euro area, Japan, the U.K., and the U.S. for the period 1978:12–2003:12. Similar to other studies, CPI data and period-ending exchange rates against the U.S. dollar are obtained from International Financial Statistics. Because the euro–dollar rate is only available from January 1999, we use the “synthetic” euro from the ECB.⁶ To construct the CPI data for the Euro area we employ the geometric weighted average method as described in Maeso-Fernandez et al. (2001, p. 11). In this calculation, Ireland is discarded because its CPI is only available as of 1997 and Luxembourg is excluded because of its currency union with Belgium. The first 25 observations of our sample are used to compute the lagged exchange rate changes needed for the ADF tests. This implies that 276 time-series observations are left for the estimation of the model and we also use $T = 276$ in the Monte Carlo simulations. For our robustness analysis, we collect similar data over the same period for Austria, Belgium, Denmark, Finland, France, Germany, Greece, Italy, the Netherlands, Norway, Portugal, Spain, Sweden, and Switzerland as well as for Cyprus, Hungary, Malta, and Turkey.

Table 1 presents summary statistics for the real exchange rates as well as results of univariate ADF unit root tests. The correlations between the real exchange rates of the euro, pound, and yen against the dollar are substantial. This underlines the importance of accounting for cross-sectional dependence in panel tests of PPP. Moreover, the bottom panel of Table 1 shows that the serial correlation properties are not identical across the real exchange rate series. The optimal lag length in the univariate unit root tests, determined with the procedure of Campbell and Perron (1991) varies from 13 for the euro–dollar series to 22 for the yen–dollar rate which is in line with previous research (see, e.g., Papell and Theodoridis, 2001; Wu and Wu, 2001). Hence, it is important to allow for heterogeneous serial correlation.

Table 1 also reports the half-life of each of the PPP deviations, which we estimate using impulse response analysis (see, e.g., Imbs et al., 2005; Murray and Papell, 2002). The half-life is defined as the

⁶ This synthetic euro series is the “ECB reference exchange rate, U.S. dollar/euro, 2:15 pm (C.E.T.), against ECU up to December 1998,” which is available at the website of the European Central Bank (www.ecb.int).

number of periods needed to absorb half of a shock administered to the real exchange rate. We use linear interpolation to approximate the half-life due to the discreteness of the data. Similar to Cheung and Lai (2000), we find non-monotonic impulse response functions, which implies that the half-life may not always be uniquely defined. In such cases, we follow Kilian and Zha (2002) and define the half-life as the highest of the estimated values. The estimated half-lives of the PPP deviations range from about two years for the U.K. to almost five years for Canada. The unit root hypothesis can be rejected for the U.K. at the 5% level and for the Euro area at the 10% level.

Interestingly, the standard deviation of the Canadian dollar–U.S. dollar real exchange rates is the lowest of the four series (and unreported graphs confirm that this series shows less pronounced deviations from the long-term average than the other series), but the estimated half-life is the greatest for this real exchange rate. It could be that there are fewer fundamental shocks to the Canadian dollar–U.S. dollar rate, which results in a lower signal-to-noise ratio in the estimation of the mean reversion parameter. Alternatively, there may be non-linear effects in the adjustment of real exchange rates to shocks, with greater mean reversion for shocks that exceed a certain threshold level. We consider a full exploration of these issues to be an interesting avenue for further research.

4. Empirical analysis

This section investigates the properties of the three panel methodologies discussed in Section 2. We do not aim to offer an exhaustive treatment of the properties of unit root tests in heterogeneous panels. Rather, the goal is to illustrate how homogeneous estimation and testing of a panel can affect the inferences, in particular when examining the PPP hypothesis in a panel consisting of real exchange rates.

For all three methodologies, we estimate the parameters of model (1) and test the PPP hypothesis by comparing the appropriate unit root test statistics to their corresponding simulated critical values. Panel A of Table 2 displays the results of estimating and testing PPP with HoHo, imposing homogeneous mean reversion. We note that although the estimated β_i is the same for all countries, the half-lives are in general different from each other due to differences in the serial correlation structure. The half-life is around four years for Canada and around three years for the other countries. Further, the unit root hypothesis is rejected (p -value = 0.023). This finding suggests that the PPP tenet is a reasonable description of the panel as a whole. We note that the τ -test and the Wald (MADF) test are equivalent in case the unit root hypothesis is tested homogeneously. Panel B of Table 2 presents the results for the HeHo and HeHe methodologies, which both involve heterogeneous estimation of the β_i parameters. Again, the estimated half-lives of PPP deviations differ importantly across the real exchange rates, varying from around two years to almost five years. The Wald test rejects the unit root null only at the 10% level for the panel as a whole (HeHo), while the ADF τ -statistic (HeHe) is only significant for the euro–dollar rate (5% level).

Inferences regarding the validity of the PPP hypothesis in our panel thus depend on the chosen methodology. While HoHo yields strong rejection of the unit root hypothesis, HeHe indicates that only one of the four real exchange rates is stationary. HeHo shows an intermediate result and rejects the unit root at the 10% level. To obtain a thorough understanding of the consequences of methodology choice, we perform a Monte Carlo study on the properties of the three methodologies. When comparing the performance of alternative methodologies to detect unit roots in panels, the most interesting experiments concern DGP's that contain both stationary series and series with unit roots (i.e., mixed panels). Hence, we follow Taylor and Sarno (1998) and analyze the properties of the methodologies when, respectively, one, two, or three of the simulated series have a root less than unity. Subsequently, we study a variety of DGP's that contain only stationary exchange rates (i.e., heterogeneous panels). In the remainder of the paper, both types of panels are referred to as panels that exhibit heterogeneous mean reversion.

4.1. Mixed panels

In this section, we examine DGP's in which one or more of the real exchange rates in the panel contain a unit root under the alternative hypothesis. First, we consider DGP's with three non-stationary series. The mean reversion parameter (β_i) of the remaining real exchange rate is set equal to values from the set {0.99, 0.975, 0.95, 0.925, 0.9}. We then apply the HoHo, HeHo, and HeHe methodologies to

Table 2

Panel unit root tests. This table presents the results of the SUR ADF unit root test under HoHo (Panel A) and the HeHo and HeHe methodologies (Panel B) of the real exchange rates of Canada, the Euro area, Japan, and the U.K. versus the U.S. over the period 1981:01–2003:12. We estimate the following system:

$$R_{i,t} = \alpha_i + \beta_i R_{i,t-1} + \sum_{k=1}^{l_i} \gamma_{i,k} \Delta R_{i,t-k} + u_{i,t},$$

where $R_{i,t}$ is the log of the real exchange rate and the value of l_i is taken from the OLS ADF unit root test results. Under HoHo, $\beta_i = \beta$ for all i . The empirical p -values as well as the critical values for the Augmented Dickey–Fuller (ADF) τ -statistic and the Multivariate ADF (MADF) Wald test statistic to test for a unit root are obtained using Monte Carlo Simulations. * and ** denote significance at the 10% and 5% level, respectively.

Country	α_i	s.e.	β_i	s.e.	Half-life	l_i	τ_i [p-value]	Wald [p-value]
<i>Panel A: Homogeneously estimated mean reversion coefficient (HoHo)</i>								
Canada	0.000	0.001	0.977	0.006	4.0	16	−4.23** [0.023]	18.34** [0.023]
Euro area	0.002	0.002			3.0	13		
Japan	−0.003	0.002			2.8	22		
U.K.	0.006	0.002			3.2	21		
10% critical value							−3.59	13.33
5% critical value							−3.95	15.99
<i>Panel B: Heterogeneously estimated mean reversion coefficient (HeHe and HeHo)</i>								
Canada	0.000	0.001	0.983	0.008	4.8	16	−2.13 [0.275]	21.69* [0.064]
Euro area	0.003	0.002	0.964	0.010	2.2	13	−3.72** [0.021]	
Japan	−0.003	0.002	0.977	0.010	2.8	22	−2.39 [0.231]	
U.K.	0.010	0.004	0.961	0.014	2.2	21	−2.90 [0.149]	
10% critical value			Canada				−2.69	20.25
			Euro area				−3.15	
			Japan				−2.80	
			U.S.				−3.10	
5% critical value			Canada				−2.98	22.39
			Euro area				−3.41	
			Japan				−3.11	
			U.S.				−3.41	

estimate the parameters in equation (1) and compute the test statistics. Comparing these statistics with the critical values yields a power function for each of the methodologies. We repeat this procedure for all four exchange rates. Fig. 1 displays the power functions of the HoHo and HeHo tests for the stationarity of the panel as a whole as well as the HeHe tests for the stationarity of the four exchange rate series individually. Chart titles indicate which of the simulated real exchange rate series is stationary. The first thing that leaps to the eye is the remarkable behavior of the power function of the unit root test of HoHo. The power of this methodology to reject the unit root null is very low and often even lower than the size of the test (5%). Furthermore, if the β of the mean reverting exchange rate decreases from 0.99 to 0.9, the power is not monotonically increasing. Especially for Canada, the power decreases when the β moves further away from the unit root null. This is remarkable, as the power of the test should increase when the panel as a whole becomes more stationary.

The behavior of the four individual unit root tests in HeHe is essentially as expected. The power to reject the unit root null for the non-stationary series is very low and roughly equal to the 5% probability of a type I error. The power function of the HeHe test for the stationary series is the highest of all methodologies and clearly increases when β moves further away from unity. The power to reject the null depends on the particular real exchange rate that is stationary. For example, when $\beta = 0.95$, the power is 0.83 for the real exchange rate of the Euro area, while it is only 0.56 for the exchange rate of Japan. Hence, the influence of the country or currency of study on the power of the test seems to be substantial. We

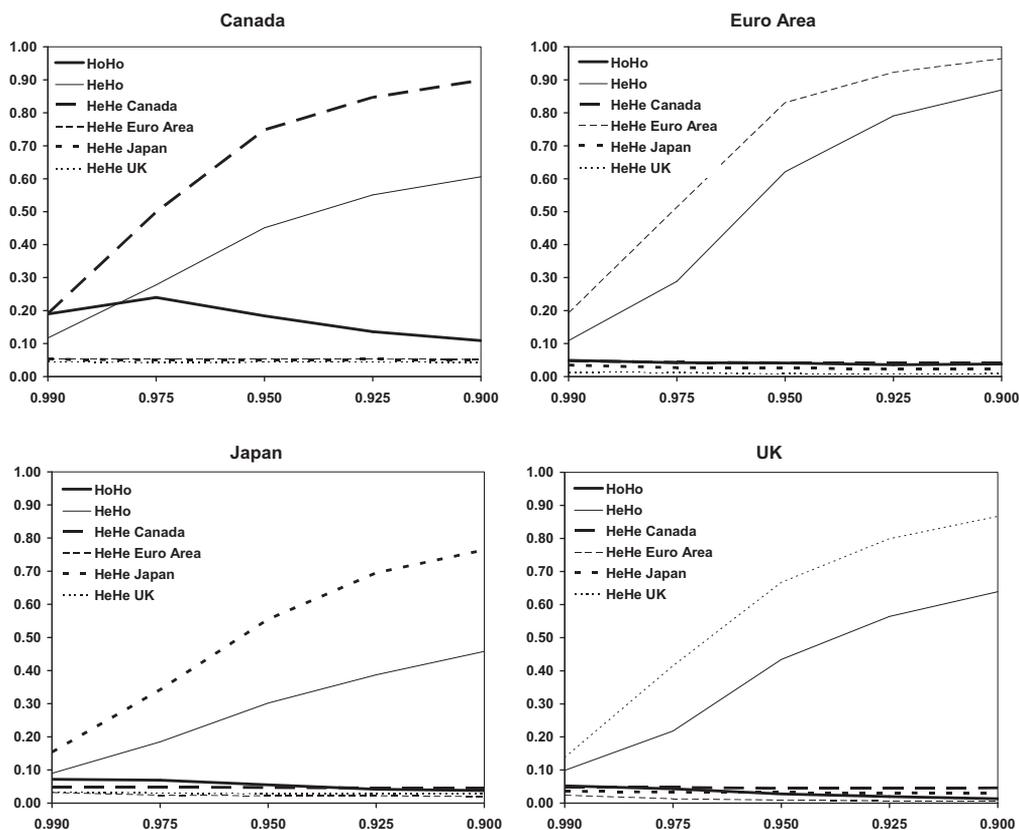


Fig. 1. Estimated power functions of the panel methodologies (mixed panel, 3 unit roots). This figure presents the estimated power functions for the panel methodologies in which (i) the mean reversion parameter is estimated homogeneously and the unit root test is performed for the entire sample of real exchange rates (HoHo methodology), (ii) the mean reversion parameters are estimated heterogeneously, but the unit root test is performed for the entire sample of real exchange rates (HeHo methodology), and (iii) the mean reversion parameters are estimated heterogeneously and the unit root test are performed for each individual real exchange rate separately (HeHe methodology). The power functions are computed with Monte Carlo simulations (see Appendix A for a detailed description) based on the real exchange rates of Canada, the Euro area, Japan, and the U.K. versus the U.S. over the period 1981:01–2003:12. The power functions are estimated under the restriction that three of the four mean reversion parameters are equal to one, while the remaining mean reversion parameter decreases from 0.99 to 0.9. Chart titles indicate the stationary real exchange rate series.

note that although the power for the HeHe test is high relative to the other methodologies, even at low values of β there is a considerable probability of failing to reject the true null-hypothesis.

The power of HeHo increases when β decreases, but is considerable lower than the power of HeHe. Although this is a natural implication of the fact that the HeHe test is an individual unit root test while HeHo evaluates the stationarity of the panel as a whole, it also underlines the role the chosen methodology can play in assessments of the PPP hypothesis.

Fig. 2 shows the power of the unit root tests for the three different methodologies when the number of exchange rates in the panel that are non-stationary is equal to two or one. The mean reversion parameters of the stationary series decrease simultaneously from 0.99 to 0.9. (To conserve space, we do not present the charts for all possible panels with one or two unit roots. The remaining charts are similar and available from the authors.) We observe that although the power of HoHo improves slightly with fewer unit roots in the sample, the power function again does not monotonically increase when β is reduced from 0.99 to 0.9. For example, when the real exchange rates of Canada and Japan are stationary, the power of the panel unit root test is twice as high when the β 's of Canada and Japan are

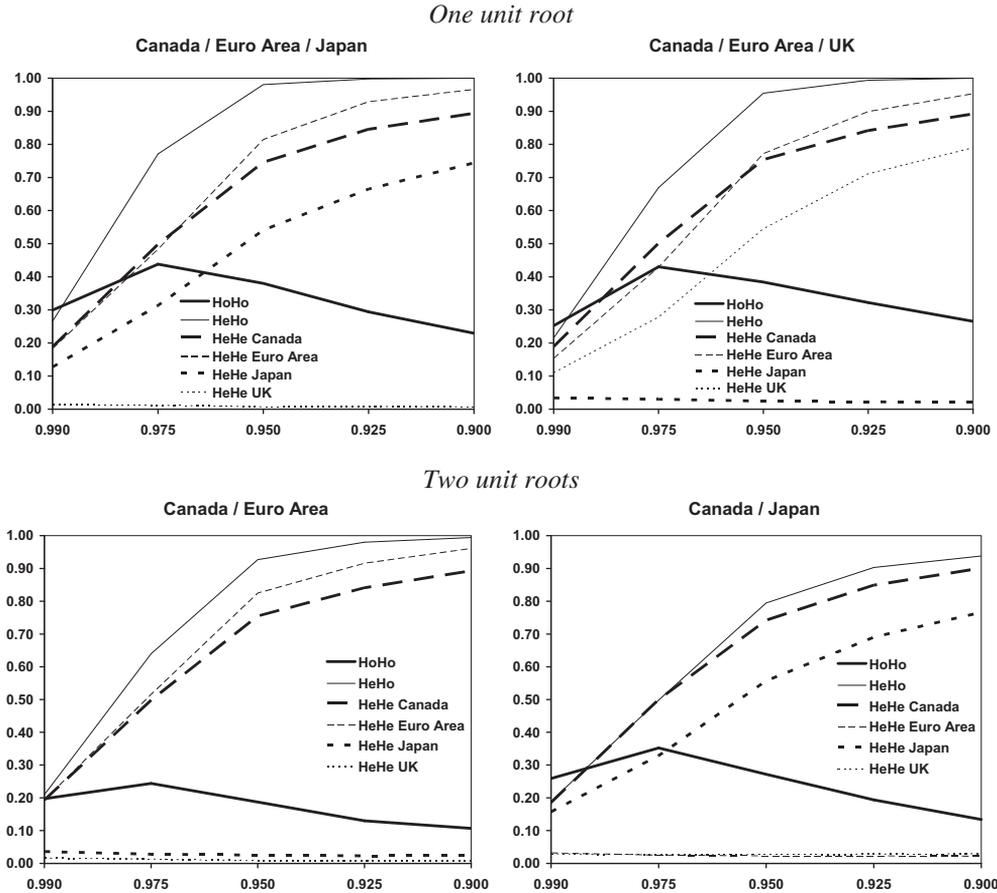


Fig. 2. Estimated power functions of the panel methodologies (mixed panel, 1 or 2 unit roots). This figure presents the estimated power functions for the HoHo, HeHo, and HeHe methodologies (described in Fig. 1). The functions are estimated under the restriction that one or two of the four mean reversion parameters are equal to one, while the remaining parameters decrease from 0.99 to 0.9. Out of the four (six) possible panels of real exchange rates with one (two) unit root(s), we present two (two) examples in this figure. The other power functions show similar shapes and are available from the authors. Chart titles indicate the stationary real exchange rate series.

equal to 0.975 compared to the case where both β 's equal 0.9. The irregular properties of HoHo seem to become more pronounced when the number of unit roots in the panel is reduced.⁷

The power functions of HeHe reveal that the power does not crucially depend on the number of non-stationary series in the panel, but the power does still vary considerably across the currencies. Again, the power of this test to reject the null is very close to the size of the test for the non-stationary series. The power of HeHo increases when the number of unit roots in the panel decreases, because the panel as a whole becomes more stationary. When comparing the power properties of HeHe and HeHo, we note that the power is similar for DGP's in which two out of the four exchange rates are stationary.

⁷ We note that the degree of heterogeneity in the panels analyzed in Fig. 2 is actually relatively limited, as the mean reversion parameters are identical for two or three out of the four series. The drawbacks of HoHo could be aggravated by introducing more heterogeneity in the panel and hence we bias the result against finding important differences between the alternative methodologies.

When only one of the exchange rates in the panel contains a unit root, the power of HeHo is generally higher.

Figs. 1 and 2 show that HeHo is generally most powerful when there are three or more stationary series in the panel. HeHe succeeds relatively well in identifying individual stationary series. The power properties of HoHo can be described as undesirable, as the power is low and often not monotonically increasing when the mean reversion parameters decrease. To gain more understanding of these adverse properties and thus of the consequences of imposing homogeneous mean reversion, we plot the distribution of estimated β coefficients under HoHo.

Fig. 3 displays the distribution of the (homogeneously) estimated mean reversion parameter β under HoHo when the panel contains two unit roots. (Again, we obtain similar results with other panels with two unit roots or with panels with one or three unit roots.) The mean reversion parameters of the stationary exchange rates in the panel are assumed to decrease simultaneously from 0.99 to 0.5.⁸ We add $\beta = 0.5$ to the set of {0.99, 0.975, 0.95, 0.925, 0.9} to highlight the impact of the homogeneity restriction in extreme cases. The histograms also present the values of μ , defined as the unweighted average of the true mean reversion parameters (used as a proxy for the “true” homogeneous mean reversion parameters, see also footnote 19), and b , the (homogeneously) estimated mean reversion parameter.

The results are striking. We observe that when the mean reversion parameter of the stationary series in the panel decreases, the distribution of the estimated β of the total panel first gradually moves to the left. This can be expected, as the panel as a whole becomes more stationary. However, when the mean reversion parameter reaches a value of 0.95 or less, there is a clear movement to the right, implying higher estimates of the homogeneous β . As the value of b slowly increases, while the mean of the true mean reversion parameters (μ) steadily drops, this implies an increasingly large bias in the homogeneous parameter estimate. Second, marked changes to the shape of the distribution occur. Moving from the top graphs to the bottom, the dispersion and skewness of the distribution initially increase notably, signifying increasing uncertainty about the parameter estimate. However, when the stationary series in the panel become even more mean reverting, the distribution again becomes more dense and less skewed. Unreported results show that this “boomerang” effect is more severe when the panel contains fewer series with unit roots. Hence, the more the panel as a whole moves further away from the unit root null-hypothesis, the more the distribution of the estimated β starts to resemble the one under the null-hypothesis. Remarkably, the homogeneous estimation methodology seems to “ignore” mean reversion parameters that deviate considerably from the non-stationary null and the unit roots in the panel dominate the stationary series in the estimation.

The extraordinary character of this feature of HoHo is best illustrated by the bottom panel of Fig. 3. This panel concerns a DGP in which only the euro–dollar rate contains a unit root, and all other three real exchange rates have a mean reversion coefficient equal to 0.5. The mean of the estimated mean reversion parameters (denoted by b) under HoHo is 0.989, implying a bias of no less than 0.364 relative to the average of the true β 's in the panel under the null-hypothesis ($\mu = 0.625$). These findings reinforce the conclusion that imposing a common mean reversion coefficient in mixed panels can lead to testing methodologies with seriously adverse properties, including low and non-monotonic statistical power and severe biases in the parameter estimates.

4.2. Heterogeneous panels

In this section, we investigate the power functions of the HoHo, HeHo, and HeHe methodologies for DGP's in which all real exchange rates are stationary. As a benchmark, we first investigate the situation in which all four exchange rates exhibit the same mean-reverting behavior. That is, all mean reversion parameters β_i are equal to the same value $\beta \in \{0.99, 0.975, 0.95, 0.925, 0.9\}$. (We obtain similar results when we set the β_i 's equal to the estimated values of the heterogeneous SUR model in Table 2 and decrease the mean reversion parameter of one of the series at the time.) The results are in Fig. 4. Not

⁸ The distributions are based on the same Monte Carlo simulations as for the analysis of the statistical power of the tests, but 10,000 instead of 1000 simulations are used to obtain a smoother representation of the distributions.

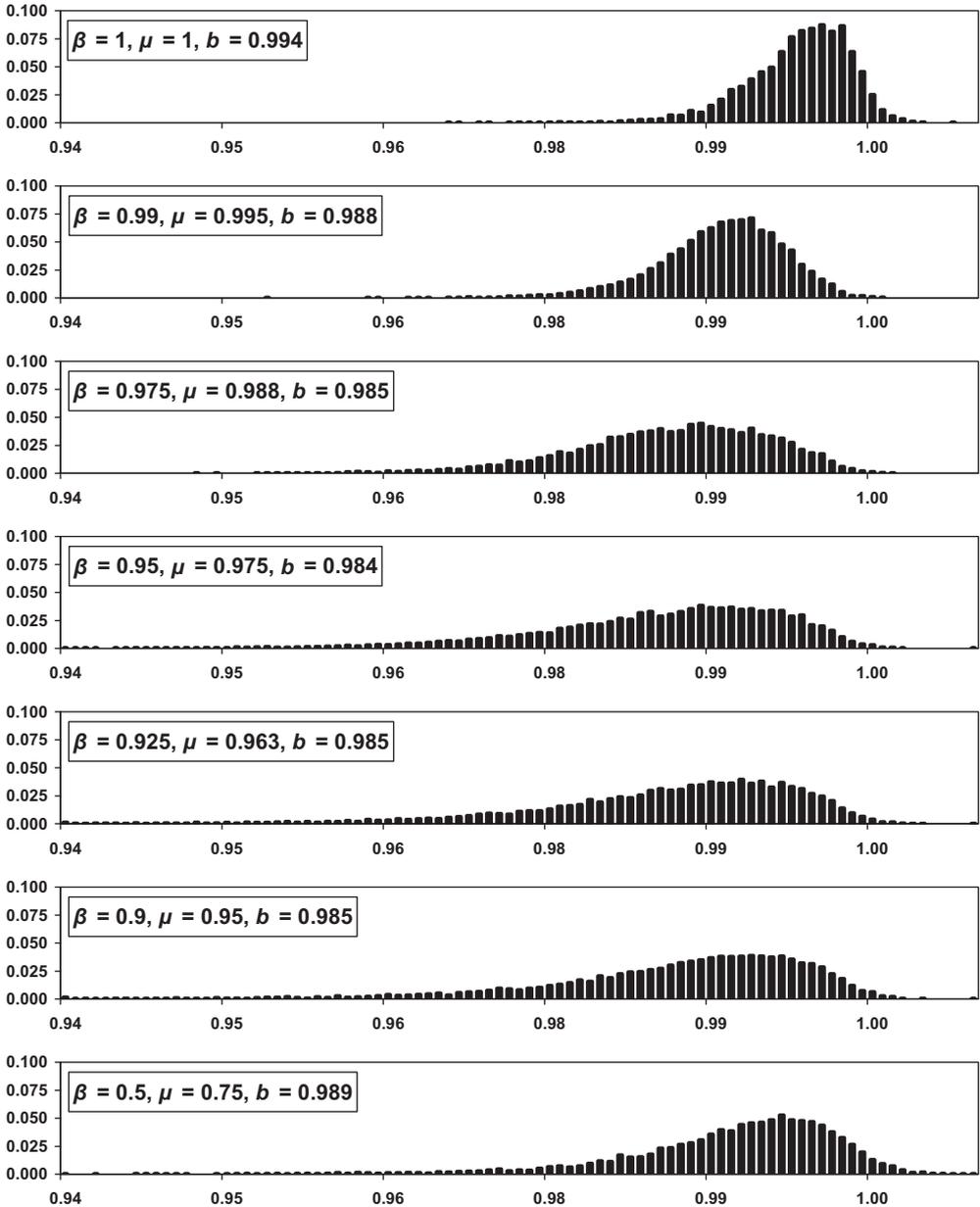


Fig. 3. Histograms of homogeneously estimated mean reversion coefficient (mixed panel, 2 unit roots). This figure presents histograms of the estimates of the common mean reversion coefficient under HoHo when the true mean reversion parameters for the set of countries {Canada, Euro area, Japan, U.K.} equal $\{\beta, 1, \beta, 1\}$, where β decreases from 1 to 0.5. In the histograms, μ represents the average of the true mean reversion parameters, while b denotes the mean of the (homogeneously) estimated mean reversion parameter.

surprisingly, HoHo performs very well. Its power is the highest of all methodologies and it monotonically increases with a decrease in the mean reversion parameter. This can be explained by the fact that the homogeneous model is not mis-specified for this DGP. HeHo is only slightly less powerful than HoHo, however. For β 's of 0.95 and lower, both testing methodologies are able to reject the unit root in

virtually all cases. When $\beta = 0.975$, the probability that non-stationarity is rejected still amounts to 92% (HoHo) and 79% (HeHo). The power functions of HeHe behave similarly to those generated under DGP's with both stationary and non-stationary series.

Fig. 5 displays the power functions of the HoHo, HeHo, and HeHe methodologies for DGP's with a decreasing value of β_i for one of the exchange rates, while we assign the values {0.99, 0.975, 0.95} to the remaining mean reversion parameters (these three values are assigned to the remaining real exchange rate series in such a way as to maximize the resemblance with the estimated values in Table 2). The plots demonstrate how sensitive the behavior of the HoHo test is to heterogeneity in the panel. The heterogeneity in the betas in the DGP's of Fig. 5 is limited but the power function of HoHo is considerably lower and is clearly non-monotonous.

We take a closer look at the empirical performance of the HoHo model by plotting the distribution of the estimated β coefficients for the DGP's analyzed in Fig. 5. (To save space, Fig. 6 only shows the distributions for the case in which the mean reversion of the Canadian dollar–U.S. dollar rate is varied. The distributions for the other exchange rates display a similar pattern and are available from the authors.) The “boomerang” effect documented in the previous section is evidently not contingent on the presence of unit roots in the panel. Again, the distribution of the estimated common mean reversion coefficient at first shifts to the left and becomes more dispersed. For lower half-lives of the Canadian dollar–U.S. dollar real exchange rate, however, the distribution moves back to the right. Even without a unit root in the panel, the least stationary series appear to dominate stationary series with a root that deviates from unity. Consequently, HoHo can also lead to serious biases in the estimated mean reversion parameters in panels with only stationary real exchange rates.

To illustrate the differences between homogeneous and heterogeneous estimation methodologies, we also plot the distribution of the estimated β 's under HeHe. Fig. 7 presents histograms for the heterogeneously estimated β 's for the same DGP as in the top-left panel of Fig. 5 and as in Fig. 6. Again, μ denotes the average of the true mean reversion parameters, and b represents the heterogeneously estimated mean reversion parameter for the specific series under examination. We note that the scale of the x-axis for Canada is different for different values of β . As the mean reversion parameter of the exchange rate of Canada decreases, the distribution of the estimated β for Canada shifts to the left, while the distributions for the other countries virtually stay the same. For Canada, the estimate b moves in lockstep with the decreasing true mean reversion parameter β . (Unreported results show a similar

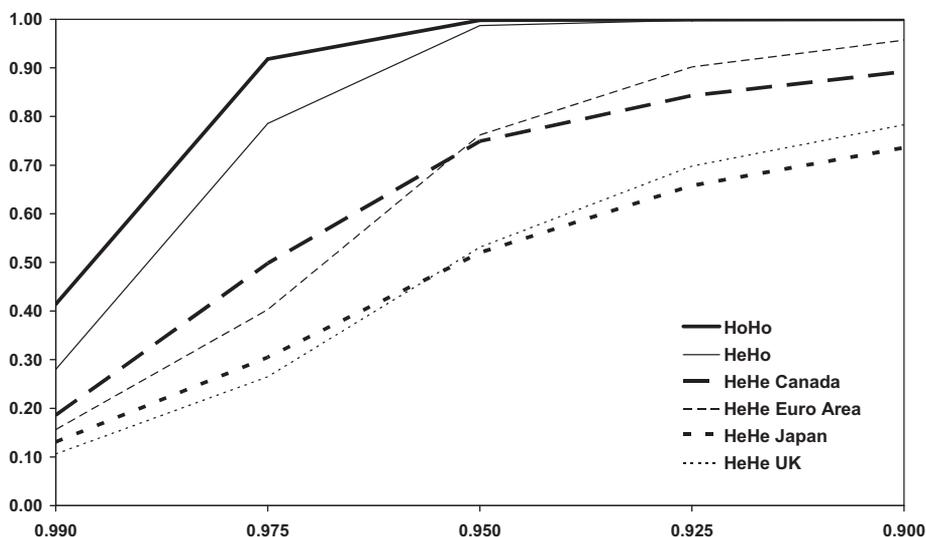


Fig. 4. Estimated power functions of the panel methodologies when mean reversion is homogeneous. This figure presents the estimated power functions for the HoHo, HeHo, and HeHe methodologies (described in Fig. 1). The functions are estimated under the restriction that all four mean reversion parameters are equal to β , where β decreases from 0.99 to 0.9.

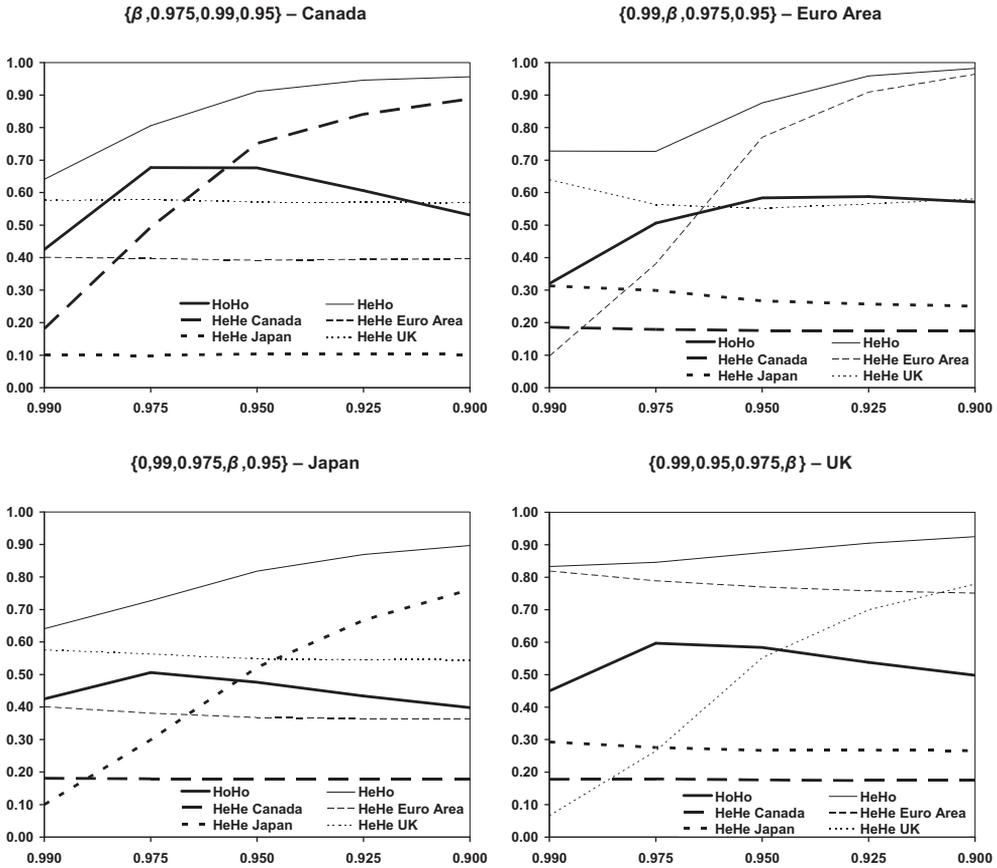


Fig. 5. Estimated power functions of the panel methodologies when mean reversion is heterogeneous (heterogeneous panel). This figure presents the estimated power functions for the HoHo, HeHo, and HeHe methodologies (described in Fig. 1). The functions are estimated under the restriction that three of the four mean reversion parameters are fixed at a level smaller than one (chosen from the set {0.99, 0.975, 0.95}), while the remaining parameter decreases from 0.99 to 0.90. Chart titles indicate the stationary real exchange rate series.

pattern when the β of one of the other countries is decreased.) For the Euro area, Japan, and the U.K. (the histograms for the latter two countries are not included in Fig. 7, but they look similar to those for the Euro area), the true mean reversion parameters are equal to {0.975, 0.99, 0.95} respectively, while – independent of the decrease in Canada’s β – the estimated mean reversion parameter b equals {0.965, 0.98, 0.934}. This indicates that HeHe leads to a relatively accurate estimate of individual mean reversion parameters in heterogeneous panels, even in small samples.

4.3. Interpreting the empirical results

The results in Sections 4.1 and 4.2 reveal two major differences between the methodologies that employ homogeneous and heterogeneous estimation of the mean reversion parameter. First, the estimated homogeneous mean reversion parameter is generally substantially larger than the unweighted average of β_i , the mean reversion parameters of the individual series in the panel. In some circumstances, this bias amounts to over 0.3 when one of the four mean reversion parameters in the panel is equal to 1. Second, we report a remarkable increase in the homogeneous mean reversion estimate when one or more of the series in the panel become more stationary. This section examines two possible explanations for these biases.

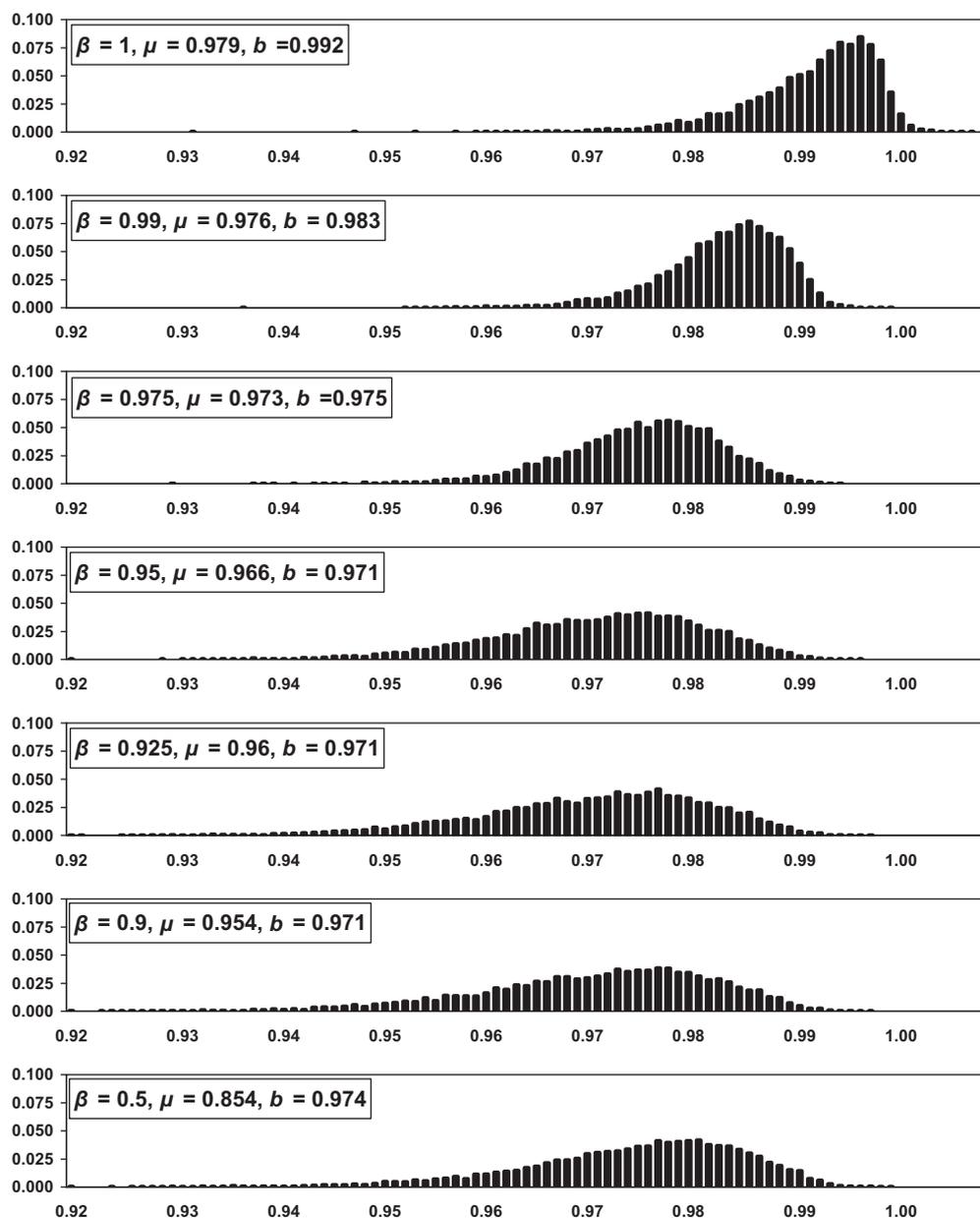


Fig. 6. Histograms of homogeneously estimated mean reversion coefficient (heterogeneous panel). This figure presents histograms of the estimates of the common mean reversion coefficient under HoHo when the true mean reversion parameters for the set of countries {Canada, Euro area, Japan, U.K.} equal $\{\beta, 0.975, 0.99, 0.95\}$, where β decreases from 1 to 0.5. In the histograms, μ represents the average of the true mean reversion parameters, while b denotes the mean of the (homogeneously) estimated mean reversion parameter.

First, non-stationary real exchange rates can dominate less persistent series in panel estimation methodologies. This implies that there is an upward bias in the estimated homogeneous mean reversion parameter (when compared to the unweighted average). In fact, it can be shown that in mixed panels, the estimate of β is asymptotically equal to one (proven in [Appendix B](#), which is available

Euro area

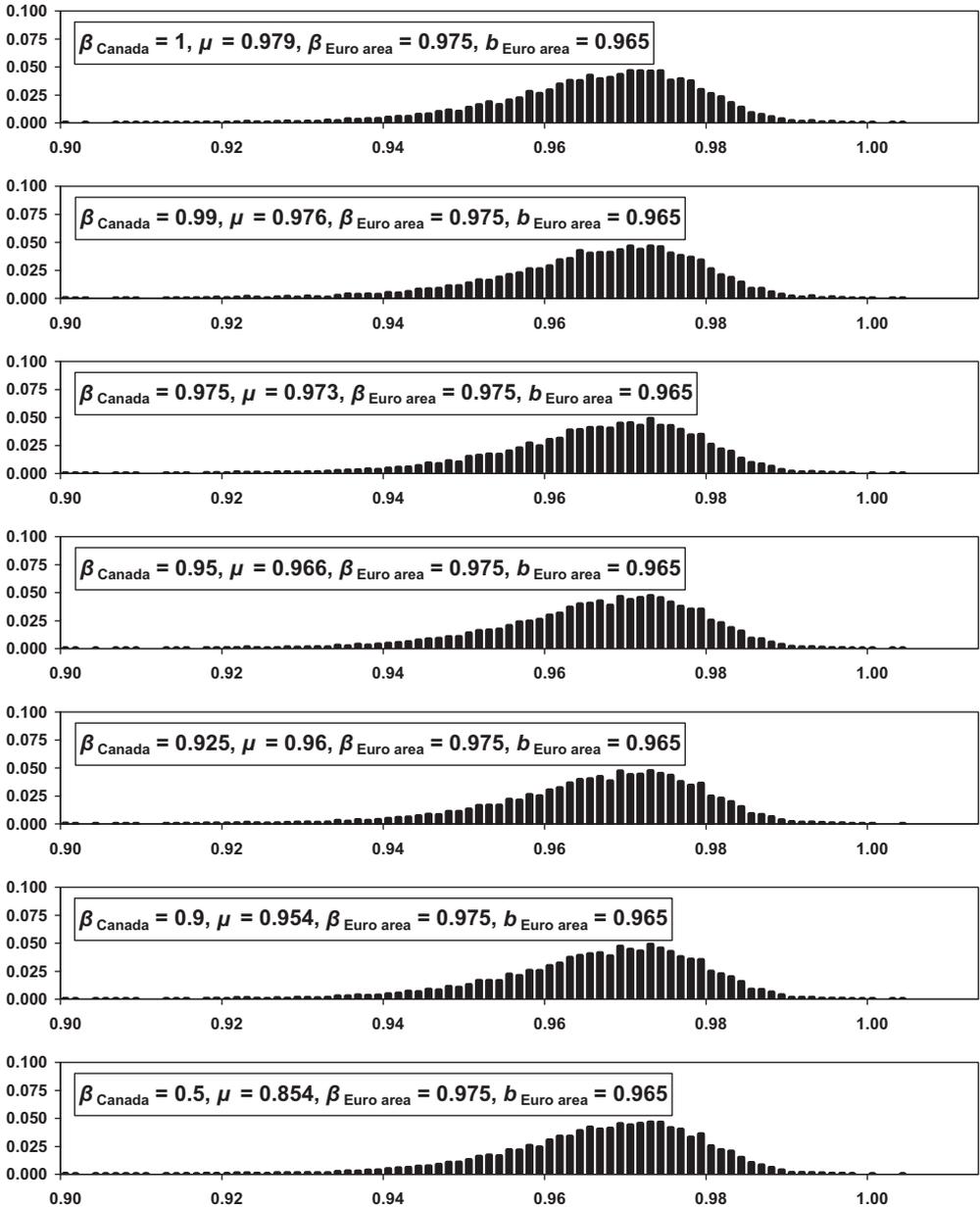


Fig. 7. Histograms of heterogeneously estimated mean reversion coefficient (heterogeneous panel). This figure presents histograms of the estimates of the heterogeneous mean reversion coefficients under the HeHe methodology when the true mean reversion parameters for the set of countries {Canada, Euro area, Japan, U.K.} equal $\{\beta, 0.975, 0.99, 0.95\}$, where β decreases from 1 to 0.5. In the histograms, μ is the average of the true mean reversion parameters, while b denotes the mean of the (heterogeneously) estimated mean reversion parameter. We note that the scale of the x-axis for Canada is different for the different values of β_{Canada} . Canada. Note: The histograms for Japan and the U.K. are similar and are available from the authors.

Euro area

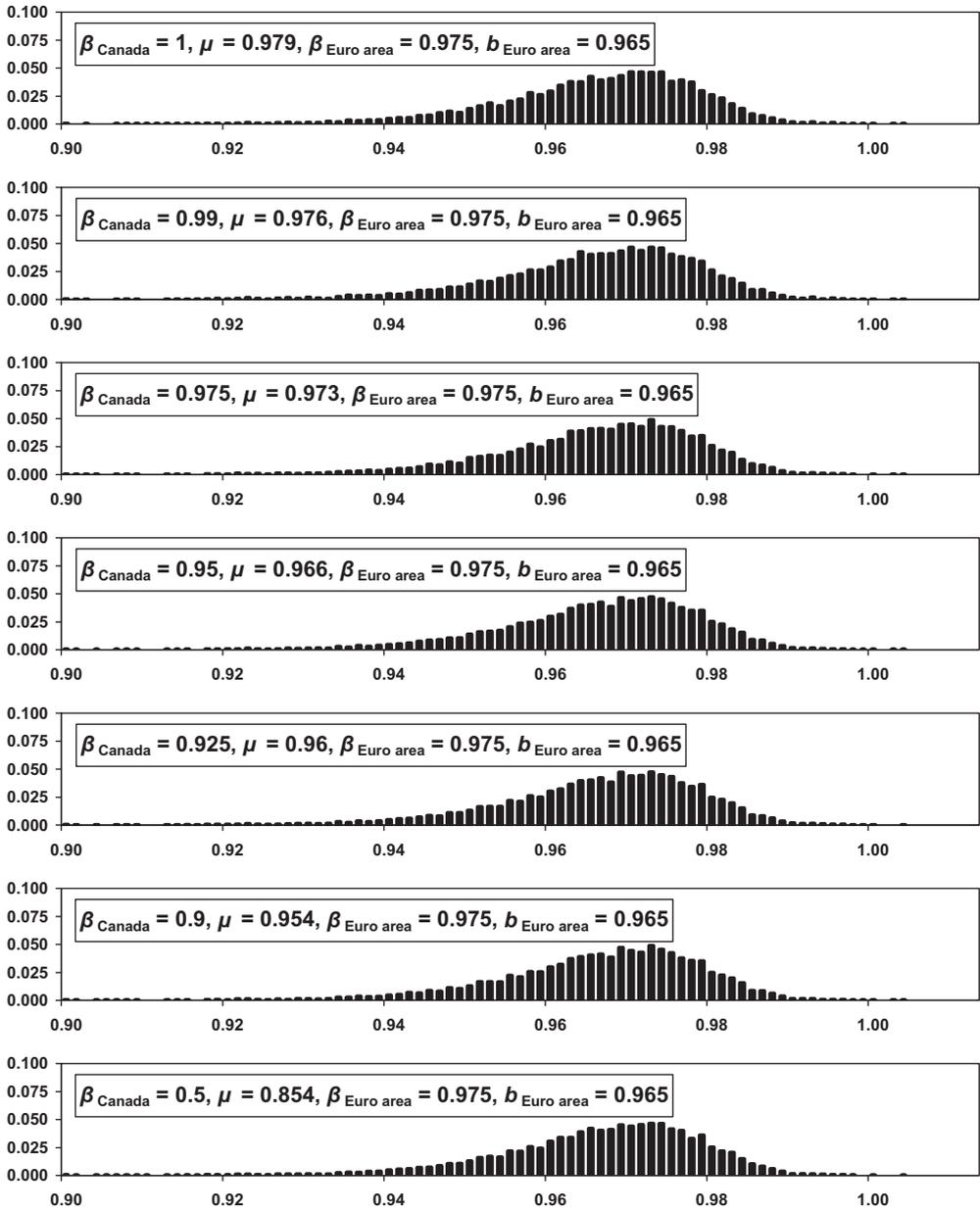


Fig. 7. (Continued).

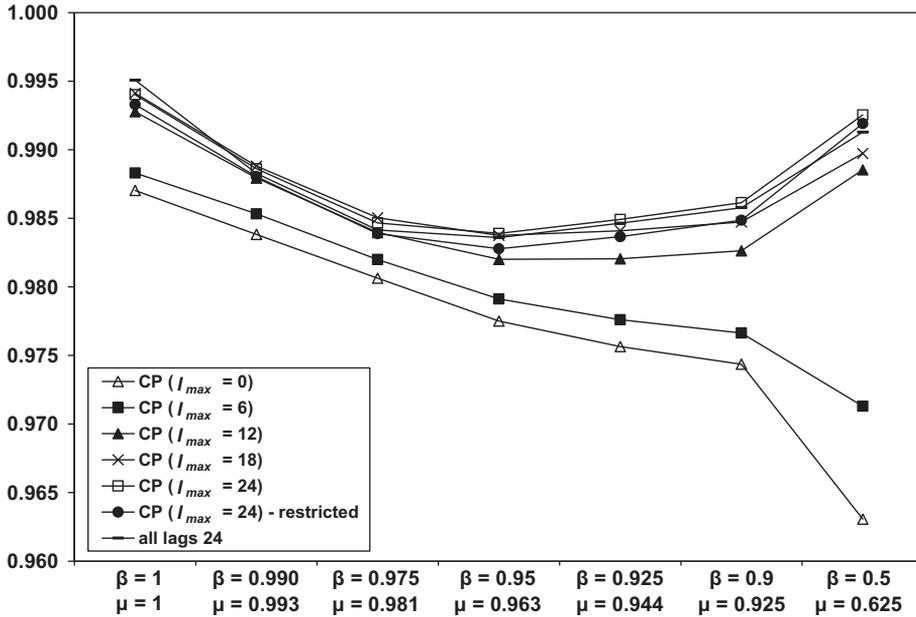


Fig. 8. Lag length and homogeneously estimated mean reversion coefficients (mixed panel, 1 unit root). This figure shows the strength of the “boomerang” effect for different variations of the lag length selection approach most commonly employed in the literature. The vertical axis depicts the value of the estimated homogeneous mean reversion coefficient. Each point in the graph represents the mean (across 10,000 simulations) of the estimated common mean reversion coefficient under HoHo when the true mean reversion parameters for the set of countries {Canada, Euro area, Japan, U.K.} equal $\{\beta, 1, \beta, \beta\}$, where β decreases from 1 to 0.5 (horizontal axis). In the figure, μ is the average of the true mean reversion parameters. The points that are generated by the same lag length selection procedure have the same symbol and are connected by lines. We consider the following different procedures to determine the lag length l_i : (i) the recursive t -statistic procedure of Campbell and Perron (1991) (CP) for l_{max} equal to 0, 6, 12, 18, and 24, (ii) the same procedure but only including the lags at which the individual serial correlation coefficients is significantly different from 0 (labeled “restricted”), and (iii) the lag length choice where l_i is set to l_{max} for all i with $l_{max} = 24$.

from the authors). In finite samples, the bias created by the dominant unit root in the panel can be large, as illustrated by Fig. 3.⁹ Furthermore, it can be demonstrated that in heterogeneous panels, the homogeneous mean reversion parameter estimate is a weighted average of the mean reversion parameters of the individual series, with the weights depending on the variance of the series (we refer to Appendix B for the proof). As the variance is determined by the persistence of shocks to the real exchange rates, the most persistent series will have the largest weight. Choi et al. (2006) demonstrate that when the number of series in a mixed panel approaches infinity (and there is constant fraction of stationary series in the panel), the OLS estimate of the homogeneous mean reversion parameter is biased upward. We show that (i) in mixed panels with a finite number of series the estimated homogeneous mean reversion parameter approaches one when T goes to infinity and (ii) as the homogeneous mean reversion parameter is a weighted average of the individual mean reversion parameters in heterogeneous panels, the series with the highest variance (i.e., with a root closest to unity) plays a dominant role in the estimate.¹⁰

⁹ In fact, when we increase values of T used in the simulation from the sample length of 276 to merely 500, we find that the estimated homogeneous mean reversion parameter is already very close to 1 for all combinations $\{\beta, 1, \beta, 1\}$ where β decreases from 1 to 0.5.

¹⁰ Choi et al. (2006) conclude that cross-sectional heterogeneity in mean reversion rates are not an important source of bias in their sample of 21 countries over 1973–1998. The use of an annual data frequency in their analysis is a likely explanation for why the authors find substantially less heterogeneity than many other papers. For example, Wu and Wu (2001) examine a sample of 20 countries on the basis of quarterly data over 1973–1997 and find half-lives that range from 5 to 15 years.

Second, Pesaran and Smith (1995) suggest that the inconsistency of pooled estimators is due to the presence of serial correlation in the regressors. Fig. 8 underscores the role of serial correlation in the boomerang effect detected in this paper. The figure depicts the magnitude of the boomerang effect for several different procedures to estimate the serial correlation structure of the simulated real exchange rates. The points indicated with “□” (labeled “CP ($l_{\max} = 24$)”) represent the mean values of the β estimates under HoHo when the true mean reversion parameters for the set of countries {Canada, Euro area, Japan, U.K.} equal $\{\beta, 1, \beta, \beta\}$, where β decreases from 1 to 0.5. The vertical axis shows the estimate of the homogeneous mean reversion parameter, while the horizontal axis displays the value of the simulated mean reversion parameters of the stationary series in the panel. The boomerang effect can be recognized by the increase in the β estimate when moving further along the horizontal axis (i.e., the U-shape in the lines connecting the homogeneous β estimates). The other symbols in the figure reflect the boomerang effect for different lag length selection procedures. The following approaches are compared: (i) the recursive t -statistic procedure of Campbell and Perron (1991) with $l_{\max} = 0, 6, 12, 18,$ and 24 (see also footnote 15), (ii) the same procedure but only including the lags at which the individual serial correlation coefficients is significantly different from 0, and (iii) the lag length choice where l_i is set to l_{\max} for all i with $l_{\max} = 24$. The boomerang effect is clearly present for the recursive t -statistic procedure with $l_{\max} > 6$ and with approach (ii) as well as (iii). Only when the maximum lag length l_{\max} is limited, the mean of the homogeneous β estimates declines monotonically, and while the estimate is still significantly larger than the unweighted average of the β_i 's (denoted with μ in the figure), there is no evidence of a boomerang effect. These patterns call attention to the potentially important role of the serial correlation structure in panels of real exchange rates. While Papell and Theodoridis (2001) and Wu and Wu (2001) stress the need for allowing for heterogeneity in the serial correlation structure across real exchange rates, remarkably little is known about the influence of the lag length. We consider a comprehensive analysis of the impact of the lag length on the estimation results of panel approaches to test the PPP hypothesis (including the examination of alternative lag length selection procedures, see, e.g., Ng and Perron, 2001) to be an interesting topic for future research.

4.4. Robustness analyses

Since empirical studies on PPP use a wide variety of panels of real exchange rates, it is important to establish that our findings do not critically depend on the panel of the real exchange rates of Canada, the Euro area, Japan, the U.K., and the U.S. that we have chosen to analyze.¹¹ Therefore, we investigate the robustness of our results to changes in the number and the composition of real exchange rates in the panel. Many earlier papers on PPP use panels consisting of around 20 (industrialized) countries (see, e.g., Choi, 2004; Papell, 1997; Papell and Theodoridis, 2001; Wu and Wu, 2001). We follow Papell (1997) and construct a panel consisting of the real exchange rates of Austria, Belgium, Canada, Denmark, Finland, France, Germany, Greece, Italy, Japan, the Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, and the U.K. against the U.S. dollar for the period 1978:12–2003:12. We let the panel contain two unit roots (the real exchange rates of Austria and Belgium), while the mean reversion parameter of the other series again decrease simultaneously from 0.99 to 0.5. Unreported results show that although only two out of the seventeen series are non-stationary, a clear boomerang effect emerges – consistent with our results in Sections 4.1–4.3. We also find nearly identical results if we restrict our sample period to the pre-euro period 1978:12–1998:12. Finally, we consider two smaller panels that contain more European real exchange rates and obtain similar results (unreported). First, we consider a panel with the same composition as in Taylor and Sarno (1998), who investigate the real exchange rates of France, Germany, Japan, and the U.K. against the U.S. dollar. Again, the results are consistent with those reported in Sections 4.1–4.3. Second, to confirm that our results are not driven by the choice of the real exchange rates of major industrialized countries but are an intrinsic attribute of panel unit root tests of PPP in many settings, we rerun all our analyses for a panel which consists of the real exchange rates of Cyprus, Hungary, Malta, and Turkey (with the Maltese lira as the numeraire currency) over the same time period

¹¹ Breuer et al. (2002) demonstrate that in SUR regressions that allow for individual inferences on each member of the panel the power increases as the cross-correlations of the residuals increase.

and we find very similar results. We conclude that neither the number nor the composition of real exchange rates in the panel has a material impact on our main findings.

4.5. Evaluation of the three methodologies

Economically, there is little reason to believe that the speed of mean reversion is homogeneous across different real exchange rates.¹² The evidence presented in Sections 4.1–4.3 indicates that the restriction of homogeneous mean reversion can also have detrimental econometric consequences. Homogeneous panel unit root tests give rise to irregularly shaped power functions and potentially large biases in the parameter estimates in both mixed and heterogeneous panels, even when the degree of heterogeneity in the panel is relatively limited. In our view, these are compelling reasons for using an econometric methodology that allows for heterogeneous *estimation* of the mean reversion parameters.

The issue whether researchers interested in the PPP hypothesis should also perform unit root tests heterogeneously depends on the purpose of the study. If the hypothesis of interest is primarily whether one or more of the real exchange rates in the panel are stationary, the researcher's main objective is probably to achieve maximum power. The relative power of the HeHo and HeHe methodologies is dependent on the number of unit roots present in the panel. As this number is unknown, it may be valuable to use both methodologies.

An important drawback of homogeneous tests of PPP is that rejecting the unit root hypothesis does not provide any guidance as to how many real exchange rates are stationary, let alone which.¹³ Furthermore, from many perspectives the relevant question is not whether a panel of real exchange rates as a whole can be considered to be stationary, but whether PPP holds between individual countries. Research that is directed at this question should use a methodology that evaluates the PPP hypothesis for individual currency pairs in the sample. An important additional advantage of such a testing methodology is that the power is independent of the number of non-stationary series in the panel.

Our study is related to recent work by Imbs et al. (2005). They demonstrate that heterogeneity in the dynamics of sectoral price indices may induce significant biases in estimates of mean reversion parameters based on aggregated price indices.¹⁴ Although these authors also point at the importance of accounting for heterogeneity in the rate of mean reversion in studies on PPP, the biases they address are of an altogether different nature than the ones we examine in this paper. While they argue that it is problematic to aggregate price indices across sectors, our analysis shows that when aggregate real exchange rates are used for testing the PPP hypothesis, it is important to recognize that mean reversion properties may differ across countries.

5. Conclusions

Froot and Rogoff (1995) show that a researcher interested in testing the PPP hypothesis needs 72 years of stationary (monthly) data to reject the null-hypothesis of a random walk when the real exchange rate has a half-life of three years. Since the early 1990s, many researchers have turned to panel unit root tests to overcome this power problem. However, inferences drawn from panel testing methodologies are adversely affected by a number of restrictions imposed on the panel model. Two of these restrictions have been challenged by O'Connell (1998), who emphasizes the importance of allowing for cross-sectional dependence, and by Papell and Theodoridis (2001) and Wu and Wu (2001), who show that it is vital to take heterogeneous serial correlation into account. However, a third critical restriction in panel tests for PPP is still widely used: numerous studies impose a homogeneous rate of mean reversion across all real exchange rates in the panel. The economic rationale for this restriction is debatable, as the validity

¹² A number of papers – e.g., Campa and Wolf (1997), Cheung et al. (2001), and Cheung and Lai (2000) – identify factors that can account for cross-sectional differences in real exchange rate persistence, most notably geographical proximity, market size, inflation, government spending, and product market competition.

¹³ This pitfall is also discussed by Taylor and Sarno (1998). They suggest a test statistic that only rejects the null-hypothesis if all real exchange rates are stationary, but this test neither reveals how many nor which series are stationary when the null is not rejected.

¹⁴ Chen and Engel (2005) dispute that heterogeneity across sectors helps to solve the PPP puzzle.

of PPP between two countries should depend on various economic, institutional, and geographic characteristics specific to those countries. Econometrically, imposing this restriction may affect the test outcomes in important ways that have not been systematically investigated in the literature.

This paper studies the properties of three different panel estimation and testing methodologies. We examine the consequences of the homogeneity restriction on tests for PPP by means of Monte Carlo simulation. The first methodology involves homogeneous estimation of the mean reversion parameters and performing a unit root test on the panel as a whole. In the second, estimation is performed heterogeneously but testing is done homogeneously. In the third, the mean reversion parameters are estimated heterogeneously and the null-hypothesis of non-stationarity is tested for each individual exchange rate in the sample individually.

Our analysis uncovers important adverse properties of the methodology with homogeneous estimation and testing in the presence of heterogeneity in the data generating process (DGP). The estimated homogeneous mean reversion parameter is generally substantially larger than the average of the individual mean reversion parameters in the DGP. Hence, the homogeneity restriction leads to potentially highly misleading estimates of the speed of mean reversion. In other words, homogeneous estimation can lead to the conclusion that the adjustment of a panel of real exchange rates to shocks is considerably slower than the individual real exchange rates actually exhibit. Moreover, we document that in many cases the power function is non-monotonic, a phenomenon we refer to as the “boomerang” effect. This effect implies that the problem of underestimating the speed of adjustment to real exchange rate shocks is even greater for panels that include real exchange rates that are more stationary. We put forward two explanations for these properties. First, the most persistent real exchange rate in the panel dominates the estimation of the homogeneous mean reversion parameter. Second, the boomerang effect can be attributed to the serial correlation structure of the real exchange rate series.

These properties are observed for both mixed and heterogeneous panels and arise even when the heterogeneity across the real exchange rates in the panel is limited. The power functions of the other two methodologies behave in a regular way. Heterogeneous estimation and homogeneous testing generally leads to substantial power, while the third, fully heterogeneous, methodology performs well in detecting stationarity for individual series, especially when the number of unit roots in the panel is high. These findings highlight the importance of taking heterogeneous mean reversion into account in panel tests of PPP.

Acknowledgements

We are grateful to Jim Lothian (the editor), an anonymous referee, Menzie Chinn, Chi-Young Choi, Bruce Hansen, Christiaan Heij, Jean Imbs, Philippe Jorion, Masao Ogaki, Richard Paap, Ron Smith, Mark Taylor, Marno Verbeek, Ken West, and seminar participants at Fordham University, Ohio State University, the University of Wisconsin–Madison, and the Econometric Institute at Erasmus University Rotterdam for helpful comments and discussion.

Appendix A: Estimation and testing procedures and Monte Carlo simulations

A.1. Estimation and testing procedures

All three methodologies rely upon Seemingly Unrelated Regression (SUR) estimation of equation (1). Before we estimate the model, we determine the number of lags l_i for each currency i by applying the recursive t -statistic procedure of Campbell and Perron (1991) – which is based on the general-to-specific method by Hall (1994) – to each individual log real exchange rate.¹⁵ The estimation procedure is then as follows. First, for each currency i , we apply OLS to equation (1). Second, the resulting covariance matrix of the error terms is used as the weighting matrix in a Feasible Generalized Least

¹⁵ This means that for currency i we choose the value of l_i by first setting l_i to some maximal value l_{\max} , then estimating equation (1) by OLS, and subsequently testing whether the last included lag is statistically significant. If so, then l_i is set to this value, else the model is estimated by setting l_i to $l_{\max} - 1$. The procedure is repeated until a significant value of l_i is found or until $l_i = 0$. Following Wu and Wu (2001), we set l_{\max} to 24 and use a 10% significance level.

Squares (FGLS) procedure to estimate the full panel.^{16,17} When applicable, we impose the homogeneity restriction in the second step.¹⁸ For the methodology in which the model is estimated and tested homogeneously, we compute the usual ADF test statistic $\tau = (\beta - 1)/\text{se}(\beta)$ to evaluate the unit root null-hypothesis. To test the unit root hypothesis homogeneously when estimation is carried out heterogeneously, we employ the Multivariate ADF (MADF) test as described in Taylor and Sarno (1998). For the methodology with heterogeneous estimation and testing, inferences about the stationarity of the individual real exchange rates are based on the individual ADF τ -statistics $\tau_i = (\beta_i - 1)/\text{se}(\beta_i)$.

We use Monte Carlo simulations to derive critical values for the test statistics of the various panel tests. In the first step, given the estimated parameters of the model α_i , β_i , $\gamma_{i,k}$, and l_i , we compute the residuals and the corresponding covariance matrix. Second, we generate N error terms $u_{i,t}$ (T times) from a panel normal distribution with mean zero and this covariance matrix. Third, given the estimated parameters $\gamma_{i,k}$ and l_i , assuming $\alpha_i = 0$, and imposing the null-hypothesis $\beta_i = 1$, we simulate exchange rate series using equation (1) on the basis of the simulated error terms $u_{i,t}$ from step 2. Fourth, given the value of l_i , we estimate the parameters in equation (1) with the simulated exchange rate series and compute the test statistic. We replicate steps 2–4 1000 times and derive critical values for the test statistic from its sample distribution. Empirical p -values can be calculated as the fraction of times the observed test statistic using the actual empirical data series is exceeded in the replications.

A.2. Monte Carlo simulations

Our purpose is to investigate the finite sample properties of the three methodologies, and specifically potential estimation biases and the statistical power of the unit root test.¹⁹ Our approach is inspired by Taylor and Sarno (1998) and is based on Monte Carlo experiments. We simulate real exchange rate series for several data generating processes (DGP's) using equation (1) as described in Section A.1 but instead of setting $\beta_i = 1$ for all exchange rates to construct the distribution of the ADF test statistic under the unit root null-hypothesis (see step 3 of Section A.1), we assign other values to the mean reversion parameters. We follow Taylor and Sarno (1998) and use the following set of values for β_i : {0.99, 0.975, 0.95, 0.925, 0.9}. We use an initial value of zero for all of the constructed real exchange rate series and discard the first 100 generated data points. We note that for both stationary and non-stationary series, this procedure implies that differences in the simulations are not driven by differences in the initial values.

We subsequently construct the power functions of the tests by comparing the simulated values of the test statistic to the 5% critical value obtained using the procedure described in Section A.1. To gain further insight into the power properties of the panel tests, we monitor the estimate of the mean reversion parameter and the estimate of the uncertainty of this parameter. In particular, we assess the bias in the estimate of β that can arise because of the homogeneity restriction (see, e.g., Pesaran and Smith, 1995; Robertson and Symons, 1992).²⁰

¹⁶ The estimation procedure can be extended by iterating between estimating parameters, constructing the corresponding residuals, and using the covariance matrix of these residuals in the FGLS step. However, following Taylor and Sarno (1998), we limit this iterative process to one iteration for computational reasons. Unreported results show that this does not have an important effect on the resulting estimates.

¹⁷ In a standard SUR model the degrees of freedom needed for the calculation of the covariance matrix of the error terms equals $T - k$, where k is the number of variables. We have to make a correction for the inclusion of lagged changes in the real exchange rate in the model. To that effect, we set the degrees of freedom used for $\sum_{ij} to $T - 2 - \text{entier}\{(l_i + l_j + 1)/2\}$, where $\text{entier}[x]$ rounds x down to the nearest integer value.$

¹⁸ Note that the homogeneity restriction only applies to the mean reversion parameters β_i and not to the intercepts α_i .

¹⁹ The aim of this paper is to compare the properties of methodologies that employ heterogeneous and homogeneous estimation of the mean reversion parameter. The literature has identified several other biases that affect both methodologies, such as the time aggregation bias (see, e.g., Choi et al., 2006) and the lagged dependent variable bias (see, e.g., Boyd and Smith, 1999). We do not address these biases here.

²⁰ Naturally, biases in estimates of coefficients are a property of the estimation method, and not of the statistical test. However, since the testing methodologies in the PPP literature are intricately linked to the way the mean reversion coefficient is estimated, we examine biases and test properties simultaneously. We also use the term "bias" in the homogeneous mean reversion coefficient in a rather loose way (i.e., relative to the unweighted average of the individual mean reversion coefficients), as the "true" value of the homogeneous coefficient is undefined.

Appendix B: Supplementary material

Supplementary data related to this article can be found online at doi:[10.1016/j.jimonfin.2010.10.004](https://doi.org/10.1016/j.jimonfin.2010.10.004).

References

- Abuaf, N., Jorion, P., 1990. Purchasing power parity in the long run. *J. Finance* 45, 157–174.
- Boyd, D., Smith, R., 1999. Testing for purchasing power parity: econometric issues and an application to developing countries. *Manchester Sch.* 67, 287–303.
- Breuer, J., McNown, R., Wallace, M.S., 2001. Misleading inferences from panel unit-root tests with an illustration from purchasing power parity. *Rev. Int. Econ.* 9, 482–493.
- Breuer, J., McNown, R., Wallace, M.S., 2002. Series-specific unit root tests with panel data. *Oxford Bull. Econ. Statist.* 64, 527–546.
- Campa, J.M., Wolf, H.C., 1997. Is Real Exchange Rate Mean Reversion Caused by Arbitrage?. In: NBER Working Paper No. 6162.
- Campbell, J.Y., Perron, P., 1991. Pitfalls and opportunities: what macroeconomists should know about unit roots. In: Blanchard, O., Fischer, S. (Eds.), *NBER Macroeconomics Annual*, vol. 6. MIT Press, Cambridge, pp. 141–201.
- Chen, S.S., Engel, C.M., 2005. Does 'aggregation bias' explain the PPP puzzle? *Pac. Econ. Rev.* 10, 49–72.
- Cheung, Y.-W., Chinn, M.D., Fujii, E., 2001. Market structure and the persistence of sectoral real exchange rates. *Int. J. Finance Econ.* 6, 95–114.
- Cheung, Y.-W., Lai, K.S., 2000. On cross-country differences in the persistence of real exchange rates. *J. Int. Econ.* 50, 375–397.
- Choi, C.-Y., 2004. Searching for evidence of long-run PPP from a post-Bretton Woods panel: separating the wheat from the chaff. *J. Int. Money Finance* 23, 1159–1186.
- Choi, C.-Y., Mark, N.C., Sul, D., 2006. Unbiased estimation of the half-life of PPP convergence in panel data. *J. Money Credit Banking* 38, 921–938.
- Dickey, D., Fuller, W., 1979. Distribution of the estimators for autoregressive time series with a unit root. *J. Am. Stat. Assoc.* 74, 427–431.
- Engel, C.M., Hendrickson, M.K., Rogers, J.H., 1997. Intranational, intracontinental, and intraplanetary PPP. *J. Jpn. Int. Economies* 11, 480–501.
- Fleissig, A.R., Strauss, J., 2000. Panel unit root tests of purchasing power parity for price indices. *J. Int. Money Finance* 19, 489–506.
- Flôres, R., Jorion, P., Preumont, P.-Y., Szafarz, A., 1999. Panel unit root tests of the PPP hypothesis. *J. Empirical Finance* 6, 335–353.
- Frankel, J.A., Rose, A.K., 1996. A panel project on purchasing power parity: mean reversion within and between countries. *J. Int. Econ.* 40, 209–224.
- Froot, K.A., Rogoff, K., 1995. Perspectives on PPP and long-run real exchange rates. In: Grossman, G., Rogoff, K. (Eds.), *Handbook of International Economics*, vol. 3. Elsevier, Amsterdam, pp. 1647–1688.
- Hall, A., 1994. Testing for a unit root in time series with pretest data-based model selection. *J. Bus. Econ. Statist.* 12, 461–470.
- Im, K.S., Pesaran, M.H., Shin, Y., 2003. Testing for unit root tests in heterogeneous panels. *J. Econometrics* 115, 53–74.
- Imbs, J., Mumtaz, H., Ravn, M.O., Rey, H., 2005. PPP strikes back: aggregation and the real exchange rate. *Quart. J. Econ.* 120, 1–44.
- Jorion, P., Sweeney, R., 1996. Mean reversion in real exchange rates: evidence and implications for forecasting. *J. Int. Money Finance* 15, 535–550.
- Kilian, L., Zha, T., 2002. Quantifying the uncertainty about the half-life of deviations from PPP. *J. Appl. Econometrics* 17, 107–125.
- Koedijk, C.G., Tims, B., van Dijk, M.A., 2004. Purchasing power parity and the Euro area. *J. Int. Money Finance* 23, 1081–1107.
- Levin, A., Lin, C.-F., Chu, C.-S.J., 2002. Unit root tests in panel data: asymptotic and finite sample properties. *J. Econometrics* 108, 1–24.
- Lopez, C., Papell, D.H., 2007. Convergence to purchasing power parity at the commencement of the euro. *Rev. Int. Econ.* 15, 1–16.
- Lothian, J.R., 1997. Multi-country evidence on the behavior of purchasing power parity under the current float. *J. Int. Money Finance* 16, 19–35.
- Maddala, G.S., Wu, S., 1999. A comparative study of unit root tests with panel data and a new simple test. *Oxford Bull. Econ. Statist.* 61, 631–652.
- Maeso-Fernandez, F., Osbat, C., Schnatz, B., 2001. Determinants of the Euro–Dollar Real Effective Exchange Rate: A Beer/Peer Approach. In: ECB Working Paper No. 85.
- Murray, C.J., Papell, D.H., 2002. The purchasing power parity persistence paradigm. *J. Int. Econ.* 56, 1–19.
- Nelson, F.D., Savin, N.A., 1990. The danger of extrapolating asymptotic local power. *Econometrica* 58, 977–981.
- Ng, S., Perron, P., 2001. Lag length selection and the construction of unit root tests with good size and power. *Econometrica* 69, 1519–1554.
- O'Connell, P.G.J., 1998. The overvaluation of purchasing power parity. *J. Int. Econ.* 44, 1–19.
- Papell, D.H., 1997. Searching for stationarity: purchasing power parity under the current float. *J. Int. Econ.* 43, 313–332.
- Papell, D.H., 2002. The great appreciation, the great depreciation, and the purchasing power parity hypothesis. *J. Int. Econ.* 57, 51–82.
- Papell, D.H., 2006. The panel purchasing power parity puzzle. *J. Money Credit Banking* 38, 447–467.
- Papell, D.H., Theodoridis, H., 2001. The choice of numeraire currency in panel tests of purchasing power parity. *J. Money Credit Banking* 33, 790–803.
- Pesaran, M.H., Smith, R., 1995. Estimating long-run relationships from dynamic heterogeneous panels. *J. Econometrics* 68, 79–113.
- Robertson, D., Symons, J., 1992. Some strange properties of panel data estimators. *J. Appl. Econometrics* 7, 175–189.
- Rogoff, K., 1996. The purchasing power parity puzzle. *J. Econ. Lit.* 34, 647–668.
- Taylor, A.M., Taylor, M.P., 2004. The purchasing power parity debate. *J. Econ. Perspect.* 18, 135–158.
- Taylor, M.P., Sarno, L., 1998. The behavior of real exchange rates during the post-Bretton Woods period. *J. Int. Econ.* 46, 281–312.
- Wu, J.-L., Wu, S., 2001. Is purchasing power parity overvalued? *J. Money Credit Banking* 33, 804–812.