Measuring Persistence of U.S. City Prices: New Evidence from Robust Tests*

Syed A. Basher[†] Josep Lluís Carrion-i-Silvestre [‡]

May 6, 2009

Abstract

We revisit the empirical analysis in Cecchetti, Mark and Sonora (2002) by applying modern panel stationary tests to an expanded data set. The estimation techniques are robust to the presence of multiple structural breaks and cross-section dependence, features that are present when analyzing long time series of integrated economies. We also discuss on the wide range of purchasing power parity (PPP) concepts that arise when the data are subject to structural breaks. Testing for PPP with structural breaks requires the definition of parametric restrictions (parity restrictions) across regimes. Our results suggest that once the structural breaks are isolated, the U.S. city price level differentials are I(0) stationary processes, with the median half-life of convergence ranged between 1.50 to 2.65 years. As a robustness check of our findings we also conduct a pairwise tests of price level convergence.

Keywords: Purchasing power parity; price level convergence; half-life; multiple structural breaks; cross-section dependence; pairwise approach.

JEL Classification: C32, C33, E31, F41

*The views expressed here are those of the authors and do not reflect the official view of the Qatar Central Bank. The usual disclaimer applies.

[†]Department of Research and Monetary Policy, Qatar Central Bank, P.O. Box 1234, Doha, Qatar. E-mail: bashers@qcb.gov.qa

[‡]Corresponding author: AQR-IREA Research Group, Department of Econometrics, Statistics, and Spanish Economy, University of Barcelona. Av. Diagonal, 690. 08034 Barcelona. Tel: +34 934024598; Fax: +34 93 4021821; E-mail: carrion@ub.edu.

"Structural change is pervasive in economic time series relationships, and it can be quite perilous to ignore. Inferences about economic relationships can go astray, forecasts can be inaccurate, and policy recommendations can be misleading or worse."

Bruce Hansen, 2001

1 Introduction

In a recent paper appeared in this *Review*, Cecchetti, Mark and Sonora (2002) (hereafter CMS), using long-term series of consumer price indices for 19 US cities, show that price index divergences across US cities are temporary but surprisingly persistent, with a half-life of nearly nine years. To uncover explanations behind the slow rate of convergence, CMS examine the role of distance, asymmetric adjustment and non-traded goods prices. However, none of the factors were shown to provide significant explanations for the slow rate of convergence.

Indeed, such long-lasting deviations from their equilibrium levels, as defined by the theory of purchasing power parity (PPP) is problematic for model with nominal rigidities, which predict faster convergence to PPP of 1-2 year half-life.[1] This problem was highlighted in Rogoff (1996), who set out an empirical puzzle[2] relating to the apparent contradiction between the slow speed of reversion to PPP (about 3-5 years) and the observed large short-term volatility of real exchange rates.[3]

In practice, estimation of the half-life of convergence is subject to different sources of bias. Two well-known biases are aggregation bias and small-sample attenuating bias (also known as 'Nickell bias'). Aggregation bias can results due to two empirical pitfalls: temporal aggregation bias (Taylor, 2001) and dynamic aggregation bias (Imbs et al., 2005).[4] While CMS correct for the Nickell bias when computing the half-life of PPP deviation, they fail to recognize and correct for the aggregation biases. Nath and Sarkar (2007) correct for both aggregate and small-sample biases applying the framework developed in Choi et al. (2006).[5] These bias-corrected estimates of autoregressive coefficients are then used to obtain various unbiased estimates of the half-life to analyze price index convergence. Based on both the original CMS data and an extended data set, Nath and Sarkar find the half-life to be about seven years – two years shorter than the estimate of CMS.

[1] The half-life is a common measure of persistence in the deviations from PPP.

[2] It is called a puzzle (commonly know as the 'PPP puzzle') because while PPP is compatible with high short-term exchange rate volatility, it also implies that deviations should be short-lived (that is 1-2 years). See Rogoff (1996) for further discussion.

[3] Since the U.S. cities share the same currency, price difference between two cities can be seen as the real exchange rate.

[4] Briefly, dynamic aggregation bias arises due to heterogeneity in the speed of price adjustment at the goods level, whereas temporal aggregation bias arises because in practice, statistical agencies report price indices over a particular interval (and thus at a higher frequency) than the point-in-time prices at which the data are actually sampled. On the other hand, small-sample estimation bias results when the dynamic autoregression is run with a constant. Generally, these biases have been addressed individually in the literature. See Choi et al. (2006) for a simultaneous treatment of all three biases with the application to country-level real exchange rates.

[5] Note that, dynamic aggregation bias as defined in Imbs et al. (2005) does not directly arise with city/country price indices. In practice, researchers essentially allow for the possibility of heterogeneity in city/country data with hindsight that heterogeneity in the speed of price adjustment at the goods level might have induced heterogeneity in city/country data.

An interesting question to be analyzed is why the deviations from US city PPP are substantially more persistent than deviations from cross-country PPP (as indicated by Rogoff's consensus of 3-5 years). One possible explanation is the use of long spans of data by CMS (and maintained in Nath and Sarkar), which are more likely to be affected by structural breaks. The structural breaks can appear either because the data have been sampled across several different monetary arrangements or by the presence of shocks such as oil price shocks. Neglect of structural breaks in the analysis can be costly and may lead to misleading results, as highlighted by Bruce Hansen in the epigraph above.

It is well-known in the time series econometric literature that the persistence of a process can be overestimated if one ignores break(s) in the data occurred at some point in time: one would take for a slow reversion to mean what in fact is a structural (permanent) change (Perron, 1989).[6] We make use of this well-known result and re-evaluate the convergence process when US city price levels are subject to multiple structural breaks. We use price level (rather than price index) to measure the cost of a given consumption basket at each point in time (see Chen and Devereux (2003)). The point we highlight in this paper is that lack of accounting for structural breaks in the computation of the half-life might be a reason behind the slow rate of convergence documented by CMS and Nath and Sarkar (2007), among other authors.

The fact that the accommodation of structural breaks in the analysis can overturn the results has not gone unnoticed in the literature. For example, whereas Lopez et al. (2005) find limited evidence of PPP for a sample of industrialized economies, Papell and Prodan (2006) find more evidence in favor of the PPP hypothesis using the same data but controlling for structural breaks. Similarly, Gadea et al. (2004) are unable to reject the null hypothesis of unit root for a set of European real exchange rates when structural breaks are omitted from the analysis, although their conclusion is revered once structural breaks were incorporated. In other study, Altissimo et al. (2006) note that the unit root can hardly be excluded over long period for inflation if the structural breaks are ignored, while inflation persistence appears to be relatively low once the discrete changes in the mean are incorporated in the model.

Even though structural breaks issue has received increasing attention in the analysis of PPP in recent years (both in time series and panel data), there is one important issue that is often overlooked: which notion of PPP to test when there are structural breaks in the data? The two widely used concepts of PPP are the ones defined by Cassel (1918), and Balassa (1964) and Samuelson (1964), which are not valid, strictly speaking, in the presence of structural breaks. For proper fulfillment of PPP, additional investigations should be conducted through imposing the so-called *parity restrictions* in price differentials when they are affected by the presence of structural breaks – see Papell and Prodan (2006) for further details.

Until recently, testing for multiple structural breaks while simultaneously addressing different concepts of PPP could not be undertaken in an unified framework. In a related paper, Basher and Carrion-i-Silvestre (2008) (hereafter BCiS) proposed an econometric framework that encompasses the presence of multiple structural breaks while simultaneously testing for different concepts of PPP that have been proposed in the literature. In particular, BCiS design individ-

[6] The presence of structural break(s) can be interpreted as real economic shocks affecting time series permanently. As movements in the real exchange rate are tantamount to deviations from PPP, Kim and Enders (1991) argue that these real economic shocks can induce permanent deviations from PPP.

ual KPSS and panel stationary tests to accommodate the parity restrictions that are required to meet the PPP hypothesis – further details are provided below. In this paper, we make use of the BCiS methods to re-evaluate the empirical analysis in CMS to an expanded data set of US city prices.

Briefly, our main results can be summarized as follows. Regardless of the parity restrictions, we find overwhelming evidence that price differentials across US cities are I(0) stationary once the structural breaks are accounted for. Using the estimated models that include the structural breaks, the median half-life of convergence ranged between 1.50 to 2.65 years, much faster than those documented in previous studies. Our results remain robust when the analysis is extended to the pairwise framework devised in Pesaran (2007).

The rest of the paper is structured as follows. Section 2 presents a brief account of the various PPP concepts and the econometric methodologies used in this paper. Section 3 presents the data and main empirical results. Section 4 presents the results based on the pairwise approach. Section 5 concludes.

2 Empirical Methods

This section is divided into two parts. First, we briefly review the different concepts of PPP that have been developed in the literature when multiple structural breaks are present in the data. Second, we outline the econometric methodologies that are used in this paper. For full details, readers are referred to BCiS.

2.1 PPP and Structural Breaks

The works of Cassel (1918), Balassa (1964) and Samuelson (1964) defined two well accepted and popular PPP definitions, which have been profusely studied in the economic literature. One way to find out whether PPP hypothesis holds is by assessing the order of integration of the real exchange rates – evidence that real exchange rates are I(1) will point against the PPP hypothesis, while finding them to be I(0) will be consistent with the PPP hypothesis. The distinction between Cassel and Balassa-Samuelson definitions can be established depending on the deterministic component that is used to assess the stochastic properties of real exchange rates. Thus, when the deterministic component that is used in the computation of the unit root and stationarity tests is given by a constant term we are dealing with Cassel's (1918) definition of the PPP. By contrast, Balassa (1964) and Samuelson (1964) devise a second concept of PPP when noticing that divergent international productivity lead to permanent deviations from the Cassel's PPP concept definition. This feature is captured through the specification of a long-run trend around which the real exchange rates would show stationary fluctuations, which defines the so-called "Trend PPP" (TPPP). Therefore, in this case unit root and stationarity test statistics have to use a linear time trend as the deterministic component when testing for TPPP.

However, these notions of PPP are not valid when structural breaks are present in the data, since they assume stable deterministic components. Therefore compatible definitions of PPP must be used for the proper fulfillment of the classical PPP hypothesis, giving rise to the following generalizations:

1. **Quasi Purchasing Power Parity (QPPP):** Testing whether the real exchange rates are I(0) stationary around a changing level.

2. **Trend Qualified Purchasing Power Parity (TQPPP):** Testing whether the real exchange rates are I(0) stationary around a deterministic component given by a linear time trend with level shifts.

As can be seen QPPP (TQPPP) is the time-varying analogue of Cassel (Balassa-Samuelson) concept of PPP that can handle the presence of structural break(s). Nevertheless, evidence in favor of QPPP or TQPPP does not imply that PPP as defined in Cassel or Balassa-Samuelson is fulfilled, since in these cases PPP requires reversion towards a constant mean or a constant trend in the long-run. Therefore, in the presence of structural breaks, QPPP or TQPPP is necessary but not sufficient condition for the classical PPP definitions to hold. In this case, when we have found evidence in favor of QPPP or TQPPP further investigations should be conducted to conclude that the PPP hypothesis is satisfied according to the classical definitions in Cassel or Balassa-Samuelson. To be specific, we require to impose the so-called parity restrictions on the coefficients of the first and last regimes so that the coefficients of these regimes are of the same sign and magnitude. Note that after imposing the parity restrictions the deterministic component does not change in the long-run. This implies that after the last break has occurred, the deterministic component of the time series equals the one previous to the first structural break. In brief, when parity restrictions are imposed we are implicitly assuming that structural breaks have only temporary effects on time series.

2.2 Econometric Methodology

Let $y_{i,j,t} = (\ln p_{i,t} - \ln p_{j,t})$ be the difference between the logarithm of two time series of prices. Assume that the data generating process (DGP) is expressed using orthogonal regressors for the model that includes a time trend with both level and slope shifts as:

$$y_{i,j} = x_{i,j}\delta_{i,j} + \varepsilon_{i,j}, \tag{1}$$

where $x_{i,j} = diag\left(x_{i,j,1}, \ldots, x_{i,j,m_{i,j}+1}\right)$, $x_{i,j,k,t} = (1, t)$, $\delta_{i,j} = \left(\delta_{i,j,1}, \ldots, \delta_{i,j,m_{i,j}+1}\right)'$ and $\delta_{i,j,k} = \left(\mu_{i,j,k}, \beta_{i,j,k}\right)'$ for $T^{i,j}_{b,k-1} < t \leq T^{i,j}_{b,k}$, $k = 1, \ldots, m_{i,j}+1$, with the convention that $T^{i,j}_{b,0} = 0$ and $T^{i,j}_{b,m_{i,j}+1} = T$, being $m_{i,j}$ the number of structural breaks for the (i,j)-th pair of time series, $i, j = 1, \ldots, N$, $i \neq j$. The parity restrictions imply that the parameters of the first regime, $\delta_{i,j,1}$, and the ones for the last regime, $\delta_{i,j,m_{i,j}+1}$, have to be equal, while the parameters of the other regimes are left free. Note that this is because the model is specified in a way that regressors are block orthogonal, so that the parameters on each orthogonal block are not affected by the values of the parameters in the previous blocks.[7] In this set-up these parity restrictions can be expressed as $R\delta_{i,j} = r$ with $R = \begin{bmatrix} I_l & 0_{l\times(m_{i,j}-1)l} & -I_l \end{bmatrix}$ and $r = 0_{(m_{i,j}+1)l\times 1}$, where I_l denotes the identity matrix, l is the number of regressors in $x_{i,j,k}$ – $l = 1$ in the constant only case and $l = 2$ in the case of the linear time trend – and $0_{a\times b}$ an $(a \times b)$-matrix of zeros. Using these elements, we can compute the restricted least squares estimator $\left(\hat{\delta}^*_{i,j}\right)$ of $\delta_{i,j}$ in (1) such that the estimator satisfies $R\delta_{i,j} = r$. Note that, at least two structural breaks are required in order to impose the parity restrictions, since parity restrictions for the one break case will imply the absence of the structural break.

To estimate the restricted least squares estimator of $\delta_{i,j}$, BCiS utilize the dynamic programming algorithm of Perron and Qu (2006), which permits the consideration of multiple structural breaks with restrictions among the parameters of the different regimes. The estimation of BCiS method proceeds in two stages: (i) estimate the number of structural breaks using the unrestricted dynamic algorithm in Bai and Perron (1998) and, conditional to the number of structural breaks, (ii) minimize the restricted sum of squared residuals to estimate the position of the structural breaks with the vector of parameters $\delta^*_{i,j}$ satisfying $R\delta_{i,j} = r$, with R and r defined above. The restricted estimated disturbance terms $\hat{\varepsilon}^*_{i,j}$ are then used to compute the individual-by-individual restricted KPSS[8] statistic:

$$\eta^*_{i,j}\left(\lambda_{i,j}\right) = \hat{\omega}^{-2}_{i,j} T^{-2} \sum_{t=1}^{T} \hat{S}^{*2}_{i,j,t}, \tag{2}$$

and the corresponding restricted panel data $(Z^*(\lambda))$ statistic in BCiS. In (2), λ is defined as the vector $\lambda_i = (\lambda_{i,1}, ..., \lambda_{i,m_i})' = \left(T^i_{b,1}/T, ..., T^i_{b,m_i}/T\right)'$, which indicates the relative position of the dates of the breaks on the entire time period, T, for each individual i, $i = 1, \ldots, N$. The limit distribution of the statistic depends on the number of structural breaks as well as on their relative position in the sample. The asymptotic critical values of the statistic are approximated using the estimated response surfaces in BCiS. The Monte Carlo experiments show that the empirical size of the statistic is close to the nominal size when the requirements of the model specification are met. Moreover, the presence of structural breaks does not cause any power loss to the test statistics – see BCiS for further details.

As the BCiS framework is based on the assumption of cross-section independence, which is unlikely to be met in the case of US city price data, we account for cross-section dependence in the following way. First, we test for the presence of cross-section dependence using two recently proposed tests by Pesaran (2004) and Ng (2006). Provided that the null hypothesis of no cross-section dependence is rejected in the first step, we then account for cross-section dependence using (i) Levin, Lin and Chu (2002) (quite restrictive) suggestion to remove the cross-section mean, which is equivalent to include temporal effects in the panel data set, and (ii) Maddala and Wu (1999) procedure to compute the empirical distribution by means of parametric bootstrap.

[7] By contrast, when the model is not specified in terms of orthogonal regressors, the restrictions can be imposed on the parameters of the model so that the sum of all the coefficients for the level shifts (and the slope shifts if we deal with a trending variable) is equal to zero.

[8] KPSS denotes the test statistic proposed in Kwiatkowsky, Phillips, Schmidt and Shin (1992).

3 Empirical results

This section begins with a short description of the data followed by discussion of our main empirical results. In the following subsections we provide detailed evidence on the dynamics of price level differences among US cities, both with and without the parity restrictions discussed above. We then present the results of half-life estimates (the *speed* of convergence), again with and without the parity restrictions. Throughout the paper, the level of significance that is used is set at the 5% level.

3.1 Data

We use the same data as BCiS which includes annual consumer price index (CPI) covering the period 1918 to 2005 ($T = 88$) for $N = 17$ US cities: Atlanta, Boston, Chicago, Cincinnati, Cleveland, Detroit, Houston, Kansas City, Los Angeles, Minneapolis, New York, Philadelphia, Pittsburgh, Portland, San Francisco, Seattle, and St. Louis.[9] All data were extracted from the Bureau of Labor Statistics's webpage (www.bls.gov). It is worth mentioning that the price indices were converted to *absolute* price levels so that they provide information on the relative cost of living across cities at a point in time – see BCiS for further details. Furthermore, as we use aggregate price levels, this may explain why the presence of a time trend is required since, unlike individual price levels, aggregate price levels may not adjust so quickly due to differences in productivity among tradable and non-tradable sectors. Besides, aggregate data is less affected by measurement error over disaggregate data. Thus, when the absolute city prices are converging, bilateral real exchange rates may contain a time trend, with or without structural breaks depending on the pattern of convergence. Figure 1 depicts the price level differentials for four cities using the aggregate US price as the benchmark. As can be seen, the evolution of these series are quite different and show that movements around a constant level for the whole period might be not suitable.

3.2 Convergence towards the US price level

3.2.1 Convergence without the parity restrictions

In this section we provide the results that are based on the use of the aggregate US price as the benchmark. The number and position of the structural breaks are estimated using the sequential procedure in Bai and Perron (1998) or the LWZ information criterion in Liu, Wu and Zideck (1997), depending on whether we specify a broken linear trend or not. The maximum number of structural breaks is set at $m^{\max} = 5$ (this maximum was never attained). The results in Panel A of Tables 1 and 2 show that at least one structural break is detected for each city regardless of the deterministic specification used. This clearly indicates the importance of accounting for structural breaks when modeling price level convergence, in particular for long span of data such as the one used in the present paper.

[9]Note that, the original Cecchetti et al. (2002)'s sample consists of 19 cities including Baltimore and Washington DC. However, since 1996, the Bureau of Labor Statistics no longer maintains separate data for these two cities. As a result, these cities are excluded from the analysis.

Let us first focus on the results for the case where the deterministic component is given by a constant term with level shifts (QPPP specification). Panel A in Table 1 presents the individual KPSS statistics along with the corresponding critical values, which indicate that, except for Kansas City and San Francisco, the null hypothesis of I(0) stationarity cannot be rejected. Panel A in Table 2 reports the results for the case where the deterministic component is given by a linear time trend with structural breaks that affect both the level and the slope (TQPPP specification). In this case the null hypothesis of I(0) stationarity is only rejected for Atlanta, Boston, Cincinnati, Houston, and Portland when using the individual KPSS statistic. Overall, these results provide favorable evidence in support of the price level convergence (either QPPP or TQPPP definitions) among the US cities since the null hypothesis of I(0) cannot be rejected for each time series at least for one of the deterministic specifications that have been essayed.

Before we analyze the panel results, we first examine the extent of cross-section dependence in the data. Table 3 reports the cross-section correlations based on Pesaran (2004) and Ng (2006). We have estimated an autoregressive model that includes dummy variables to account for the presence of level shifts – the order of the autoregressive correction is selected using the t-sig criterion in Ng and Perron (1995) with up to ten lags. As can be seen, for both deterministic specifications the CD statistic strongly rejects the null hypothesis of cross-section independence. Likewise, the Ng (2006) statistic for the whole sample of correlations ($svr^{W}(\hat{\eta})$) strongly rejects the null hypothesis of independence. Note, however, that the null hypothesis of cross-section dependence cannot be rejected for the small ($svr^{S}(\hat{\eta})$) and large ($svr^{L}(\hat{\eta})$) groups of correlations, which possibly indicates that the extent of correlation detected in the whole sample is not strong and pervasive. Taken together, these results suggest that we have to take into account the presence of cross-section dependence across units when testing the null hypothesis of panel I(0) stationarity.

Panel B in Tables 1 and 2 presents the panel data statistics in Carrion-i-Silvestre et al. (2005) that permit the presence of the estimated structural breaks. The results that are based on the independence assumption are not reliable given the evidence of cross-section dependence reported by Pesaran (2004) and Ng (2006) statistics. Let us first focus on Table 1, where the QPPP hypothesis specification is investigated. When we take into account the presence of cross-section dependence either by removing the cross-section mean or by the use of the bootstrapped distribution, both versions of the $Z(\lambda)$ tests lead to the non-rejection of the null hypothesis of I(0) stationarity, which indicates that price convergence took place during the analyzed period. Table 2 reports the results corresponding to the TQPPP hypothesis. In this case the $Z(\lambda)$ tests reject the null hypothesis of I(0) stationarity when cross-section is accounted for through cross-section demeaning, although it is not rejected when we use the bootstrap critical values. It is worth mentioning that cross-section demeaning is quite restrictive since this approach assumes that the cross-section dependence is driven only by one common factor that, in turn, affects all the time series in the panel data set with the same magnitude. Therefore, we give more emphasis on the results that are based on the use of the bootstrap distribution.

Finally, we have also computed the panel data statistics that consider a mixture of model specifications, assuming that for some real exchange rates the QPPP hypothesis may be better than the TQPPP hypothesis and the converse. In order to select between the QPPP and the TQPPP specifications for each time series we have used the BIC information criterion – the same selection is achieved using the AIC information criterion. The specification that has been chosen in almost all cases is the TQPPP one, with the exception of Cincinnati, Minneapolis and Saint Louis for which the QPPP specification is selected. Detailed results are reported in the companion Appendix available from the authors upon request.

Panel A in Table 4 reports the panel data statistics based on the mixture of QPPP and TQPPP specifications. Results show that the null hypothesis of I(0) cannot be rejected when using the bootstrap critical values, although it is rejected when cross-section dependence is accounted for using the cross-section demeaning procedure. Provided that accounting for cross-section dependence through cross-section demeaning is quite restrictive, we rely on the results of the bootstrap approach.

To sum up, we can conclude that there is evidence in favor of the QPPP and TQPPP hypotheses, either at individual or panel data level, when the price differentials are computed using the US as the benchmark.

3.2.2 Convergence with the parity restrictions

As mentioned above, evidence in favor of QPPP or TQPPP hypotheses does not imply that PPP is satisfied. Therefore, in this section we consider the parity restrictions that allow us to test the PPP hypothesis when there are structural breaks. Again, we have dealt with both specifications, i.e. the constant term with multiple level shifts and the time trend with multiple level and slope shifts – to the best of our knowledge, this is the first time that this second specification is used in the literature of the PPP hypothesis. Note that parity restrictions can only be imposed in those situations where at least two structural breaks have been detected. Thus, in cases where only one structural break is found (e.g. New York), imposing the parity restrictions will nullify the presence of structural break. Therefore, we only report the results for those cities for which $\hat{m}_{i,US} > 1$ (see Tables 1 and 2). The number of structural breaks is the one estimated in the previous section where parity restrictions were not considered.

Tables 5 and 6 present the individual KPSS statistics, respectively, for the QPPP and TQPPP hypothesis with the parity restrictions. Table 5 indicates that the null hypothesis of I(0) is rejected for Atlanta, Cleveland, Houston, Kansas City, San Francisco, and Seattle – i.e. in six out of fourteen cases – when using the change in level specification. Besides, when the time trend is considered, the null hypothesis of I(0) is rejected for Atlanta, Boston, Chicago, Houston, New York, Portland, and Seattle – i.e. in seven out of fourteen cases – see Table 6. Taken together, we can see that for Atlanta, Houston, New York, and Seattle none of the deterministic specifications avoid the rejection of the null hypothesis of I(0), which shows evidence against the PPP fulfilment in these cases. We have to include Los Angeles in this group since only one structural break has been detected and, hence, parity restrictions cannot be imposed. For the other situations we have found some evidence supporting either the PPP or the TPPP hypotheses.

Panel B in Tables 5 and 6 report the panel data based statistics. These results show strong evidence against the PPP hypothesis provided that the null hypothesis of I(0) is rejected when parity restrictions are imposed to the coefficients of the estimated break points. However, panel data based evidence is favorable to the TPPP hypothesis since the null hypothesis is not rejected when the statistics are compared with the bootstrap critical values. As above, we can impose the parity restrictions to the model that arises from the use of the BIC information criterion and define a panel data set with a mixture of model specifications – the model with the constant with multiple level shifts is specified only for Cincinnati, Minneapolis and Saint Louis. In this case the null hypothesis of I(0) is not rejected – see Panel B in Table 4.

In all, our analysis indicates that the more flexible specification that considers changes in the time trend tends to support the Balassa-Samuelson definition of the PPP hypothesis. The Balassa-Samuelson effect indicates that productivity differential between tradable and nontradable sectors will lead to changes in real costs and relative prices, bringing about divergences in exchange-rate adjusted national (*city*, in our case) price levels.[10] In a recent much-cited paper, Engel (1999) challenges this conventional view and concludes that, nearly 100% of U.S. based CPI real exchange rate variation is explained by deviations from the law of one price for tradable goods, and virtually *none* by differentials in the relative price of nontradables across countries.[11] This implies that the Balassa-Samuelson effect does not help to explain movements in real exchange rates. However, we expect the Engel effect to be weaker within the United States, as the rate of convergence of actual prices (not price indexes) is shown to much faster for tradable goods over the nontradable goods (Parsley and Wei, 1996). Although our analysis does not incorporate disaggregate price levels, one would expect the movements in relative price of nontradables – due to differential in productivity growth – within the United States may have largely worked their way to explain the variations in price levels among U.S. cities.

3.3 Half-life estimates

In this section we re-evaluate the speed of price level convergence among US cities using the framework designed in the previous sections. The remarkable slow rate of convergence as documented by CMS should come as a little surprise if we realize that long span aggregate price data are more susceptible to structural instability because unlike individual price level aggregate price level differentials may not adjust so quickly due to differences in productivity among tradable and non-tradable good sectors. Further, neglecting structural breaks in the computation of the speed of convergence will clearly lead to a model misspecification, which, in turn, will bias upward shocks persistence of price differentials. This important issue has not received much attention in the PPP literature. Our aim here is to fill that gap.

Following CMS we have estimated the persistence of price level adjustment using the popular half-life (HL) measure – i.e., the time it takes for 50% of a shock to the price level to dissipate. As is well known that least squares (LS) estimators of the AR(p) model with time trend generate substantial biases, we have employed the approximate median unbiased (MU) estimators of Andrews and Chen (1994), which provides a bias-correction for the LS estimator.[12]

[10] Demand factors – notably real government consumption – may be an alternative cause for regional price differentials if there is a bias towards the service sector, since this will tend to raise the relative price of nontradables.

[11] Engel's (1999) study examines five high-income countries (for CPI based real exchange rates), but finds similar results using output price indexes (seven countries) and producer price indexes (sixteen countries).

[12] Formally, idea behind the concept of median-unbiasedness can be explained as follows. Let $m(\alpha_i)$ denote

To implement the MU estimator, we have estimated an AR(p) model for each time series using the number and position of the structural breaks that have been obtained in the previous sections. For each time series, we have selected between the PPP and TPPP sort of models using the BIC information criterion.[13] Thus the results presented in Table 7 point to the best model (either QPPP or TQPPP) for each series. These results are further categorized as unrestricted (when PPP is not imposed, second column in Table 7) and restricted (when PPP is imposed, third column in Table 7) models.

Table 7 presents the HL estimates when US price level is used as the benchmark. The results are very encouraging, with point estimates between one to three years for majority of the cities. More importantly, the average and median speed of convergence are not only faster than the estimates obtained by CMS and Nath and Sarkar (2007), they are also well below than Rogoff's (1996) consensus estimates of 3–5 years. Results are not upset when the PPP restriction is imposed in the computation of HLs. As can be seen, except for Minneapolis and San Francisco, the estimated HLs for remaining cities are within the neighborhood of consensus range. Furthermore, while the average HL is affected due to presence of outliers, the median point estimate of 2.654 years is still below the consensus estimates. This is reassuring for the different approaches used in the analysis, as both restricted and unrestricted models show evidence of faster reversion to price level parity in general.

There is little work in the literature addressing (all) potential sources of biases in the estimation of the half-life to PPP convergence. While several recent papers – e.g. Choi et al. (2006), Nath and Sarkar (2007) – make use of recent methodologies to control for some well-known biases, these papers remain silent about how the presence of structural breaks can provide imprecise and biased estimates of the autoregressive parameters, and thereby providing inaccurate picture of the speed of adjustment to PPP. Our analysis sheds light on this often neglected issue and raises a warning flag about econometric work on the PPP puzzle.

4 Extensions: Pairwise analysis

The preceding analysis assumes that US city price levels converge towards the aggregate US price level. Some authors have pointed out that the main drawback of this approach is that results can be sensitive to the choice of the benchmark and, as a result, can lead to misleading conclusion.[14] For example, it could be that the price deviations between a pair of cities is I(0) stationary, but their deviations computed separately against the aggregate US price level could be non-stationary. The fact that price levels converge between this pair would be lost by just

the median function of an arbitrary estimator, $\widehat{\alpha}_i$ say, of α_i. This function is defined by $P(\widehat{\alpha}_i < m(\alpha_i)) = 0.5$, which can be inverted to obtain another estimator $m^{-1}(\widehat{\alpha}_i)$ of α_i. By construction, this estimator satisfies $P(m^{-1}(\widehat{\alpha}_i) < \alpha_i) = 0.5$ so the probability of underestimation is equal to the probability of overestimation. An estimator that has this property is said to be median-unbiased.

[13] Results remain unaffected if we use the AIC information criterion. A companion appendix containing unreported results is available on request from the corresponding author.

[14] In cross-country PPP studies the problem of benchmark effect can be quite damaging. For example, Papell and Theodoridis (2001) show that PPP is more likely to hold with European (e.g. German mark) than non-European (e.g. U.S. dollar) as the numeraire currency. A common explanation for this behavior is that the former is generally less volatile than the latter. Besides, European countries are relatively more open to international trade than the USA and Japan. The geographical proximity of European countries makes goods arbitrage more effective since transaction costs are lower.

focusing on the aggregate US price level (Pesaran et al. (2007)). To overcome this limitation, we use pairwise tests of price convergence à la Pesaran (2007), which is not sensitive to the choice of the benchmark. Thus, if we analyze N time series of prices, the pairwise tests focus on all possible $N(N-1)/2$ price deviation pairs between the time series in the panel, which in addition can consistently estimate the proportion of pairs that do not converge. The pairwise approach allows us to obtain more robust results than the ones based on a particular benchmark (e.g. aggregate US price level). In this section, we check the robustness of our results reported above by applying the pairwise tests of price level convergence. We first present evidence of (pairwise) price level convergence followed by the half-life of (pairwise) convergence.

4.1 Pairwise convergence with and without the parity restrictions

Evidence of pairwise convergence among the US city price levels is elaborately discussed in BCiS. We therefore briefly review their results in this section. The interested reader is referred to BCiS for further details.

Tables 8 and 9, adopted from BCiS, present the summary results of pairwise convergence. Under the pairwise approach, a natural way to test convergence is by asking what proportion of price-deviation pairs are I(0) stationary. Let us first review the results when the parity restriction is not imposed in the convergence process. The "non-restricted" panel in Table 8 summarizes the proportion of rejections of the null hypothesis of I(0) in each case. We can see that more evidence against the null hypothesis of I(0) is found when the TQPPP specification is considered. The null hypothesis of I(0) is rejected up to 13.2% of the time under QPPP specification as compared to around 35.3% of the time with the TQPPP specification. When choosing between the QPPP and TQPPP hypothesis specifications for each time series pair (mixed case), the proportion of rejections stands at 33.8%. Overall, the magnitude of rejection against the null hypothesis of I(0) stationarity is not large with either QPPP, TQPPP or the mixture QPPP/TQPPP versions of the PPP hypotheses, which lends favorable evidence of price level convergence.

The results on the computation of the panel data statistics are reported in Panel A of Table 9. As can be seen, when cross-section dependence is taken into account through cross-section demeaning, the null hypothesis of I(0) cannot be rejected for the QPPP specification, although the null hypothesis is clearly rejected for the TQPPP and the mixture of the QPPP/TQPPP specifications. Conclusive results are obtained if we base the inference on the bootstrapped critical values. In this case, the null hypothesis of I(0) is not rejected for either the QPPP, the TQPPP or the mixture of the QPPP/TQPPP hypotheses regardless of the statistic that is used.[15]

Panel B in Tables 8 and 9 reports the results of pairwise convergence when the parity restriction is imposed. First of all, we can see that restricting the parameters of the QPPP model yield a proportion of rejections of the null hypothesis of I(0) of 31.5%, while the proportion for the TQPPP hypothesis is of the 8.8%. These results show that imposing parity restrictions on

[15] As above, the selection of the model that is used to compute the panel data statistics with the mixture of the QPPP and TQPPP specifications is based on the BIC information criterion. Only for eighteen out of the one hundred and thirty-six pairs of cities the QPPP specification is selected.

the QPPP model specification may imply incredible restrictions that are not to be satisfied in practice. However, the converse is found for the TQPPP model specification. Thus, the more flexible specification that defines the TQPPP hypothesis is more likely to satisfy the parity restrictions. As expected, when choosing between these two specifications for each individual gives a proportion that lies between both situations – the QPPP specification is selected only 23.8% of the cases.

The picture based on the individual statistics can be completed with the results from the panel data statistics. Table 9 indicates that the null hypothesis of I(0) is rejected for both the restricted QPPP and TQPPP hypotheses when using the bootstrap critical values. These results show that there is no evidence in favor of the PPP hypothesis when we consider the whole panel data set. Therefore, we have to conclude that the PPP hypothesis does not hold for all the pairs of price differentials that have been considered in this paper.

4.2 Pairwise half-life estimates

Let us first focus on the results for the unrestricted specifications. Provided that the panel data statistics that have been applied in the previous section indicates that the null hypothesis of I(0) cannot be rejected for either the QPPP, TQPPP or the combination of the QPPP/TQPPP specification, we have computed MU HLs estimates for all pairs of price differentials. In order to save space, we only report detailed results for the combination of the QPPP/TQPPP specifications in Table 10 – results for the QPPP and TQPPP specifications are available upon request. As can be seen, most of the HL estimates are below the consensus range of 3–5 years mentioned in Rogoff (1996). Overall, the median half-life estimate shows a faster adjustment to PPP than the consensus view. Similar results are also found when using the QPPP and TQPPP specifications.

In order to get a complete picture we have summarized in Table 11 the percentage of HLs that are below, within and above the 3–5 years consensus. Note that the vast majority of HLs are below or within the consensus for the three different situations that we consider, which indicates that taking into account the presence of structural breaks that might be affecting the time series reduces the persistence of the difference in the price levels of the US cities.

A key point made by Taylor and Taylor (2004) is that the equilibrium exchange rate is moving gradually over time,[16] therefore estimates of the speed of reversion based on models with fixed PPP exchange rate will be biased, which may partly be responsible for Rogoff's PPP puzzle. Hence, allowance for long run trends can make a noticeable difference in resolving the puzzles about whether and how fast the exchange rate moves to its PPP level. Nevertheless, misspecifications in the deterministic component of the model in which the unit root or stationary statistics are based can lead to misleading conclusions. The analysis presented above highlights these features in a more formal way. Overall, the differences between our results and those in previous studies are partly attributable to conceptual as well as methodological improvements.

[16] Bergin et al. (2004) suggest that the Balassa-Samuelson effect may have been variable over time, perhaps due to variations in relative productivity differentials themselves, or other factors. For a graphical illustration of this point, see Figure 3 in Taylor and Taylor (2004).

4.3 Distance

As a by-product of our analysis, we have estimated regression models where the HL estimates are explained in terms of different functions of the (great circle) geographical distance. These sorts of equations were also estimated in CMS for the alternative measures of persistence that they compute. The estimation results of these equations are reported in Table 12. The analysis has removed two observations that have been characterized as outliers, and none of the observations have shown as influent. Further, we have computed the White's robust estimates for the standard errors.

As can be seen, the F-test of joint statistical significance shows that the slope parameters are not significant. These results are also obtained when computing the bootstrap confidence intervals for the estimated parameters. Therefore, the distance is not relevant when explaining the short-lived effects of the shocks that we have estimated. This indicates that price differentials adjustment is not driven by geographical distance but for the existence of nominal rigidities in the economies of the different US cities.

5 Concluding remarks

We have explicitly shown how structural breaks introduce conceptual and econometric difficulties that complicate the interpretation of PPP, and thereby impacting the computations of the half-life of PPP deviations. The crux of the issue is that structural breaks introduce an upward bias in the autoregressive coefficients of the AR models that have been adjusted to price level differentials, which in turn implies upward bias in the persistence measure of the PPP deviations, falsely leading us to conclude that shocks affecting price differentials are highly persistent. Existing work focuses on the implications of various econometric biases on the computations of the half-life of PPP deviations, they are however silent about the consequence of neglecting structural breaks in the data generating process. The present paper fills this gap.

Several interesting findings emerge from our analysis. Regardless of the parity restrictions, we find overwhelming evidence that price differentials across US cities are I(0) stationary. Once the structural breaks are accounted for, the median half-life of convergence ranged between 1.50 to 2.65 years, much faster than those documented in previous studies. As a check of robustness, we have extended the analysis to conduct a pairwise tests of price level convergence. The finding is broadly consistent with those obtained from the aggregate US price level.

References

Altissimo, F., Bilke, L., Levin, A., Mathä, T. and Mojon, B. (2006). Sectoral and aggregate inflation dynamics in the euro area, *Journal of the European Economic Association* 4, 585-593.

Andrews, D.W.K. and Chen, H.Y. (1994). Approximately median-unbiased estimation of autoregressive models, *Journal of Business and Economic Statistics* 12, 187-204.

Bai, J. and Perron, P. (1998). Estimating and testing linear models with multiple structural changes, *Econometrica* 66, 47-78.

Balassa, B. (1964). The purchasing parity power doctrine: A reappraisal, *Journal of Political Economy* 72, 584-596.

Basher, S.A. and Carrion-i-Silvestre, J.L. (2008). Price level convergence, purchasing power parity and multiple structural breaks in panel data analysis: An application to US cities, *Journal of Time Series Econometrics* 1(1), Article 3.

Bergin, P., Glick, R. and Taylor, A.M. (2004). Productivity, tradability, and the long-run price puzzle, NBER Working Paper No. 10569.

Cassel, G. (1918). Abnormal deviation in international exchanges, *Economic Journal* 28, 413-415.

Carrion-i-Silvestre, J.L., Del Barrio-Castro, T. and López-Bazo, E. (2005). Breaking the panels: An application to the GDP per capita, *Econometrics Journal* 8, 159-175.

Cecchetti, S.G., Mark, N.C. and Sonora, R.J. (2002). Price index convergence among United States cities, *International Economic Review* 43, 1081-1099.

Chen, L.L. and Devereux, J. (2003). What can US city price data tell us about purchasing power parity? *Journal of International Money and Finance* 22, 213-222.

Choi, C-Y., Mark, N. and Sul, D. (2006). Unbiased estimation of the half-life to PPP convergence in panel data, *Journal of Money, Credit and Banking* 38, 921-938.

Engel, C. (1999). Accounting for U.S. real exchange rate changes, *Journal of Political Economy* 107, 507-38.

Gadea, M.D., A. Montañés, and Reyes, M. (2004). The European union currencies and the US dollar: From post-Bretton-Woods to the euro, *Journal of International Money and Finance* 23, 1109-1136.

Hansen, B.E. (2001). The new econometrics of structural change: Dating breaks in US labor productivity, *Journal of Economic Perspective* 15, 117-128.

Imbs, J., Mumtaz, H., Ravn, M.O. and Rey, H. (2005). PPP strikes back: Aggregation and the real exchange rate, *Quarterly Journal of Economics* 120, 1-43.

Kim, J-O. and Enders, W. (1991). Real and monetary causes of real exchange rate movements in the pacific rim, *Southern Economic Journal* 57, 1061-1070.

Kwiatkowski, D., Phillips, P.C.B., Schmidt, P.J. and Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root, *Journal of Econometrics* 54, 159-78.

Levin, A., Lin, C.F. and Chu, J. (2002). Unit root tests in panel data: Asymptotic and finite-sample properties, *Journal of Econometrics* 108, 1-24.

Liu, J., Wu, S. and Zidek, J.V. (1997). On segmented multivariate regressions, *Statistica Sinica* 7, 497-525.

Lopez, C., Murray, C.J. and Papell, D.H. (2005). State of the art unit root tests and purchasing power parity, *Journal of Money, Credit, and Banking* 37, 361-369.

Maddala, G.S. and 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, 631-52.

Nath, H.K. and Sarkar, J. (2007). Unbiased estimation of the half-life to price index convergence among US cities, forthcoming in *Journal of Money, Credit and Banking.*

Ng, S. and Perron, P. (1995). Unit root tests in ARMA models with data dependent methods for the selection of the truncation lag, *Journal of the American Statistical Association* 90, 268-281.

Ng, S. (2006). Testing cross-section correlation in panel data using spacings, *Journal of Business and Economic Statistics* 24, 12-23.

Papell, D.H. and Theodoridis, H. (2001). The choice of numeraire currency in panel tests of purchasing power parity, *Journal of Money, Credit and Banking* 33, 790-803.

Papell, D.H. and Prodan, R. (2006). Additional evidence of long-run purchasing power parity with restricted structural change, *Journal of Money, Credit and Banking* 38, 1329-1349.

Parsley, D. and Wei, S-J. (1996). Convergence to the law of one price without trade barriers or currency fluctuations, *Quarterly Journal of Economics* 61, 1211-1236.

Perron, P. (1989). The great crash, the oil price shock, and the unit root hypothesis, *Econometrica* 57, 1361-1401.

Perron, P., and Qu, Z. (2006). Estimating restricted structural change models, *Journal of Econometrics* 134, 373-399.

Pesaran, M.H. (2004). General diagnostic tests for cross section dependence in panels, *Cambridge Working Papers in Economics*, No. 435, University of Cambridge.

Pesaran, M.H. (2007). A pair-wise approach to testing for output and growth convergence, *Journal of Econometrics* 138, 312-355.

Pesaran, M.H., Smith, R.P., Yamagata, T. and Hvozdyk L. (2007). Pairwise tests of purchasing power parity, forthcoming in *Econometric Reviews.*

Rogoff, K. (1996). The purchasing power parity, *Journal of Economic Literature* 34, 647-668.

Samuelson, P.A. (1964). Theoretical notes on trade problems, *Review of Economics and Statistics* 46, 145-154.

Taylor, A.M. (2001). Potential pitfalls for the purchasing-power-parity puzzle? Sampling and specification biases in mean-reversion tests of the law of one price, *Econometrica* 69, 473-498.

Taylor, A.M. and Taylor, M.P. (2004). The purchasing power parity debate, *Journal of Economic Perspective* 18, 135-158.

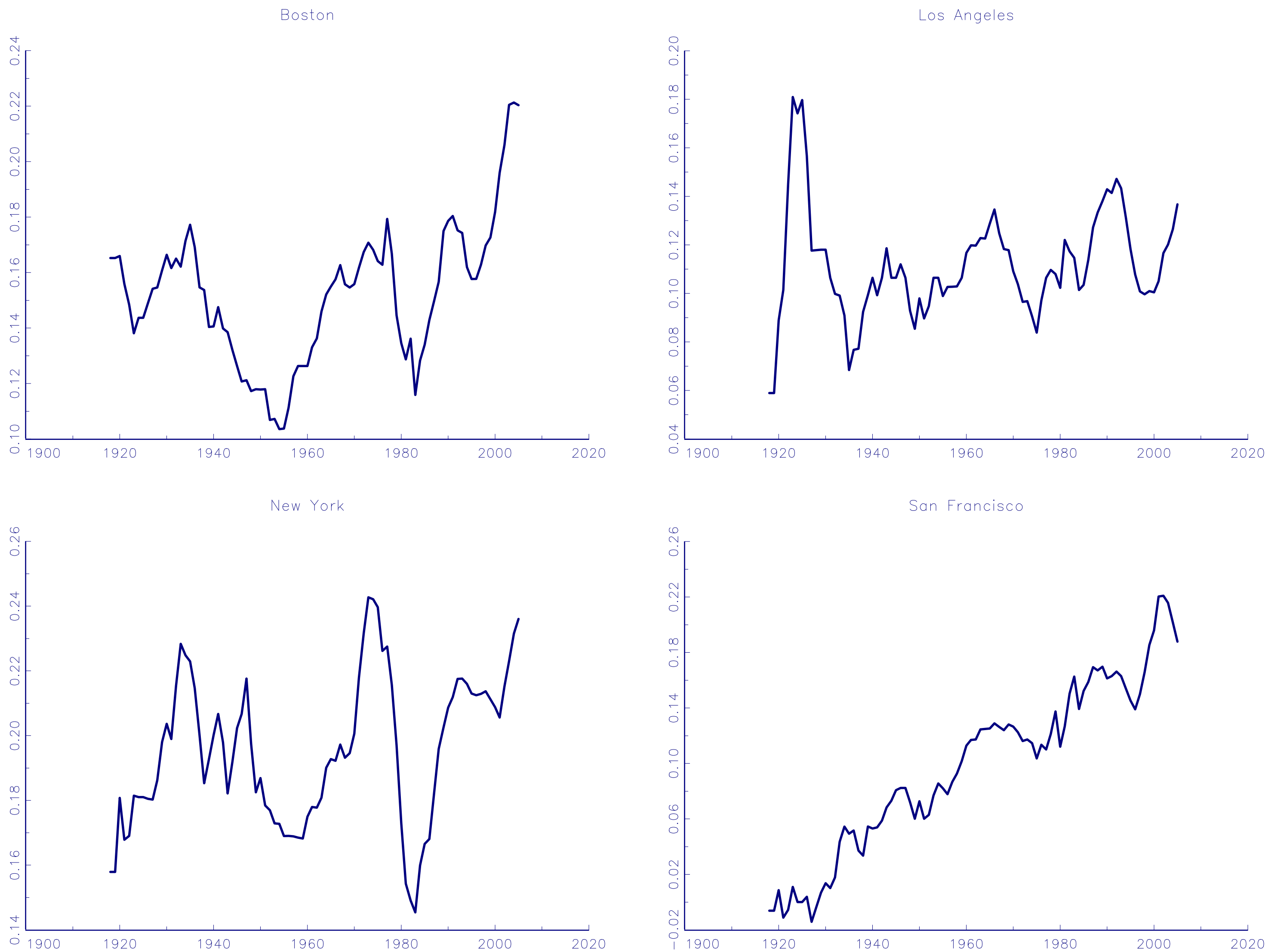

Figure 1. Price differentials against US aggregate price level

Table 1: Individual and panel data stationarity tests with multiple structural breaks: QPPP hypothesis (without parity restrictions)

Panel A: Individual statistics

	$\hat{\eta}_{i,US}\left(\hat{\lambda}_{i,US}\right)$	Critical values 10%	5%	$\hat{T}_{b,1}^{i,US}$	$\hat{T}_{b,2}^{i,US}$	$\hat{T}_{b,3}^{i,US}$	$\hat{T}_{b,4}^{i,US}$
Atlanta	0.090	0.106	0.135	1930	1947	1961	
Boston	0.024	0.072	0.083	1943	1962	1988	
Chicago	0.039	0.056	0.064	1933	1949	1970	1983
Cincinnati	0.050	0.066	0.080	1930	1961	1978	1991
Cleveland	0.069	0.069	0.081	1937	1962	1981	
Detroit	0.097	0.156	0.206	1930	1986		
Houston	0.069	0.064	0.077	1947	1960	1975	1988
Kansas City	0.140	0.071	0.083	1940	1964	1989	
Los Angeles	0.098	0.227	0.296	1986			
Minneapolis	0.034	0.130	0.167	1932	1981		
New York	0.057	0.245	0.315	1989			
Philadelphia	0.065	0.116	0.147	1964	1978	1991	
Pittsburgh	0.033	0.097	0.120	1932	1946	1987	
Portland	0.072	0.196	0.258	1942			
San Francisco	0.145	0.059	0.069	1932	1945	1958	1981
Seattle	0.041	0.091	0.113	1939	1978	1992	
St. Louis	0.042	0.079	0.096	1951	1971	1990	

Panel B: Panel Stationarity Tests

	Independence		CS demeaned		Bootstrap distribution			
	Test	p-val	Test	p-val	90%	95%	97.5%	99%
$Z(\lambda)$ Hom.	0.343	0.366	0.093	0.463	4.646	5.612	6.565	7.729
$Z(\lambda)$ Het.	0.772	0.220	-0.115	0.546	3.439	4.053	4.593	5.243

Note: $\hat{\eta}_i$ refers to the individual KPSS statistic, while $Z(\lambda)$ is the corresponding panel stationary statistics; $\hat{\lambda}_i$ indicates the relative position of the dates of the breaks on the entire time period, T, for each individual i; and $T_{b,k}^i$ denotes the k-th date of the break for i-th individual. Hom.: Homogenous; Het.: Heterogenous. 'Independence' refers to cross section independence; 'CS demeaned' refers to cross section demeaning of Levin et al. (2002); 'Bootstrap distribution' is the Maddala and Wu (1999) parametric bootstrap to accommodate cross section dependence. In the QPPP (Qualified Purchasing Power Parity) specification the deterministic component includes only level shifts, whereas in the TQPPP (Trend Qualified Purchasing Power Parity) specification the deterministic component includes both changes in the level and in the slope of the time trend.

Table 2: Individual and panel data stationarity tests with multiple structural breaks: TQPPP hypothesis (without parity restrictions)

Panel A: Individual statistics

	$\hat{\eta}_{i,US}\left(\hat{\lambda}_{i,US}\right)$	Critical values 10%	5%	$\hat{T}_{b,1}^{i,US}$	$\hat{T}_{b,2}^{i,US}$	$\hat{T}_{b,3}^{i,US}$	$\hat{T}_{b,4}^{i,US}$
Atlanta	0.036	0.028	0.032	1932	1951	1983	
Boston	0.079	0.026	0.029	1936	1956	1978	
Chicago	0.022	0.022	0.024	1932	1947	1960	1979
Cincinnati	0.073	0.041	0.047	1935	1978		
Cleveland	0.029	0.041	0.048	1938	1982		
Detroit	0.040	0.065	0.079	1931	1944		
Houston	0.057	0.021	0.023	1934	1949	1971	1984
Kansas City	0.031	0.041	0.048	1931	1964		
Los Angeles	0.088	0.092	0.113	1930			
Minneapolis	0.031	0.088	0.109	1931			
New York	0.015	0.023	0.026	1931	1958	1971	1984
Philadelphia	0.020	0.031	0.035	1937	1973	1986	
Pittsburgh	0.017	0.040	0.046	1932	1967		
Portland	0.027	0.022	0.024	1931	1946	1970	1983
San Francisco	0.015	0.058	0.068	1971			
Seattle	0.025	0.029	0.034	1937	1971	1985	
St. Louis	0.023	0.054	0.065	1972	1988		

Panel B: Panel Stationarity Tests

	Independence		CS demeaned		Bootstrap distribution			
	Test	p-val	Test	p-val	90%	95%	97.5%	99%
$Z(\lambda)$ Hom.	2.203	0.000	3.319	0.001	10.084	12.197	14.400	17.339
$Z(\lambda)$ Het.	3.325	0.000	4.138	0.000	5.896	6.515	7.148	7.937

Note: See Table 1.

Table 3: Cross-section correlation tests

	constant with level shift	time trend with level and slope shifts
CD Statistic	2.90 (0.002)	2.376 (0.009)
$\hat{\eta}$	62	102
$svr^W(\hat{\eta})$	3.362 (0.000)	2.909 (0.002)
$svr^L(\hat{\eta})$	1.239 (0.108)	0.530 (0.298)
$svr^S(\hat{\eta})$	0.876 (0.191)	1.353 (0.088)

Note: All the statistics in the table specify the null hypothesis of cross-section independence. CD refers to the test statistic proposed in Pesaran (2004). $svr^W(\hat{\eta})$, $svr^L(\hat{\eta})$, and $svr^S(\hat{\eta})$ report Ng's (2006) spacing variance ratio for the whole, large, and small sample, respectively. P-values between parentheses. $\hat{\eta}$ indicates the estimated break point in the sample, which splits the sample of correlations in two groups (small and large correlations). We use the $t - sig$ criterion in Ng and Perron (1995) to select the order of the autoregressive correction with up to ten lags.

Table 4: Panel data stationarity tests with multiple structural breaks: Mixture of the QPPP and TQPPP hypothesis specifications (with and without parity restrictions)

Panel A: Without parity restrictions

	Independence		CS demeaned		Bootstrap distribution			
	Test	p-val	Test	p-val	90%	95%	97.5%	99%
$Z(\lambda)$ Hom.	1.263	0.103	2.149	0.016	9.492	10.894	12.242	13.944
$Z(\lambda)$ Het.	1.991	0.023	2.869	0.002	5.628	6.283	6.882	7.569

Panel B: With parity restrictions

	Independence		CS demeaned		Bootstrap distribution			
	Test	p-val	Test	p-val	90%	95%	97.5%	99%
$Z(\lambda)$ Hom.	11.662	0.000	29.684	0.000	10.611	11.875	12.963	14.457
$Z(\lambda)$ Het.	3.945	0.000	4.392	0.000	3.697	4.308	4.919	5.678

Note: See Table 1.

Table 5: Individual and panel data stationarity tests with multiple structural breaks: QPPP hypothesis (with parity restrictions)

Panel A: Individual statistics

	$\hat{\eta}_{i,US}\left(\hat{\lambda}_{i,US}\right)$	Critical values 10%	5%	$\hat{T}_{b,1}^{i,US}$	$\hat{T}_{b,2}^{i,US}$	$\hat{T}_{b,3}^{i,US}$	$\hat{T}_{b,4}^{i,US}$
Atlanta	1.867	0.362	0.489	1930	1949	1958	
Boston	0.098	0.380	0.519	1938	1962	1988	
Chicago	0.378	0.281	0.385	1933	1948	1971	1983
Cincinnati	0.112	0.231	0.312	1928	1961	1978	1991
Cleveland	0.343	0.247	0.328	1938	1982	1996	
Detroit	0.695	0.599	0.828	1937	1983		
Houston	0.323	0.234	0.317	1926	1938	1976	1986
Kansas City	0.669	0.430	0.581	1931	1940	1996	
Los Angeles							
Minneapolis	0.339	0.519	0.696	1926	1977		
New York							
Philadelphia	0.032	0.272	0.354	1969	1978	1987	
Pittsburgh	0.181	0.459	0.620	1932	1946	1987	
Portland							
San Francisco	0.571	0.198	0.267	1926	1956	1981	1996
Seattle	0.673	0.257	0.340	1942	1978	1996	
St. Louis	0.230	0.284	0.380	1926	1972	1990	

Panel B: Panel Stationarity Tests

	Independence		CS demeaned		Bootstrap distribution			
	Test	p-val	Test	p-val	90%	95%	97.5%	99%
$Z(\lambda)$ Hom.	13.470	0.000	13.120	0.000	3.575	5.010	6.418	8.115
$Z(\lambda)$ Het.	11.805	0.000	8.523	0.000	3.405	4.436	5.309	6.573

Note: See Table 1.

Table 6: Individual and panel data stationarity tests with multiple structural breaks: TQPPP hypothesis (with parity restrictions)

Panel A: Individual statistics

	$\hat{\eta}_{i,US}\left(\hat{\lambda}_{i,US}\right)$	Critical values 10%	Critical values 5%	$\hat{T}_{b,1}^{i,US}$	$\hat{T}_{b,2}^{i,US}$	$\hat{T}_{b,3}^{i,US}$	$\hat{T}_{b,4}^{i,US}$
Atlanta	0.124	0.044	0.053	1927	1951	1981	
Boston	0.070	0.048	0.058	1929	1956	1977	
Chicago	0.104	0.043	0.052	1933	1947	1958	1979
Cincinnati	0.055	0.069	0.082	1928	1991		
Cleveland	0.048	0.063	0.075	1938	1980		
Detroit	0.046	0.113	0.142	1931	1941		
Houston	0.068	0.032	0.037	1926	1949	1975	1986
Kansas City	0.036	0.075	0.092	1926	1964		
Los Angeles							
Minneapolis							
New York	0.057	0.036	0.042	1931	1959	1974	1984
Philadelphia	0.040	0.048	0.056	1938	1978	1987	
Pittsburgh	0.072	0.072	0.087	1932	1967		
Portland	0.062	0.037	0.044	1926	1942	1971	1982
San Francisco							
Seattle	0.097	0.048	0.057	1939	1971	1983	
St. Louis	0.063	0.070	0.084	1926	1990		

Panel B: Panel Stationarity Tests

	Independence		CS demeaned		Bootstrap distribution			
	Test	p-val	Test	p-val	90%	95%	97.5%	99%
$Z(\lambda)$ Hom.	5.210	0.000	4.042	0.000	5.782	6.640	7.432	8.448
$Z(\lambda)$ Het.	6.195	0.000	4.880	0.000	5.828	6.427	6.999	7.683

Note: See Table 1.

Table 7: Half-life estimates of price level convergence towards US CPI: QPPP and TQPPP hypotheses specification

	Non-restricted	Restricted
Atlanta	1.750	2.201
Boston	3.305	4.481
Chicago	1.760	2.252
Cincinnati	1.166	0.979
Cleveland	4.226	8.312
Detroit	1.502	3.246
Houston	5.721	2.683
Kansas City	1.381	1.080
Los Angeles	1.263	
Minneapolis	1.936	99.964
New York	3.559	3.959
Philadelphia	1.554	1.531
Pittsburgh	1.111	1.260
Portland	1.493	2.604
San Francisco	0.826	567.807
Seattle	1.268	2.624
St. Louis	1.452	7.398
Mean	2.075	44.524
Median	1.502	2.654

Note: QPPP: Qualified Purchasing Power Parity; TQPPP: Trend Qualified Purchasing Power Parity.

Table 8: Proportion of rejections for the pairwise analysis. Non-restricted and restricted QPPP and TQPPP hypotheses.

		Proportion of rejections of the null hypothesis using critical values at the 5% level of significance	10% level of significance
Non-restricted	QPPP	0.132	0.191
	TQPPP	0.353	0.426
	Mixed	0.338	0.404
Restricted	QPPP	0.315	0.518
	TQPPP	0.088	0.212
	Mixed	0.238	0.478

Note: QPPP: Qualified Purchasing Power Parity; TQPPP: Trend Qualified Purchasing Power Parity.

Table 9: Panel data statistics for the pairwise analysis. QPPP and TQPPP cases.

		Panel A: Non-restricted							
		Independence		CS demeaned		Bootstrap distribution			
		Test	p-val	Test	p-val	90%	95%	97.5%	99%
QPPP	$Z(\lambda)$ Hom.	-0.126	0.550	-1.501	0.933	9.573	11.090	12.510	14.242
	$Z(\lambda)$ Het.	5.027	0.000	1.577	0.057	6.695	7.813	8.831	10.082
TQPPP	$Z(\lambda)$ Hom.	11.033	0.000	8.868	0.000	17.263	19.740	22.371	25.803
	$Z(\lambda)$ Het.	9.721	0.000	9.768	0.000	10.646	11.646	12.552	13.649
Mixed	$Z(\lambda)$ Hom.	9.080	0.000	9.981	0.000	12.149	14.164	16.291	19.442
	$Z(\lambda)$ Het.	9.608	0.000	10.273	0.000	8.924	9.862	10.799	12.004
		Panel B: Parity restrictions							
		Independence		CS demeaned		Bootstrap distribution			
		Test	p-val	Test	p-val	90%	95%	97.5%	99%
QPPP	$Z^*(\lambda)$ Hom.	29.792	0.000	29.187	0.000	8.620	11.984	15.544	20.004
	$Z^*(\lambda)$ Het.	27.367	0.000	28.233	0.000	8.010	10.575	13.143	16.861
TQPPP	$Z^*(\lambda)$ Hom.	16.922	0.000	22.753	0.000	15.026	16.623	18.155	19.883
	$Z^*(\lambda)$ Het.	19.762	0.000	23.075	0.000	14.636	15.770	16.871	18.276
Mixed	$Z^*(\lambda)$ Hom.	45.394	0.000	39.327	0.000	53.025	54.480	55.768	57.299
	$Z^*(\lambda)$ Het.	15.278	0.000	20.136	0.000	13.222	13.832	14.412	15.162

Note: Table 1

Table 10: HL estimates for the pairwise analysis. QPPP and TQPPP hypotheses specifications

	HL		HL		HL		HL
ATL-BOS	7.314	CHI-PIT	1.388	DET-SL	1.280	NY-SL	1.751
ATL-CHI	1.502	CHI-POR	1.976	HOU-KAN	1.352	PHI-PIT	2.075
ATL-CIN	1.037	CHI-SF	1.456	HOU-LA	2.253	PHI-POR	1.705
ATL-CLE	4.115	CHI-SEA	3.049	HOU-MIN	1.391	PHI-SF	1.915
ATL-DET	1.456	CHI-SL	1.496	HOU-NY	3.015	PHI-SEA	1.938
ATL-HOU	1.767	CIN-CLE	1.157	HOU-PHI	2.528	PHI-SL	1.907
ATL-KAN	1.031	CIN-DET	1.388	HOU-PIT	2.180	PIT-POR	1.757
ATL-LA	1.548	CIN-HOU	4.032	HOU-POR	2.215	PIT-SF	1.257
ATL-MIN	2.154	CIN-KAN	1.611	HOU-SF	1.598	PIT-SEA	1.465
ATL-NY	3.840	CIN-LA	2.647	HOU-SEA	6.672	PIT-SL	0.499
ATL-PHI	2.007	CIN-MIN	2.476	HOU-SL	1.903	POR-SF	1.561
ATL-PIT	1.504	CIN-NY	2.762	KAN-LA	1.403	POR-SEA	2.110
ATL-POR	2.214	CIN-PHI	1.759	KAN-MIN	0.787	POR-SL	1.479
ATL-SF	1.634	CIN-PIT	4.294	KAN-NY	0.973	SF-SEA	1.694
ATL-SEA	2.900	CIN-POR	2.347	KAN-PHI	1.322	SF-SL	0.765
ATL-SL	16.348	CIN-SF	297.389	KAN-PIT	1.102	SEA-SL	1.954
BOS-CHI	3.069	CIN-SEA	1.534	KAN-POR	2.772		
BOS-CIN	121.025	CIN-SL	1.834	KAN-SF	1.247		
BOS-CLE	3.521	CLE-DET	1.151	KAN-SEA	8.913		
BOS-DET	2.937	CLE-HOU	4.760	KAN-SL	0.798		
BOS-HOU	3.503	CLE-KAN	2.383	LA-MIN	1.437		
BOS-KAN	1.162	CLE-LA	1.912	LA-NY	2.211		
BOS-LA	2.490	CLE-MIN	1.818	LA-PHI	1.402		
BOS-MIN	2.332	CLE-NY	4.692	LA-PIT	1.257		
BOS-NY	1.705	CLE-PHI	1.549	LA-POR	2.371		
BOS-PHI	12.380	CLE-PIT	5.243	LA-SF	2.010		
BOS-PIT	2.545	CLE-POR	1.610	LA-SEA	2.855		
BOS-POR	1.736	CLE-SF	5.801	LA-SL	1.784		
BOS-SF	3.029	CLE-SEA	1.412	MIN-NY	1.757		
BOS-SEA	2.378	CLE-SL	1.568	MIN-PHI	1.247		
BOS-SL	1.722	DET-HOU	1.437	MIN-PIT	1.926		
CHI-CIN	0.918	DET-KAN	2.098	MIN-POR	1.391		
CHI-CLE	9.554	DET-LA	2.143	MIN-SF	2.670		
CHI-DET	1.677	DET-MIN	1.265	MIN-SEA	1.524		
CHI-HOU	2.463	DET-NY	2.069	MIN-SL	1.009		
CHI-KAN	1.912	DET-PHI	1.730	NY-PHI	1.326		
CHI-LA	1.701	DET-PIT	1.232	NY-PIT	3.796		
CHI-MIN	1.537	DET-POR	1.408	NY-POR	1.905		
CHI-NY	3.909	DET-SF	1.422	NY-SF	1.579		
CHI-PHI	1.617	DET-SEA	1.673	NY-SEA	1.764		

	HL
Mean	5.419
Median	1.766

Note: ATL (Atlanta), BOS (Boston), CHI (Chicago), CIN (Cincinnati), CLE (Cleveland), DET (Detroit), HOU (Houston), KAN (Kansas City), LA (Los Angeles), MIN (Minneapolis), NY (New York), PHI (Philadelphia), PIT (Pittsburgh), POR (Portland), SF (San Francisco), SEA (Seattle), and SL (St. Louis). HL refers to half-life.

Table 11: Proportion of pairwise half-life that are below, within and above the Rogoff's (1996) 3-5 years consensus view

	Non-restricted specifications		
	$HL < 3$	$3 \leq HL \leq 5$	$5 < HL$
QPPP	67.6%	15.4%	16.9%
TQPPP	79.7%	6.8%	13.5%
Mixed	82.4%	10.3%	7.4%
	Restricted specifications		
	$HL < 3$	$3 \leq HL \leq 5$	$5 < HL$
PPP	41.9%	12.2%	45.9%
TPPP	61.2%	17.5%	21.4%
Mixed	59.1%	13.9%	27%

Note: QPPP: Qualified Purchasing Power Parity; TQPPP: Trend Qualified Purchasing Power Parity. HL: Half-life.

Table 12: HL estimates versus distance. Robust estimates

	(1)	(2)	(3)	(4)
ln(distance)	-0.149	1.493		
	(0.223)	(2.330)		
ln(distance)^2		-0.117		
		(0.167)		
ln(ln(distance))			-0.955	23.900
			(1.524)	(28.112)
ln(ln(distance))^2				-6.473
				(7.407)
Intercept	3.446	-2.209	4.254	-19.506
	(1.700)	(8.055)	(3.091)	(26.553)
Obs	134	134	134	134
R^2	0.003	0.005	0.003	0.006
F test	0.448	0.445	0.393	0.477

Note: Robust standard errors between parentheses.