PERSISTENCE IN LAW-OF-ONE-PRICE DEVIATIONS: EVIDENCE FROM MICRO-DATA

by

Mario J. Crucini and Mototsugu Shintani

Working Paper No. 02-W22R

December 2002
Revised July 2004

DEPARTMENT OF ECONOMICS
VANDERBILT UNIVERSITY
NASHVILLE, TN 37235

www.vanderbilt.edu/econ

Persistence in Law-of-One-Price Deviations: Evidence from Micro-data

Mario J. Crucini and Mototsugu Shintani*

Original version: December 2001
Latest version: July 2004

Abstract

We study the dynamics of good-by-good real exchange rates using a micro-panel of 270 goods prices drawn from major cities in 71 countries and 245 goods prices drawn from 13 major U.S. cities. We find half-lives of deviations from the Law-of-One-Price for the average good is about 1 year; somewhat lower for U.S. cities and somewhat higher for cities in the OECD with LDC cities in between. This speed of adjustment is well below the concensus range of estimates of 3 to 5 years for purchasing power parity deviations yet consistent with plausible 'price-stickiness.' We further construct price indices using our micro data and find that aggregation bias combined with small sample bias accounts for a large part of the difference between micro and macro estimates for the OECD.

*Department of Economics, Vanderbilt University. Crucini gratefully acknowledges the financial support of the National Science Foundation (SES-0136979) and the able research assistance provided by Inkoo Lee under the grant. We are especially thankful for detailed and constructive comments provided by Charles Engel, Yanqin Fan, David Parsley, John Rogers, Andy Rose and Randy Verbrugge. We enjoyed productive interaction with seminar and conference participants at the NBER Summer Institute (2004), International Monetary Fund, Louisiana State University, the University of Houston, the University of Virginia, the University of Wisconsin, University of Tokyo, the 2002 Midwest Macroeconomics Conference in Nashville, the 2002 North American Summer Meeting of Econometric Society in Los Angeles and the 2003 ASSA Meetings in San Diego.

1. Introduction

Our research is motivated by the well-known and puzzling behavior of aggregate real exchange rates following the breakdown of Bretton Woods in the early 1970s. In a paper which sparked a flurry of empirical work, Mussa (1986) examined the time series properties of bilateral nominal and real exchange rates before and shortly after the collapse of Bretton Woods. He found that real and nominal exchange rates had comparable volatility and moved very closely with each other after 1973. An enduring interpretation of Mussa's findings is that prices are sticky in domestic currency units.

Following more than 15 years of intense exploration and debate, however, a broad consensus has emerged which places the half-lives of deviations from purchasing power parity (PPP) in the neighborhood of 3 to 5 years (see Rogoff (1996)). For example, Frankel and Rose (1996) utilize a panel of 150 countries and obtain strong evidence of mean-reversion of aggregate real exchange rates with a half-life of about four years. Murray and Papell (2003) conduct bias-corrections using similar data and arrive at a point estimate of 4.6 years. The lower and upper confidence bounds are 2.7 and 5.3 years, which nicely embraces the so-called consensus range.

Even the lower bound of this consensus range is longer than what economists believe is consistent with plausible lengths of price rigidity. Moreover, a limited but growing number of studies on the speed of adjustment for individual goods reinforces this belief. For example, a prominent recent study investigating within-country price adjustment at the level of individual goods by Bils and Klenow (2004) finds that the median good underlying the U.S. CPI experiences a price change every 4.3 months, while Blinder et al. (1998) use firm-level surveys and find that the median firm changes its price once a year. Assuming that prices are being adjusted close to their flexible price equilibrium following the change, these studies place the half-life of deviations at between 2 to 6 months.

Most existing international micro-evidence comes from a collection of studies that focus on either an individual good or a subset of goods within the CPI-basket. For example, Goldberg and Verboven (2004) estimate half-lives of relative price deviations for automobiles in the range of 1.3 to 1.6 years; they focus on Europe. Cumby (1997) finds the half-life of international price deviations in Big Mac hamburgers to be about 1 year. While not their primary focus, Haskel and Wolf (2001) provide some evidence of convergence in international prices of IKEA furniture; pooling all their data, the estimated autoregressive coefficient is 0.11.[1]

Taken together, the microeconomic and macroeconomic evidence presents an empirical puzzle: How do we reconcile the seemingly long half-lives of deviations from PPP with the apparent high speed with which prices adjust at the level of individual goods? To adequately address this question requires a sufficiently large cross-section of goods to capture what the aggregate CPI basket represents and a sufficiently large panel to assure statistical reliability of the results. Much of the existing work has been very limited in coverage of goods and somewhat limited in coverage of locations (i.e., focusing on Europe). Moreover, going from estimates of individual price dynamics to that of an aggregate CPI is not merely a matter of examining the persistence of the median good: as we shall see aggregation bias plays an important role.

This paper uses a recently released world-wide retail price survey conducted

[1] Related studies in this vien are: wheat, butter and charcoal (Froot, Kim and Rogoff (1995)), the Economist Magazine (Ghosh and Wolf (1994)). More comprehensive coverage of goods is found in Crownover, Pippenger and Steigerwald (1996), Isard (1977), Giovannini (1988), Engel (1993), Engel and Rogers (1996), and Rogers and Jenkins (1995).

by the Economist Intelligence Unit to overcome some of the data limitations of existing work and directly answer the question posed above. Our data consists of an 11 year micro-panel of 270 goods prices drawn from major cities in 71 countries and 245 goods prices drawn from 13 major U.S. cities.

When we estimate a dynamic panel model by pooling all international cities, the Law-of-One-Price deviation of an average good in our sample has a half-life of about 1 year. Some notable differences across goods and across groups of countries are also observed. Pooling all international locations, the average half-life for traded goods is about 11 months (0.9 years), while that for non-traded goods is about 19 months (1.6 years). Persistence is somewhat higher for the OECD group than the LDC group. The half-life for the average good when we restrict the country sample to the OECD is close to the world average, at 13 months (1.1 years), compared to just 10 months (0.8 years) for the LDC. Going from a traded good in the LDC to a non-traded good in the OECD, the half-life doubles from 9 months (0.8 years) to 20 months (1.6 years). Not surprisingly, the adjustment of price deviations across U.S. cities turns out to be even faster than what is found internationally.

Taken at face value, these estimates of half-lives of international (and intranational) price deviations at the level of individual goods and services seem consistent the existing microeconomics literature on the frequency of price adjustment with countries and the limited evidence on micro-price dynamics internationally. But how do we go from half-lives of in the neighborhood of 1 year at the level of individual goods to half-lives of CPI-based PPP deviations in the range of 3 to 5 years? The answer we present in this paper is a combination of small sample bias and aggregation bias.

Our baseline results are based on a short annual panel of 11 years and thus, in estimating the persistence parameter, we must rely upon asymptotic theory in the direction of locations to compensate for the limited time dimension. While our estimator of persistence for a particular good is asymptotically unbiased as the number of locations approaches infinity, the fact is we have a relatively limited number of locations: the median number of cities is 54 for our international estimates and as low as 13 for the intranational U.S. case. This leaves room for small sample bias not from the finite time series dimension, but from the finite cross-sectional dimension. To address this issue, we employ a Monte Carlo procedure, allowing key parameters of the data-generating-process to differ good-by-good. We use the output of the Monte Carlo procedure to adjust our baseline estimates for each individual good. While this small sample bias adjustment elevates the persistence estimates, they remain below what is found in the PPP literature. For all international cities, the half-life increases from 1 year to 18 months (1.5 years). When focusing on OECD countries, the group that has received the most empirical study in the existing literature, the half-life rises from 13 months (1.1 years) to 27 months (2.2 years). Thus the half-life of the typical good is still below the bottom of the consensus range in the PPP literature.

We turn, next, to the effect of aggregation in persistence estimates. For the purpose of obtaining an aggregate price index which resembles the actual CPI we utilize consumption expenditure data and take expenditure weighted averages of individual prices. Such an index is constructed for sub-groups of countries, namely the OECD, LDC and U.S. We find that these aggregate real exchange rates do indeed have longer half-lives than the average good used in their construction. Combining this aggregation bias with finite sample bias correction, we arrive at an estimated CPI persistence from our micro-data that equals 5 years for the OECD and 2.9 years for the LDC. This finding suggests that at least part of the answer to the PPP puzzle lies in the difference between price dynamics at the level of individual goods and price dynamics of an aggregate price index. Our results are consistent, qualitatively with a recent study by Imbs et al. (2002) who find that disaggregated price indices display less persistence than CPI indices across European countries.

2. The data

The source of our micro-data on retail prices is the *Worldwide Cost of Living Survey* coordinated and compiled by the Economist Intelligence Unit. The target market for this data source is corporate human resource managers who find it useful in determining compensation levels of employees residing in different cities of the world. While the goods and services reflect this objective to some extent (one obvious example is foreign language instruction of children), the sample is extensive enough to overlap significantly with what would appear in a typical urban consumption basket.

What makes the data particularly attractive for research purposes is the fact that the prices are in absolute, domestic currency, terms and the survey is conducted by a single agency in a consistent manner over time. Other important features are the significant number of cities surveyed, including multiple cities in same country. The maximum number of goods and services priced in any given year is 301 and the maximum number of cities surveyed is 122. The cities are located in 78 countries. The greater number of cities than countries reflects the fact that in a few countries price data is gathered for more than one city. We conduct our international analysis using one city from each country (though which city we choose varies in a minor way across goods for reasons we describe below). We chose the continental U.S. for our intranational analysis for the simple reason that it contains by far the largest number of cities surveyed at 13, compared to the next largest number of cities surveyed which equals 5 in Australia, China and Germany.

In our dynamic estimation we pool our data across locations and time and run a separate regression for each good. Since the raw data contain a number of missing observations and we want to work with balanced panels, we select goods and locations in the following way. First, if the country underwent a currency reform we eliminate it from the sample.[2] Second, for each good, cities that contain missing observations are removed. In selecting the city to use when more than one city is available our default choice is the city that comes first alphabetically. When the price observations are not available for the first city in the alphabet, we move on down through the alphabet until we either find a price observation or exhaust the available cities in that particular country.

After applying these selection rules, the number of cities utilized for an individual good ranges from 22 to 62 (the median number of cities used is 54). In the U.S. context, the number of cities utilized for an individual good ranges from 10 to 13 (13 is the total number available). The median (after rounding) is also 13, reflecting the fact that the U.S. panel contains very few missing observations.

Table 1 presents the 90 cities from 71 countries used in our study, organized into seven geographic regions ordered as follows: Asia, Africa, Europe, South America, Oceania, Central America, and North America. Following these regions, are the 13 U.S. cities used for the intranational analysis. The number of goods for which a particular city is used in the panel estimation is noted in parentheses. Note that in most cases the city coming first in the alphabet is used for most

[2] In an earlier version of this paper we included countries that had experienced currency reforms and the results are similar to those reported here. See Crucini and Shintani (2002) for comparisons.

goods. For example, Berlin is used as the German city in our analysis for 265 goods and Dusseldorf is used for the remaining 5 goods.

Table 2 presents the descriptions of 154 individual goods and services used in our analysis. We actually utilize 270 prices, but 116 of the 154 items are priced in two different types of retail outlets. To conserve space in the table we record the description only once in these cases and make a notation in the table indicating the items for which two price observations are available. What we classify as non-traded goods are indicated in bold-face.

Before turning to the estimation of a dynamic model of price adjustment, we present some basic features of our data. Throughout, we transform the raw price data by subtracting the logarithm of the common currency price of a good from the geometric world average price of the same good:

$$q_{it}^{j} = \ln P_{it}^{j} - N^{-1} \sum_{i=1}^{N} \ln P_{it}^{j} \tag{2.1}$$

where P_{it}^{j} is the U.S. dollar price of good j in city i in period t and N is the number of cities for which prices are available for good j. The normalization to the average world price avoids an arbitrary choice of numeraire location.[3] We refer to the q_{it}^{j}'s as 'Law-of-One-Price deviations': when multiplied by 100, they represent the percentage by which the price of an individual good in a particular location at a particular date differs from the average world price of the same good at the same date. The Law-of-One-Price proposition predicts that in the absence of natural or official barriers to trade: $q_{it}^{j} = 0$ for all i, j and t.

One way to summarize how our data line up with this prediction is to plot

[3] We utilize the deviations from the geometric mean because the definition fits well to the retail price model we use. However, while not reported here, we also conducted all the statistical analysis using deviations from the arithmetic mean and found that the main results were not sensitive to the choice of normalization.

the empirical distribution of the Law-of-One-Price deviations at a particular date $t = \tau$, $\widehat{F}(q^j_{i\tau})$. If the Law-of-One-Price held exactly across all goods and locations at date $t = \tau$, all of the mass of the distribution would be located at 0. Apart from internationally integrated primary commodity markets (e.g., gold bullion traded on centralized exchanges) we would not expect to come very close to this prediction. More interesting is a comparison of how good an approximation the Law-of-One-Price is for different locations, goods or time periods.

For example, Figure 1 presents the smoothed density estimates of the world and U.S. prices using a Gaussian kernel for benchmark years – 1990, 1995, and 2000 – pooling all goods and locations. The tendency toward the Law-of-One-Price prediction is much more evident across U.S. cities than across cities of the world economy. The visual impression is confirmed by the standard deviations in the data; using data from 2000, the standard deviation of price deviations averages 0.60 internationally, about twice that found across U.S. cities, where the standard deviation is 0.25. We conclude from this that the U.S. retail market is considerably more integrated than the international retail market.

To get a sense of the role of city-specific versus good-specific sources of price variation we average the deviations across goods for each city and data and plot the resulting densities in Figure 2. Effectively, we are computing a crude estimate of the purchasing power parities of the cities in our sample. These densities are much more concentrated around zero reflecting the fact that Law-of-One-Price deviations tend to average out across goods on a country-by-country basis (a property first pointed out by Crucini, Telmer and Zachariadis (2001), in the context of European cities). However, the U.S. density becomes much more concentrated around zero than the world density, suggesting that PPP holds to a much closer approximation across U.S. cities than across cities of the world.

A number of theoretical perspectives predict that we should observe more price dispersion in non-traded goods than in traded goods. Figure 3 elucidates this property of the micro-data, plotting the distribution of Law-of-One-Price deviations in 2000 for traded and non-traded goods; again with separate densities estimated for the U.S. and international price data. Consistent with expectations, non-traded goods exhibit greater geographic price dispersion than traded goods. One surprising feature of the data is the fact that the dispersion of non-traded goods prices is lower across U.S. cities than are the prices of traded goods internationally (their respective standard deviations are 0.32 and 0.53). It is likely that this anomaly would disappear if we concentrate the international analysis on cities with similar wealth variation to that existing across U.S. cities.

The next section focuses on the time series properties of Law-of-One-Price deviations across countries and within the U.S.

3. Dynamics of Law-of-One-Price deviations

3.1. Testing for a unit root in Law-of-One-Price deviations

We begin with tests of unit roots in Law-of-One-Price deviations good-by-good, pooling the data across locations. With a small number of time series observations ($T = 11$) for each real exchange rate, conventional panel unit root tests are not applicable. Instead, we employ a panel unit root test with fixed time dimension as developed by Harris and Tzavalis (1999). Table 3 summarizes the results of the unit root tests. In the international data, we are able to reject the null of a unit root for virtually every good. We reject the null of a unit root for 99% of the goods at the 1% level based on the tests that include a constant. Even with the tests without a constant, we reject the unit root null for 94% of the goods at the 1% level of significance. Somewhat weaker evidence is obtained when tests are applied to OECD and LDC sub-groups.

The null hypothesis of a unit root is also rejected in many cases in the U.S.; the fraction of goods for which the null is rejected is 74% for at 1% level test with a constant. Comparison of the international and intranational results is puzzling, since we would naturally expect stationarity of real exchanges in a highly integrated market with a common currency. Recall, however, that the median number of cities available is 13 for the U.S. in contrast to 54 internationally, thus the results may simply reflect lower power of the test with a small sample.

In summary, our evidence provides overwhelming support for the hypothesis that when a disturbance alters the relative price of a good from its location-specific mean, the deviations are temporary, not permanent. This conclusion holds both within and across countries. The natural question to ask next is: How persistent are these temporary deviations?

3.2. How persistent are Law-of-One-Price deviations?

We consider the convergence of international (and intranational) prices under two modes of convergence, which we refer to as the conditional and absolute convergence hypotheses. The *conditional convergence* hypothesis posits that the common currency price of a good reverts back to a level unique to each location in the long-run. The stronger hypothesis, *absolute convergence*, posits that the common currency price of a good reverts back to a common level in all locations in the long-run. Since the prices studied are final consumer prices of retail goods as opposed to source and destination prices of homogenous traded commodities, we expect conditional convergence to be the rule rather than the exception.

We estimate persistence of the Law-of-One-Price deviations q_{it}^{j} for each good

$j(=1,...,M)$ using the maximum available locations for that good. In what follows, we drop the good index j since all of our estimation is good-by-good pooling across cities $i = 1, ..., N$. We consider the following autoregressive model of order one,

$$q_{it} = \eta_i + \rho q_{it-1} + v_{it} \tag{3.1}$$

where $|\rho| < 1$, η_i is a zero mean time-invariant individual (city) specific effect with variance σ_η^2, and v_{it} is an idiosyncratic shock with mean zero conditional on η_i and lagged q_{it}'s and with variance σ_v^2. The presence of η_i allows the long-run price level of good j to differ location-by-location. To be more specific, $\eta_i/(1-\rho)$ may be viewed as the steady state level in the sense that it is a mean of q_{it} conditional on η_i for large t. Therefore, provided that $\sigma_\eta^2 > 0$, the model is consistent with the conditional convergence hypothesis. Evidently, it reduces to the model of absolute convergence with the restriction $\sigma_\eta^2 = 0$. Our main interest is ρ which represents the speed of adjustment in prices or the persistence of the price deviations. We employ the two-step generalized method of moments (GMM) estimator of ρ based on the first difference transformation,

$$q_{it} - q_{i,t-1} = \rho\left(q_{i,t-1} - q_{i,t-2}\right) + \left(v_{it} - v_{i,t-1}\right) \quad \text{for } t = 3, ..., T, \tag{3.2}$$

with instruments selected from the orthogonality conditions,

$$E\left[q_{is}\left(v_{it} - v_{i,t-1}\right)\right] = 0 \quad \text{for } s = 1, ..., t-2 \text{ and } t = 3, ..., T. \tag{3.3}$$

This choice of instruments, originally proposed by Holtz-Eakin, Newey, and Rosen (1988) and Arellano and Bond (1991), is known to provide a consistent estimator for fixed T and large N under fairly general assumptions. We follow Arellano and Bond's (1991) suggestion for the choice of a weighting matrix in the first step

GMM estimation.

We begin with the persistence of international deviations from the Law-of-One-Price. As reported in the top panel of Table 4, the average persistence parameter across goods in the international data is 0.51, indicating relatively fast adjustment with a half-life of 1 year. For each good, the standard error is small and the confidence interval has an upper bound below unity, consistent with our prior rejection of unit roots. The first and third quartiles are 0.42 and 0.60, respectively. When we focus on the traded goods only, the average persistence reduces to 0.46 with the corresponding half-life being 0.9 years. On the other hand, the average persistence of non-traded goods is 0.65 implying slower adjustment compared to traded goods.

The story remains much the same when we restrict our estimation to sub-panels involving OECD countries. In contrast, the deviations in LDC countries tend to be less persistent than in OECD countries with half-life being less than a year for the average good with persistence estimates averaging 0.44. This finding is consistent with the view that greater volatility nominal exchange rates and more rapid inflation gives rise to faster price adjustment in the LDC countries.[4] Going from a traded good in the LDC to a non-traded good in the OECD, the half-life doubles from 0.8 years to 1.6 years.

Turning to the U.S. city data, the average persistence level is 0.41, considerably lower than the average for the OECD and slightly lower than the average for the LDC. The first and third quartiles are 0.23 and 0.57. While the convergence rates in the U.S. city prices are faster, standard deviations are also large mainly due to limited availability of the locations ($N = 13$).

[4] This faster adjustment in LDCs was also observed in the cross country study of PPP by Cheung and Lai (2000).

The fact that we are using micro data collected across quite distinct markets makes the possibility of measurement error a concern in our estimation. The presence of measurement error in q_{it} produces a correlation between instruments dated $t-2$ and the error in first difference, $v_{it} - v_{i,t-1}$, and thus the moment conditions (3.3) become invalid (see Holtz-Eakin, Newey, and Rosen (1988) for more on this issue). To examine the effect of measurement error on our results, we employ an alternative GMM estimator based on the instruments dated $t-3$ and before. The results based on this measurement-error-robust estimator are provided in the second panel of Table 4.

The distribution of the good-by-good estimates is virtually unaffected in the U.S. case, where we might expect measurement error to be less of a concern. The international estimates do change somewhat in distribution but not much in terms of the mean. The last panel of Table 4 reports the result of Hausman test for the null hypothesis of no measurement error constructed from the difference between the baseline estimates using (3.3) and the measurement-error-robust estimates. For both the international and intranational data, the hypothesis is rejected for only a small fraction of the goods in our sample. Given the insensitivity of the estimates to the choice of the two alternative estimators as well as results of the formal test for measurement errors, we conclude that measurement error is a small part of the story.

3.3. Absolute versus conditional convergence

Before turning to the stronger hypothesis of absolute convergence, it should be noted that all the estimation methods and results discussed so far are valid under both absolute and conditional convergence. However, in the case of absolute convergence, the lack of individual effect, or $\sigma_\eta^2 = 0$, implies $T-1$ additional valid moment conditions, namely, $E\left[q_{it}v_{i,t-1}\right] = 0$ for $t = 2, ..., T$ (see Holtz-Eakin (1988) for more on this issue). Thus we can employ another GMM estimator that incorporates the new moment conditions in addition to (3.3) to estimate the persistence under absolute convergence. The results are provided in Table 5.

When we impose the restriction that prices to converge to a common level across all locations good-by-good, the estimated persistence increases dramatically in both the international and the U.S. data. The persistence of the deviations of average good in the international data increases from 0.51 to 0.91, which translates into a half-life of more than 7 years. The Law-of-One-Price deviations remain slightly less persistent in the LDC than the OECD. The persistence among U.S. cities is lower than that of the international data, yet the half-life of deviation in average goods is as large as 2.9 years.

These persistence estimates are much closer to what we see in the macroeconomic literature than was the case for our conditional convergence estimates. Thus one is tempted to conclude that imposing the restriction of absolute convergence is key to resolving the divergent results between our original estimates and those found in the macroeconomic literature. However, we find this explanation very unlikely, for two reasons.

First, since the difference between the GMM estimates under conditional convergence and the GMM estimates under absolute convergence can be considered as an indication of an individual effect, a Hausman-type test statistic can be used as a formal test for the null hypothesis of absolute convergence. The lower panel of Table 5 reports these test results. We reject the absolute convergence hypothesis resoundingly in the international context: for 86% of the goods we reject at the 1% level of significance. Evidence against the null is considerably weaker in the U.S. case, where the null is rejected for only 7% of the goods. While the U.S.

results are certainly consistent with the view that long-run Law-of-One-Price deviations within a country are smaller than across countries, the finding may also be the result of low power of the test since we have far fewer U.S. cities than international cities. In any case, it seems fair to say that the absolute convergence hypothesis is flagrantly violated across international cities and at best weakly support by the U.S. data. Thus, the observed high persistence in the case of absolute convergence is most likely the result of statistical misspecification; resembling the upward bias that would result when a constant term is omitted from a simple, univariate, auto-regressive model.

The second reason that it is inappropriate to compare absolute convergence estimates to the existing PPP literature is that constant terms are always included in the regression equation when researchers utilize CPI data. The role of the constant term in the PPP literature is different however, reflecting the fact that when index number data are used the level of the absolute deviation from PPP is unknown. Our results using absolute price data indicate that the resting point of most Law-of-One-Price deviations is not zero as implied by the theoretical proposition, even across cities within countries.

While there are many plausible hypotheses for the existence of permanent deviations from the absolute version of the Law-of-One-Price, we have found a simple model of retail price determination a useful conceptual device for organizing our thinking about the importance of individual effects relative to dynamic price adjustment. Specifically, consider a retail firm selling a good or service in a single location i and taking the prices of inputs as given. Non-traded inputs are locally provided while traded inputs are either exported or imported. The solution to the cost minimization problem assuming a Cobb-Douglas production technology (in non-traded and traded inputs) implies the long-run price level is a linear combination of the long-run levels of markups, wages and trade costs across countries (see Crucini and Shintani (2002) for a more complete derivation).

Using the notation of our estimated model, the steady state level of the Law-of-One-Price deviation for a particular good in location i is $\eta_i/(1-\rho)$ and would correspond to the long-run deviation predicted by the retail model:

$$b_i + \alpha w_i + (1-\alpha)\widetilde{q}_i \tag{3.4}$$

where b_i is the long-run level of markup over marginal cost relative to the world mean (in logs), w_i is the long-run level of wage cost (representing the non-traded inputs) relative to the world mean (in logs), $\widetilde{q}_i$ is the relative price of traded input in the long-run, and α is the share of the non-traded inputs in the Cobb-Douglas production. This simple model predicts the presence of individual effects since b_i and w_i are likely to differ across locations even when $\widetilde{q}_i = 0$ holds for the traded inputs.

Our results to this point present us with two surprising findings and a key remaining puzzle. The first surprising result is the rapid convergence of relative prices to their sample means for most goods. Moreover, the speed of relative price adjustment is more similar across and within countries than we expected. The second surprising result is an implication of rapid price adjustment in the context of substantial price dispersion at a point in time: most of the variance in relative prices is accounted for by individual-specific effects (long-run deviations) not the stochastic variation around them.

The puzzle arises from the fact that while our micro-estimates appear in line with direct measures of price rigidities coming from micro studies of the frequency of absolute price adjustment within countries (such as Bils and Klenow (2004)), they are a full 2 years below the minimum of the consensus range coming from the PPP literature (1 years versus 3 years).

Taking our estimates literally, Law-of-One-Price deviations do not convey the substantial price inertia suggested by the existing PPP literature. The goal of the next section is to reconcile these conflicting results and, in a sense, re-establish the link between Purchasing Power Parity and its basic building block – the Law-of-One-Price.

4. Reconciliation with macroeconomic evidence

4.1. Finite sample bias

We pursue two directions in our attempt to reconcile our microeconomic evidence with the existing macroeconomic evidence. The first direction is to look for evidence of small sample bias in our micro-estimates. For a simple within-group estimator (also known as least square dummy variable estimator) of a dynamic panel model, the finite time series observation T is known to cause downward bias, even as the number of cross sectional observation N tends to infinity. Correcting bias caused by finite T in the least-squares persistence estimates of real exchange rates has been conducted in studies by Andrews (1993), Murray and Papell (2002), and Choi, Mark and Sul (2003), among others.[5] In contrast, our baseline results are based on the GMM estimator which is asymptotically unbiased with fixed T. While our estimator does not suffer from the bias caused directly by finite T (as long as N tends to infinity), the fact is we have a finite number of locations for each good. Moreover, the number of cross-sectional observations becomes particularly small when we focus on sub-groups of countries or the U.S.

[5] An earlier version of this paper included the results based on the least square dummy variable estimator in addition to the GMM estimator. Both methods provided very similar persistence estimates. Small sample bias correction was also applied to the least square dummy variable estimator. See Crucini and Shintani (2002).

cities.

To quantify the amount of bias from finite N in our sample, we employ a Monte Carlo procedure. We allow key parameters of the data-generating-process to differ depending on the individual good, and use the corresponding output of the experiment to make a bias correction to the baseline persistence estimate of each individual good. First, note that for a given T, the extent of the bias in the GMM estimator depends on: 1) N – the number of locations in the cross-section, 2) ρ – the magnitude of the true underlying persistence parameter, and iii) σ_η^2/σ_v^2 – the relative importance of the individual effects relative to the "shocks" in the data-generating process. In the Monte Carlo experiment, we set a grid for the persistence parameter ρ spanning 0.2 to 1 and for the variance ratio σ_η^2/σ_v^2 spanning 0.02 to 1.[6] For each cell in this table we obtain GMM estimates of ρ using artificially generated data from (3.3) with $v_{it} \sim$ i.i.d. $N(0,1)$, $\eta_i \sim$ i.i.d. $N(0, \sigma_\eta^2/\sigma_v^2)$, $T = 11$ and a given N. Bias in each cell is obtained by subtracting ρ from the average of GMM estimates across 500 independent runs. We repeat this process for $N = 54$, 23, 13; corresponding to the median number of locations for panels we use in this section (the world, the OECD and LDC, and the U.S.). Figure 4 plots the bias surface for the number of locations set equal to 23 (about the size of the OECD and LDC panels we use below). The base axes are the true persistence and variance ratio of the data-generating-process and the vertical axis is the estimated bias. For a fixed ratio of variance of individual effects to innovation variance, the bias is low at the extremes (zero persistence or high persistence). The variance ratio bends the shape of the function so that the

[6] We consider a finer grid for ρ above the 0.5 mark moving from 10ths to 100ths because of the highly non-linear features of the contour of the bias function. The variance ratio is an evenly spaced grid with increments of 0.02. The upper bound for the variance ratio is selected based on the actual estimates from the data.

maximum bias follows a trough that deepens as individual effects vanish and persistence rises.

Our bias corrected results for individual goods are summarized in Table 6. The upper-panel contains the baseline estimates, which were previously discussed and are reproduced here for comparison purposes. It also includes the summary of good-by-good estimates of the variance ratio σ_η^2/σ_v^2.[7] The lower-panel contains the bias adjusted estimates. We see, first of all, that small sample bias is substantial in all cases but none of the relative rankings of persistence are affected by the bias correction. The half-life of the average good increases from 1 year to 1.5 years. Turning to the OECD, which is the most studied group, the half-life more than doubles from about 1 year to slightly above 2 years; the half-life of deviations for the LDC group also doubles, but from a lower base of 9 months.

For the purpose of investigating the effect of bias correction to different types of goods, the table also reports estimates for traded and non-traded goods separately. We see that non-traded goods have more persistent deviations across the board and bias correction seems to have a larger effect. For the OECD, the half-live of a deviation for non-traded goods is elevated from 1.6 year to 4.6 years. However, for traded goods, the bias-corrected estimates of half-life is still below 2 years. In summary, the bias seems to have a substantial effect but it alone is not sufficient to reconcile the microeconomic and macroeconomic evidence.

4.2. Aggregation bias

In the previous section, we computed the average persistence across goods and compared this value to the persistence estimates coming from macroeconomic

[7] Note that each indiviual effect η_i cannot be consistently estimated under fixed T large N assumption. However, its variance σ_η^2 can be consistently estimated. See Arellano (2003) for the formula of the estimator of σ_η^2 and σ_v^2.

studies using the CPI-based real exchange rate. While this is a useful starting point, the correct comparison to make is between aggregates and aggregates. We do so in this section by constructing aggregate real exchange rates directly from our micro-prices.

Recently, Imbs et al. (2002) and Chen and Engel (2004), focus their attention explicitly on the effect of aggregation on the persistence of PPP deviations across European countries and arrive at substantively different conclusions (from each other). Our approach differs from theirs in that we construct aggregate indices from micro-data while they use official statistical data at different levels of aggregation to infer the role of aggregation bias. Moreover, they focus on Europe, whereas we study the entire OECD and a large group of LDC countries. Unfortunately, we cannot directly resolve the disagreement in their conclusions for Europe for two reasons. First, aggregation bias is not monotonic in the level of aggregation. Thus, even if we find substantial aggregation bias (and we do) in going from individual goods to the aggregate CPI we cannot claim consistency with Imbs et al. (2002) who start with more aggregative data than we do. Second, if we reduce the cross-section of locations to Europe alone, we lose our ability to conduct accurate inference and reliable small sample bias adjustment. Thus we cannot produce results for Europe to compare to these two interesting recent studies.

Due to the level of detail that we are able to study aggregation bias, it is helpful to first review some analytics of aggregation bias. The first type of aggregation bias arises when the aggregate data is a simple, equally-weighted, average of the underlying time series. Since more persistence goods have more variability in an unconditional sense, it turns out that the persistence of the aggregate real exchange rate will be skewed in the direction of these goods. It is important to keep in mind that this does not mean that the estimated persistence of the aggregate CPI is wrong, just that it's persistence becomes unrepresentative of the micro-dynamics in Law-of-One-Price deviations when a small number of goods have highly persistent deviations (in an absolute and relative sense).

To illustrate the above mechanism at work, consider the following simple example. Suppose there were only two prices used to construct the aggregate CPI, and both are first-order autoregressive processes:

$$q_{it}^{j} = \rho^{j} q_{it-1}^{j} + v_{it}^{j}$$

for $j = H, L$, and $\rho^H \geq \rho^L$. A simple computation reveals that the first-order autocorrelation of the aggregate of the individual Law-of-One-Price deviations, $q_{it} = 0.5q_{it}^H + 0.5q_{it}^L$, becomes

$$\rho = \psi\rho^H + (1-\psi)\rho^L = \psi(\rho^H - \rho^L) + \rho^L$$

where $\psi \equiv Var(q_{it}^H)/\left(Var(q_{it}^H) + Var(q_{it}^L)\right)$. Note that if $\rho^H = \rho^L = \rho^*$, then trivially $\rho = \rho^*$. Problems arise when the persistence estimates differ across goods. Suppose we have 100 goods, each with an innovation variance of unity, but 10 of these goods have persistence of 0.98 and 90 have persistence of 0.5. Based on the formula above, the weight of each of the high (low) persistent goods in the aggregate persistence level would be 0.67 (0.33), rather than 0.10 (0.90) implied by their sample proportions.[8] The persistence of the aggregate would equal 0.82, much higher than the simple average persistence levels of the individual Law-of-One-Price deviations, which equals 0.59.

[8]The more persistent price gets a higher weight here because we are assuming that more persistence implies higher unconditional variation. If it were true that less persistence processes had higher innovation variances, that would lessen the impact of pure aggregation bias.

The second issue to address is that expenditure weights differ across goods. Continuing to assume that expenditure shares do not differ across locations, the persistence of the aggregate real exchange rate becomes:

$$\rho = \psi\omega\rho^H + (1-\omega)(1-\psi)\rho^L$$

where ω is the consumption expenditure weight on the first of the two goods. As is obvious from this equation, if goods exhibiting high levels of persistence make up the majority of expenditure, the aggregate level of persistence will be higher than implied by the simple aggregation bias formula.

The third issue arises from the fact that expenditure shares differ across countries. Thus, to some extent, the CPI baskets differ across countries and PPP could be rejected even if the Law-of-One-Price holds exactly good-by-good. A simple computation shows that a bilateral CPI-based (log) real exchange rate q_t (between country 1 and 2) can be decomposed as:

$$q_t = q_t^* + \omega_{12} z_{1t}$$

where q_t^* is the average Law-of-One-Price deviation computed at a common set of weights (here chosen to be those of country 1, namely, ω_1^H and ω_1^L), ω_{12} is the difference in the weights across locations for good H $(= \omega_1^H - \omega_2^H)$ and z_{1t} is the price of good H relative to good L in country 1 $(= \ln P_{1t}^H - \ln P_{1t}^L)$.

Thus the real exchange rate is the sum of two terms. The first, is what we had earlier, the expenditure-weighted average of the underlying Law-of-One-Price deviations. The second term, is proportional to the relative price of the two goods *within* a country. In a more general derivation this term would involve a weighted average of relative price movements with the weights being the difference in expenditure shares across locations.

Basically the issue boils down to the fact that if a country consumes a higher fraction of goods with prices rising at more than the average rate of inflation across all goods, it will experience a rising relative price of its consumption basket (i.e., aggregate real exchange rate) even if the Law-of-One-Price is satisfied for every good, location and time period.

4.3. Aggregate real exchange rate with expenditure shares

The textbook presentation of how a consumer price index is constructed suggests a simple method for moving from Law-of-One-Price analysis to PPP analysis. Armed with consumption expenditure data, one simply takes weighted averages of individual prices and then uses the resulting price indices to define the aggregate real exchange rate. In reality, of course, the construction of the CPI is much more complex and researchers lack the necessary data to actually replicate the procedures of national statistical agencies. Our goal is to approximate the CPI-based aggregate real exchange rate to the extent possible given the available data.

For the international weights we use the International Comparison Project (ICP) consumption expenditure weights. The shares are available for 24 (OECD) countries in 1990 and 115 countries in 1996. Reconciliation of the cities in our international panel and these weights resulted in our including 23 OECD countries using 1990 weights and 26 LDCs using 1996 weights. The number of expenditure categories are 78 for OECD and 26 for LDC.

Our consumption expenditure weights for U.S. cities are obtained from the Bureau of Labor Statistics (BLS).[9] The weights are those used by the BLS at the Primary Sampling Unit (PSU) level, which, fortunately, correspond very closely to cities. Until very recently, the weights utilized by the BLS changed infrequently.

[9] We thank Randy Verbrugge at the BLS for providing us with these data.

Over our sample period of observation there has been one change in expenditure weights, it occurred in 1994 with the weights maintained from that date through 2000, the end of our sample. We chose to use a fixed set of weights over the entire sample period and therefore utilize the weights for 1994 since they overlap with the greatest time span of our price data. The expenditure categories are what the BLS refers to as ELI (Entry Level Items); they are the lowest level of aggregation in terms of expenditure weights and there are approximately 210 of them.

The tedious part of our work involved the reconciliation of each good in our micro-sample with both an ICP expenditure category and a BLS expenditure category. For the OECD and LDC samples we apply a common ICP weight to each good falling within a particular expenditure category; we repeat this process for the U.S. samples using the BLS weights and categories. Note that the BLS are more appropriate weights since they are applicable to cities and therefore correspond to the locations of the price surveys. In contrast, the country weights in the ICP are meant to reflect national consumption patterns, not consumption patterns by urban dwellers.

Even our straightforward method of aggregation (using country and good-specific expenditure shares) obscures the role of good and country-specific weighting in the end results. For this reason, we will build up our CPI-construct in steps to demonstrate the separate roles played by three distinct factors explained in the previous subsection. The first factor is pure aggregation bias in the sense that an equally-weighted average of underlying series will tend to have different persistence properties than the average persistence of the series that are aggregated. The second factor will reflect that fact that some prices get larger weights (i.e., consumption expenditure shares) than others in the CPI-aggregation. The third factor is that consumption weights differ across countries, at the level of individual goods or, more precisely, categories of goods since weights rarely, if ever, extend down to the level of an individual good.

The upper panel of Table 7 contains all of the aggregation results for the OECD, LDC and U.S. groupings, without small sample bias correction. We do not have a world grouping here because we lack consumption expenditure shares for a number of non-OECD countries. For the baseline estimates, we see that the average persistence level across goods is indeed a poor predictor of the persistence of PPP deviations in all cases except the LDC. Moreover, the (aggregation) bias is in the direction we expected based on our assumption that more persistent goods would effectively receive a larger weight (again the LDC is an exception, though the impact is small to begin with). The half-lives are elevated from 13 months (1.1 years) to 22 months (1.8 years) in the OECD case and from 9 months (0.8 years) to 20 months (1.7 years) in the U.S. case.

Accounting for the fact that goods are consumed in different proportions, but maintaining common weights across locations gives us the third row of results. Here the impact is somewhat mixed. The persistence in the international data is elevated in all cases and the LDC Purchasing Power Parity deviations are now more persistent than what the average good conveys. However, the U.S. aggregate (city level) persistence falls significantly. Moving to the aggregation method that most closely approximates the actual CPI (country- and good-specific weights) causes persistence to fall in both the OECD and U.S. case relative to common weighting. Comparing the starting and ending points, the half lives increase from 13 months (1.1 years) to 23 months (1.9 years) in the OECD, from 10 months (0.8 years) to 13 months (1.1 years) in the LDC and from 9 months (0.8 years) to 10 months (0.8 years) for U.S. cities.

Turning to the results of aggregation combined with small sample bias correction shown in the lower panel of Table 7, the half-lives are elevated as we would expect from the bias correction at the level of individual goods. The combination of aggregation bias and small sample bias now places the half-lives well within the consensus estimate of 3 to 5 years.

Whereas the mean half-life for an individual good was elevated by 1 year from 1 to 2 years in the OECD case, the aggregate PPP deviations go from a sample estimate of over 2 years without bias correction to almost exactly 5 years with bias correction. The LDC mean half-life at the level of individual goods was about 9 months without bias correction and increased to 1.6 years after bias correction, whereas the PPP half-life increases from 1 year to 3 years. For the U.S., we see the bias correction moves them closer to the LDC panel than the OECD panel. The baseline estimates tend to display less persistence than the LDC, but following bias adjustment they look very similar. Keep in mind however, that our bias correction is based on the variance ratio σ_η^2/σ_v^2 which may not be precisely estimated with only 13 locations. Thus we are much less certain about U.S. price dynamics than international price dynamics.

The bottom line of the aggregation results, after correcting for small sample bias, is that we have largely reproduced the persistence levels in the existing literature using our CPI constructs from micro-data. Doing so demonstrates the quantitative importance of aggregation bias in our sample of goods, under our aggregation methods. The fact that we do replicate the 3 to 5 year range in the existing literature that uses official CPI data leads us to suspect that aggregation bias is quantitatively important for interpretations of the existing literature too. Our conclusions offer some suggestions for future research.

5. Conclusions

Our analysis began with the startling observation that deviations from the Law-of-One-Price exhibited half-lives of a year or less. This result seemed plausible in light of the rate at which absolute prices adjust in the U.S. (i.e., the evidence obtained by Bils and Klenow (2004) and Blinder et al. (1998)). The result was also robust even if we incorporated the presence of measurement error. This left us with a new puzzle, namely how one goes from Law-of-One-Price deviations of 1 year to CPI half-lives of 3 to 5 years. As a potential answer to this question, we showed that aggregation bias combined with small sample bias could fill the gap. What implications does this have for economic theories of relative price adjustment? We believe there are four main implications.

First, it is patently unrealistic to abstract from the difference between traded and non-traded goods. For each subset of locations or countries that we examine we find absolute deviations from the Law-of-One-Price tend to be smaller for traded goods than for non-traded goods and persistence of price deviations is lower for traded goods than non-traded goods.

Second, we have demonstrated that the notion – reiterated again and again in the existing literature – that PPP holds in the long-run when the aggregate real exchange rate is stationary is simply false. That fact of the matter is, index numbers do not provide the answer to the question: To what levels do PPP deviations converge?

Third, we have introduced a new dimension to the relationship between the magnitude of price deviations and the speed at which prices adjust. Much of the existing literature emphasizes a positive association (e.g., as emphasized in the threshold autoregression models). In fact, in our panel data we see both rapid convergence to large deviations and slow convergence to small deviations. A perfect example of this is a comparison of deviations from the Law-of-One-Price to deviations from PPP. Aggregation tends to increase the persistence of deviations (aggregation bias) whereas deviations from the Law-of-One-Price tend to average out across goods so that PPP tends to hold to a better approximation 'on average.'

Fourth, our advice to quantitative theorists is to take advantage of the information in the micro-data to calibrate and estimate models based on micro-foundations. The aggregate implications of these models should, then, endogenously generate the types of aggregation bias and averaging out of deviations that are found in the micro-data. The existing approach – starting from the persistence found in the aggregate CPI, while also treating all goods identically – most likely leads the quantitative theorist to under-estimate deviations from the Law-of-One-Price (the most common assumption is that they are actually zero in the long-run) and over-estimate their persistence/inertia.

We intend, in future work, to build on the evidence presented here. The most pressing need is to find either higher frequency price surveys or a longer sample period of observation (preferably both). The more aggregated Penn World Tables is one possible source for longer time spans, national statistical agencies are a possible source for higher frequency data. We also hope to augment the intranational data since the limited number of locations combined with the short time sample limited our ability to conduct accurate inference for the U.S. case. Aside from comparisons of international and intranational price dynamics, such data may shed light on the differences in microeconomic and macroeconomic half-lives of real exchange rates across U.S. cities, reported by Parsley and Wei (1996) and Cecchetti, Mark and Sonora (2002) of about 1 year and 9 years, respectively.

Data Appendix

The purpose of this appendix is to describe the procedure used to clean up the raw data.

1. First, we compute annual price changes good-by-good and compare these changes with the aggregate inflation rate. More specifically, if an individual price change relative to overall inflation – $|\ln(P^j_{it+1}/P^j_{it}) - \ln(P_{it+1}/P_{it})|$ – exceeds 1 we flag the observation. Using this criterion we flagged 5,480 price observations, which is amounts to 1.25% of total sample of price observations in EIU database.

2. Next, we examine the role of changes in the currency denomination in accounting for the flagged observations. The table below reports the countries involved, dates of the denomination change, terms of the change and the number of observations involved.

Country	Date of change	currency before →after	Conversion rate	Observations eliminated
Poland	January 1,1995	zloty	10,000	778
Mexico	January 1, 1993	peso	1,000	287
Uruguay	March 1, 1993	peso	1,000	268
Peru	July 1, 1991	inti→nuevo sol	1,000,000	18
Brazil	March 16, 1990	new cruzado→cruzeiro	1	594
	August 1, 1993	cruzeiro→cruzeiro real	1,000	316
	July 1, 1994	cruzeiro real→real	2,750	
Ecuador	March 20, 2000	sucres→USD	25,000	562
	April 30, 2000	dollarization		
Russia	January 1, 1998	ruble	1,000	458
Argentina	1992		10,000	269

The total number of observations affected by these reforms, that also satisfied our criteria was 3,550, accounting for 64.8% of the 5,480 flagged price changes. We do not use the countries above in our analysis.

3. The next step was more tedious, explicitly examining each of the remaining 1,930 price observations for evidence of errors. In doing so, we use the following protocol:

We first searched for evidence of dropped digits by looking for percentage changes in the raw nominal prices (year-to-year) approximating x=10, 100, 1000, 10000, and so forth. After flagging this subset, we took one of the following actions: i) if it is obvious that units have changed we adjust to common units; ii) if it is obvious that leading digit is missing, we enter the digit which is same as the leading digit of previous and subsequent series; and iii) if it is not obvious whether it is a unit change or a digit is missing we do one of two things. When we were convinced that correction should be made, we take average of adjacent observations. When we were not convinced a correction should be made, we leave the data as is (this is the result in most cases of type 3.)

The end result was to make 305 adjustments to the data. Of the 305 adjustments, in 179 cases we change the units or add the missing digit(s) and in 126 cases of obvious outliers (more than 100% change relative to aggregate inflation from period $t-1$ to t with reversion back to a level close to that at date $t-1$) we replaced the period t observation with the average price across periods $t-1$ and $t+1$. The bottom line is that we end up correcting 15.8% of 1,930 flagged observations that remained after dealing with currency denomination issues (item 2., above).

References

[1] Andrews, Donald W. K., 1993, "Exactly median-unbiased estimation of first order autoregressive/unit root models," *Econometrica* 61(1), 139-165.

[2] Arellano, Manuel, 2003, *Panel Data Econometrics*, Oxford: Oxford University Press.

[3] Arellano, Manuel and Stephen Bond, 1991, "Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations," *Review of Economic Studies* 58, 277-297.

[4] Bils, Mark and Peter J. Klenow, 2004, "Some evidence on the importance of sticky prices," *Journal of Political Economy*, forthcoming.

[5] Blinder, Alan S., Elie R. D. Canetti, David E. Lebow and Jeremy B. Rudd, 1998, *Asking about prices: a new approach to understanding price stickiness*, Russell Sage Foundation, New York.

[6] Cecchetti, Stephen, Nelson Mark and Robert Sonora, 2002, "Price index convergence among United States cities," *International Economic Review* 43(4), 1081-1099.

[7] Chen, Shiu-Sheng and Charles Engel, 2004, "Does "aggregation bias" explain the PPP puzzle?"mimeograph, University of Wisconsin.

[8] Cheung, Yin-Wong and Kon S. Lai, 2000, "On cross-country differences in the persistence of real exchange rates," *Journal of International Economics* 50, 375-397.

[9] Choi, Chi-Young, Nelson C. Mark and Donggyu Sul, 2003, "The dominance of downward bias in half-life estimates of PPP deviations," mimeograph, University of New Hampshire.

[10] Crownover, Collin, John Pippenger and Douglas G. Steigerwald, 1996, "Testing for absolute purchasing power parity," *Journal of International Money and Finance* 15, 783-796.

[11] Crucini, Mario J. and Mototsugu Shintani, 2002, "Persistence in Law-of-One-Price deviations: Evidence from micro-data," Vanderbilt University Working Paper No. 02-22.

[12] Crucini, Mario J., Chris I. Telmer and Marios Zachariadis, 2001, "Understanding European real exchange rates," Vanderbilt University Working Paper No. 01-20.

[13] Cumby, Robert E., 1997, "Forecasting exchange rates and relative prices with the hamburger standard: Is what you want what you get with McParity?" mimeo, Georgetown University.

[14] Engel, Charles, 1993, "Real exchange rates and relative prices: An empirical investigation," *Journal of Monetary Economics* 32, 35-50.

[15] Engel, Charles and John H. Rogers, 1996, "How wide is the border?" *American Economic Review* 86, 1112-1125.

[16] Frankel, Jeffrey A. and Andrew K. Rose, 1996, "A panel project on purchasing power parity: Mean reversion within and between countries," *Journal of International Economics* 40, 209-224.

[17] Froot, Kenneth A., Michael Kim and Kenneth Rogoff, 1995, "The law of one price over 700 years," NBER Working Paper No. 5132.

[18] Ghosh, Atish R. and Holger C. Wolf, 1994, "Pricing in international markets: Lessons from the Economist," NBER Working Paper No. 4806.

[19] Giovannini, Alberto, 1988, "Exchange rates and traded goods prices," *Journal of International Economics* 24, 45-68.

[20] Goldberg, Pinelopi K. and Frank Verboven, 2004, "Market integration and convergence to the Law of One Price: evidence from the European car market," *Journal of International Economics* forthcoming.

[21] Harris, Richard D.F. and Elias Tzavalis, 1999, "Inference for unit roots in dynamic panels where the time dimension is fixed," *Journal of Econometrics*, 91, 201-226.

[22] Haskel, Jonathan and Holger C. Wolf, 2001, "The law of one price - a case study," *Scandinavian Journal of Economics* 103(4), 545-558.

[23] Holtz-Eakin, Douglas, 1988, "Testing for individual effects in autoregressive models," *Journal of Econometrics* 39, 297-307.

[24] Holtz-Eakin, Douglas, Whitney Newey and Harvey S. Rosen, 1988, "Estimating vector autoregressions with panel data," *Econometrica* 56(6), 1371-1395.

[25] Imbs, Jean, Haroon Mumtaz, Morten Ravn and Helene Rey, 2002, "PPP strikes back: Aggregation and the real exchange rate," NBER Working Paper 9372.

[26] Isard, Peter, 1977, "How far can we push the Law of One Price," *American Economic Review* 67(3), 942-948.

[27] Murray, Christian J. and David H. Papell, 2002, "The purchasing power parity persistence paradigm," *Journal of International Economics* 56, 1-19.

[28] Murray, Christian J. and David H. Papell, 2003, "Do panels help solve the purchasing power parity puzzle?" mimeograph, University of Houston.

[29] Mussa, Michael, 1986, "Nominal exchange rate regimes and the behavior of real exchange rates: Evidence and implications," *Carnegie-Rochester Conference Series on Public Policy* 25, 117-214.

[30] Parsley, David and Shang-Jin Wei, 1996, "Convergence to the law of one price without trade barriers or currency fluctuations," *Quarterly Journal of Economics* 61, 1211-1236.

[31] Rogers, John H. and Michael Jenkins, 1995, "Haircuts or hysteresis? Sources of movements in real exchange rates," *Journal of International Economics* 38, 339-360.

[32] Rogoff, Kenneth, 1996, "The purchasing power parity puzzle," *Journal of Economic Literature* 34(2), 647-668.

TABLE 1 – LOCATIONS

City, Country (No. of goods)	City, Country (No. of goods)	City, Country (No. of goods)
(1) International Data	Europe [28 cities, 24 nations]	Oceania [7 cities, 2 nations]
[90 cities, 71 nations]	Vienna, Austria (263) *	Adelaide, Australia (251) *
	Brussels, Belgium (263) *	Brisbane, Australia (12) *
Asia [24 cities, 20 nations]	Prague, Czech (188) †	Melbourne, Australia (2) *
Bahrain, Bahrain (230) †	Copenhagen, Denmark (264) *	Perth, Australia (2) *
Dhaka, Bangladesh (133) †	Helsinki, Finland (255) *	Sydney, Australia (2) *
Beijing, China (144)	Lyon, France (261) *	Auckland, New Zealand (257) *
Hong Kong, Hong Kong (242) †	Paris, France (7) *	Wellington, New Zealand (5) *
New Delhi, India (57)	Berlin, Germany (265) *	Central America [4 cities, 4 nations]
Mumbai, India (146)	Dusseldorf, Germany (5) *	San Jose, Costa Rica (230)
Jakarta, Indonesia (183) †	Athens, Greece (247) *	Guatemala City, Guatemala (221)
Tehran, Iran (181) †	Budapest, Hungary (255) †	Mexico City, Mexico (238) †
Tel Aviv, Israel (255) †	Dublin, Ireland (248) *	Panama City, Panama (242) †
Osaka Kobe, Japan (244) *	Milan, Italy (263) *	North America [8 cities, 2 nations]
Tokyo, Japan (7) *	Rome, Italy (5) *	Calgary, Canada (250) *
Amman, Jordan (137) †	Luxembourg, Luxembourg (260) *	Montreal, Canada (15) *
Seoul, Korea (167) †	Amsterdam, Netherlands (260) *	Toronto, Canada (3)
Kuala Lumpur,Malaysia (244)	Oslo, Norway (233) *	Atlanta, USA (249) *
Karachi, Pakistan (192) †	Warsaw, Poland (232) †	Boston, USA (11)
Manila,Philippines (211) †	Lisbon, Portugal (267) *	Chicago, USA (5)
Al Khobar, Saudi Arabia (203)	Bucharest, Romania (1)	Cleveland, USA (3)
Jeddah, Saudi Arabia (17)	Moscow, Russia (116) †	New York, USA (1)
Singapore, Singapore (256) †	Barcelona, Spain (268) *	
Colombo, Sri Lanka (212) †	Stockholm, Sweden (252) *	
Taipei, Taiwan (215)	Geneva, Switzerland (262) *	*(2) US Data* [13 cities]
Bangkok, Thailand (257) †	Zurich, Switzerland (6) *	
Abu Dhabi, UAE (238)	Istanbul, Turkey (253) *	Atlanta (248)
Dubai, UAE (11)	London, UK (261) *	Boston (257)
Africa (10 cities, 10 nations)	Belgrade, Yugoslavia (105)	Chicago (251)
Abidjan, Cote dIvoire (242) †	South America [9 cities, 9 nations]	Cleveland (249)
Cairo, Egypt (197) †	Buenos Aires, Argentina (253) †	Detroit (260)
Nairobi, Kenya (233) †	Sao Paulo, Brazil (255) †	Houston (250)
Tripoli, Libya (51)	Santiago, Chile (257) †	Los Angeles (248)
Casa Blanca, Morocco (199) †	Bogota, Columbia (235)	Miami (253)
Lagos, Nigeria (204) †	Quito, Ecuador (177) †	New York (234)
Dakar, Senegal (197) †	Asuncion, Paraguay (250)	Pittsburgh (235)
Johannesburg, South Africa (253)	Lima, Peru (1)	San Francisco (230)
Tunis, Tunisia (186) †	Monte Video, Uruguay (257) †	Seattle (252)
Harare, Zimbabwe (200) †	Caracas, Venezuela (238) †	Washington DC (255)

Note: Entries are the city in which the price data are collected, the country to which the city belongs and the number of goods in the analysis for which that city is used. * (†)indicates a city used later in our CPI construct for the OECD (LDC) group. We have expenditure weights for all the U.S. cities

TABLE 2 – LIST OF INDIVIDUAL GOODS

Goods-name		Goods-name	
White bread, 1 kg (supermarket)	*	Fresh fish (1 kg) (supermarket)	*
Butter, 500 g (supermarket)	*	Instant coffee (125 g) (supermarket)	*
Margarine, 500g (supermarket)	*	Ground coffee (500 g) (supermarket)	*
White rice, 1 kg (supermarket)	*	Tea bags (25 bags) (supermarket)	*
Spaghetti (1 kg) (supermarket)	*	Cocoa (250 g) (supermarket)	*
Flour, white (1 kg) (supermarket)	*	Drinking chocolate (500 g) (supermarket)	*
Sugar, white (1 kg) (supermarket)	*	Wine, fine quality (700 ml) (supermarket)	*
Cheese, imported (500 g) (supermarket)	*	Beer, local brand (1 l) (supermarket)	*
Cornflakes (375 g) (supermarket)	*	Beer, top quality (330 ml) (supermarket)	*
Milk, pasteurised (1 l) (supermarket)	*	Scotch whisky, six years old (700 ml) (supermarket)	*
Olive oil (1 l) (supermarket)	*	Gin, Gilbey's or equivalent (700 ml) (supermarket)	*
Peanut or corn oil (1 l) (supermarket)	*	Vermouth, Martini & Rossi (1 l) (supermarket)	*
Potatoes (2 kg) (supermarket)	*	Cognac, French VSOP (700 ml) (supermarket)	*
Onions (1 kg) (supermarket)	*	Liqueur, Cointreau (700 ml) (supermarket)	*
Tomatoes (1 kg) (supermarket)	*	Soap (100 g) (supermarket)	*
Carrots (1 kg) (supermarket)	*	Laundry detergent (3 l) (supermarket)	*
Oranges (1 kg) (supermarket)	*	Toilet tissue (two rolls) (supermarket)	*
Apples (1 kg) (supermarket)	*	Dishwashing liquid (750 ml) (supermarket)	*
Lemons (1 kg) (supermarket)	*	Insect-killer spray (330 g) (supermarket)	*
Bananas (1 kg) (supermarket)	*	Light bulbs (two, 60 watts) (supermarket)	*
Lettuce (one) (supermarket)	*	Batteries (two, size D/LR20) (supermarket)	*
Eggs (12) (supermarket)	*	Frying pan (Teflon or good equivalent) (supermarket)	*
Peas, canned (250 g) (supermarket)	*	Electric toaster (for two slices) (supermarket)	*
Tomatoes, canned (250 g) (supermarket)	*	**Laundry (one shirt) (standard high-street outlet)**	*
Peaches, canned (500 g) (supermarket)	*	**Dry cleaning, man's suit (standard high-street outlet)**	*
Sliced pineapples, canned (500 g) (supermarket)	*	**Dry cleaning, woman's dress (standard high-street outlet)**	*
Beef: filet mignon (1 kg) (supermarket)	*	**Dry cleaning, trousers (standard high-street outlet)**	*
Beef: steak, entrecote (1 kg) (supermarket)	*	Aspirins (100 tablets) (supermarket)	*
Beef: stewing, shoulder (1 kg) (supermarket)	*	Razor blades (five pieces) (supermarket)	*
Beef: roast (1 kg) (supermarket)	*	Toothpaste with fluoride (120 g) (supermarket)	*
Beef: ground or minced (1 kg) (supermarket)	*	Facial tissues (box of 100) (supermarket)	*
Veal: chops (1 kg) (supermarket)	*	Hand lotion (125 ml) (supermarket)	*
Veal: fillet (1 kg) (supermarket)	*	Lipstick (deluxe type) (supermarket)	*
Veal: roast (1 kg) (supermarket)	*	**Man's haircut (tips included) (average)**	
Lamb: leg (1 kg) (supermarket)	*	**Woman's cut & blow dry (tips included) (average)**	
Lamb: chops (1 kg) (supermarket)	*	Cigarettes, Marlboro (pack of 20) (supermarket)	*
Lamb: Stewing (1 kg) (supermarket)	*	Cigarettes, local brand (pack of 20) (supermarket)	*
Pork: chops (1 kg) (supermarket)	*	Pipe tobacco (50 g) (average)	
Pork: loin (1 kg) (supermarket)	*	**Telephone and line, monthly rental (average)**	
Ham: whole (1 kg) (supermarket)	*	**Telephone, charge per local call from home (3 mins) (average)**	
Bacon (1 kg) (supermarket)	*	**Electricity, monthly bill (average)**	
Chicken: frozen (1 kg) (supermarket)	*	**Gas, monthly bill (average)**	
Chicken: fresh (1 kg) (supermarket)	*	**Water, monthly bill (average)**	
Frozen fish fingers (1 kg) (supermarket)	*	**Heating oil (100 l) (average)**	

TABLE 2 (CONTINUED)

Goods-name		Goods-name	
Business suit, two piece, medium weight (chain store)	*	**Four best seats at cinema (average)**	
Business shirt, white (chain store)	*	Low priced car (900-1299 cc) (low)	*
Men's shoes, business wear (chain store)	*	Compact car (1300-1799 cc) (low)	*
Mens raincoat, Burberry type (chain store)	*	Family car (1800-2499 cc) (low)	*
Socks, wool mixture (chain store)	*	Deluxe car (2500 cc upwards) (low)	*
Dress, ready to wear, daytime (chain store)	*	**Yearly road tax or registration fee (low)**	*
Women's shoes, town (chain store)	*	**Cost of a tune up (but no major repairs) (low)**	*
Women's cardigan sweater (chain store)	*	**Annual premium for car insurance (low)**	*
Women's raincoat, Burberry type (chain store)	*	Regular unleaded petrol (1 l) (average)	
Tights, panty hose (chain store)	*	**Taxi: initial meter charge (average)**	
Child's jeans (chain store)	*	**Taxi rate per additional kilometre (average)**	
Child's shoes, dresswear (chain store)	*	**Taxi: airport to city centre (average)**	
Child's shoes, sportswear (chain store)	*	**Furnished residential apartment: 1 bedroom (moderate)**	*
Girl's dress (chain store)	*	**Furnished residential apartment: 2 bedroom (moderate)**	*
Boy's jacket, smart (chain store)	*	**Unfurnished residential apartment: 2 bedrooms (moderate)**	*
Boy's dress trousers (chain store)	*	**Unfurnished residential apartment: 3 bedrooms (moderate)**	*
Hourly rate for domestic cleaning help (average)		**Unfurnished residential apartment: 4 bedrooms (moderate)**	*
Maid's monthly wages (full time) (average)		**Furnished residential house: 3 bedrooms (moderate)**	*
Babysitter's rate per hour (average)		**Unfurnished residential house: 3 bedrooms (moderate)**	*
Compact disc album (average)		**Unfurnished residential house: 4 bedrooms (moderate)**	*
Television, colour (66 cm) (average)		**Business trip, typical daily cost**	
Kodak colour film (36 exposures) (average)		**Hilton-type hotel, single room, one night including breakfast (average)**	
Cost of developing 36 colour pictures (average)		**Moderate hotel, single room, one night including breakfast (average)**	
International foreign daily newspaper (average)		**One drink at bar of first class hotel (average)**	
Daily local newspaper (average)		**Two-course meal for two people (average)**	
International weekly news magazine (Time) (average)		**Simple meal for one person (average)**	
Paperback novel (at bookstore) (average)		**Hire car, weekly rate for lowest price classification (average)**	
Three course dinner for four people (average)		**Hire car, weekly rate for moderate price classification (average)**	
Visit of four people to a nightclub (average)		**One good seat at cinema (average)**	
Four best seats at theatre or concert (average)		**Average cost of labour per hour (pay and non-pay costs)**	

Notes: Non-traded goods are indicated by bold fonts. * indicates the good with multiple price observations.

TABLE 3 – PANEL UNIT ROOT TESTS FOR INDIVIDUAL GOODS

	World	OECD	LDC	US
Panel A: Rejection frequencies of test without constant				
10% level	0.98	0.90	0.96	0.94
5% level	0.97	0.82	0.93	0.89
1% level	0.94	0.62	0.83	0.81
Panel B: Rejection frequencies of test with constant				
10% level	0.99	0.96	0.97	0.90
5% level	0.99	0.92	0.97	0.84
1% level	0.99	0.82	0.92	0.74
Number of tests/goods	270	269	263	245
Median number of cities	54	23	22	13

Notes: Based on 11 year panel (1990-2000). The number shows the proportion of goods for which the null hypothesis of unit root is rejected using the panel unit root test of Harris and Tzavalis (1999). The number of tests/goods is the total number of unit root test statistics computed for each good while the median of the number of cities refers to the number of cross-section sample used to compute each test statistic.

TABLE 4 – DYNAMIC PANEL ESTIMATES OF PERSISTENCE FOR INDIVIDUAL GOODS UNDER CONDITIONAL CONVERGENCE

	World	OECD	LDC	US
Panel A: Baseline GMM estimates				
Mean	0.51	0.53	0.44	0.41
	(0.01)	(0.04)	(0.04)	(0.12)
25th percentile	0.42	0.39	0.33	0.23
50th percentile (median)	0.51	0.55	0.44	0.41
75th percentile	0.60	0.67	0.55	0.57
Number of goods	270	269	263	245
Median number of cities	54	23	22	13
Panel B: Measurement error robust GMM estimates				
Mean	0.53	0.51	0.44	0.40
	(0.03)	(0.04)	(0.04)	(0.12)
Panel C: Rejection frequencies of Hausman test for H_0: No measurement error				
10% level	0.09	0.21	0.23	0.07
5% level	0.05	0.15	0.20	0.05
1% level	0.03	0.19	0.14	0.02
Number of tests/goods	255	139	149	117

Notes: Based on 11 year panel (1990-2000). Mean shows the averages of good-specific estimates of ρ. Numbers in parentheses are the averages of the standard errors. The number of goods is the total number of good-specific estimates of ρ while the median of the number of cities refers to the number of cross section sample used in the estimation of each good-specific ρ. The lowest panel contains the proportion of goods for which the null hypothesis of no measurement error is rejected using Hausman test constructed from GMM estimates with and without the measurement error correction.

TABLE 5 – DYNAMIC PANEL ESTIMATES OF PERSISTENCE FOR INDIVIDUAL GOODS UNDER ABSOLUTE CONVERGENCE

	World	OECD	LDC	US
Panel A: GMM estimates without individual effects				
Mean	0.89	0.91	0.87	0.79
	(0.00)	(0.01)	(0.01)	(0.06)
25th percentile	0.86	0.88	0.83	0.70
50th percentile (median)	0.90	0.92	0.89	0.81
75th percentile	0.92	0.95	0.92	0.89
Number of goods	270	269	263	245
Median number of cities	54	23	22	13
Panel B: Rejection frequencies of Hausman test for H_0: Absolute convergence				
10% level	0.98	0.65	0.81	0.18
5% level	0.94	0.55	0.71	0.12
1% level	0.86	0.36	0.55	0.07
Number of tests/goods	262	268	254	245

Notes: See notes to Table 4. The lower panel contains the proportion of goods for which the null hypothesis of no individual effects is rejected using Hausman test constructed from GMM estimates under conditional and absolute convergence.

TABLE 6. – SMALL SAMPLE BIAS CORRECTION FOR INDIVIDUAL GOODS

		World	OECD	LDC
		Panel A: Baseline estimates		
All goods	ρ (SE)	0.50 (0.01)	0.52 (0.04)	0.44 (0.04)
	σ_η^2/σ_v^2	0.54	0.56	0.57
	Half-life	**1.0 yrs.**	**1.1 yrs.**	**0.8 yrs.**
Traded goods	ρ (SE)	0.46 (0.01)	0.49 (0.04)	0.40 (0.04)
	σ_η^2/σ_v^2	0.57	0.59	0.58
	Half-life	**0.9 yrs.**	**1.0 yrs.**	**0.8 yrs.**
Non-traded goods	ρ (SE)	0.65 (0.01)	0.65 (0.03)	0.55 (0.04)
	σ_η^2/σ_v^2	0.42	0.44	0.51
	Half-life	**1.6 yrs.**	**1.6 yrs.**	**1.2 yrs.**
		Panel B: Bias adjusted estimates		
All goods	ρ	0.63	0.73	0.65
	Half-life	**1.5 yrs.**	**2.2 yrs.**	**1.6 yrs.**
Traded goods	ρ	0.57	0.70	0.61
	Half-life	**1.2 yrs.**	**1.9 yrs.**	**1.4 yrs.**
Non-traded goods	ρ	0.84	0.86	0.79
	Half-life	**3.9 yrs.**	**4.6 yrs.**	**2.9 yrs.**

Notes: Based on 11 year panel (1990-2000). ρ and σ_η^2/σ_v^2 shows the averages of good-specific estimates of ρ and σ_η^2/σ_v^2, respectively. Numbers in parentheses are the averages of the standard errors. Half-life is based on the corresponding average of estimated ρ.

TABLE 7 – COMBINED EFFECTS OF AGGREGATION AND SMALL SAMPLE BIAS ON PERSISTENCE ESTIMATES

Weighting method		OECD	LDC	US
Panel A: Baseline estimates				
Average across goods	ρ (SE)	0.52 (0.04)	0.44 (0.04)	0.41 (0.12)
	σ_η^2/σ_v^2	0.56	0.57	0.42
	Half-life	**1.1 yrs.**	**0.8 yrs.**	**0.8 yrs.**
Equal weighting	ρ (SE)	0.68 (0.01)	0.42 (0.01)	0.67 (0.12)
	σ_η^2/σ_v^2	0.57	0.71	0.69
	Half-life	**1.8 yrs.**	**0.8 yrs.**	**1.7 yrs.**
Good-specific weighting	ρ (SE)	0.73 (0.01)	0.49 (0.01)	0.50 (0.09)
	σ_η^2/σ_v^2	0.46	0.68	0.88
	Half-life	**2.2 yrs.**	**1.0 yrs.**	**1.0 yrs.**
Good and country	ρ (SE)	0.69 (0.01)	0.52 (0.01)	0.43 (0.11)
specific weighting	σ_η^2/σ_v^2	0.53	0.63	0.91
	Half-life	**1.9 yrs.**	**1.1 yrs.**	**0.8 yrs.**
Panel B: Bias adjusted estimates				
Average across goods	ρ	0.73	0.65	0.73
	Half-life	**2.2 yrs.**	**1.6 yrs.**	**2.2 yrs.**
Equal weighting	ρ	0.86	0.72	0.88
	Half-life	**4.6 yrs.**	**2.1 yrs.**	**5.4 yrs.**
Good-specific weighting	ρ	0.89	0.77	0.79
	Half-life	**6.0 yrs.**	**2.7 yrs.**	**2.9 yrs.**
Good and country	ρ	0.87	0.79	0.77
specific weighting	Half-life	**5.0 yrs.**	**2.9 yrs.**	**2.7 yrs.**
Number of cities		23	26	13

Notes: Based on 11 year panel (1990-2000). The point estimates of ρ and σ_η^2/σ_v^2 are shown for each price index except for the first row in each panel where the average of estimates across individual goods is provided. Numbers in parentheses are the standard errors. Price indexes are constructed as follows: i) equal weighting involves taking simple averages of prices across good, ii) good-specific weights means that each goods receives a common weight across cities within each city group, but the weights differ across goods (the average consumption weight applicable across cities within each group is the good-specific weight), iii) a CPI-like methodology, applying a good and country-specific weight to each price.

Figure 1
International and Intranational Price (Individual Goods)

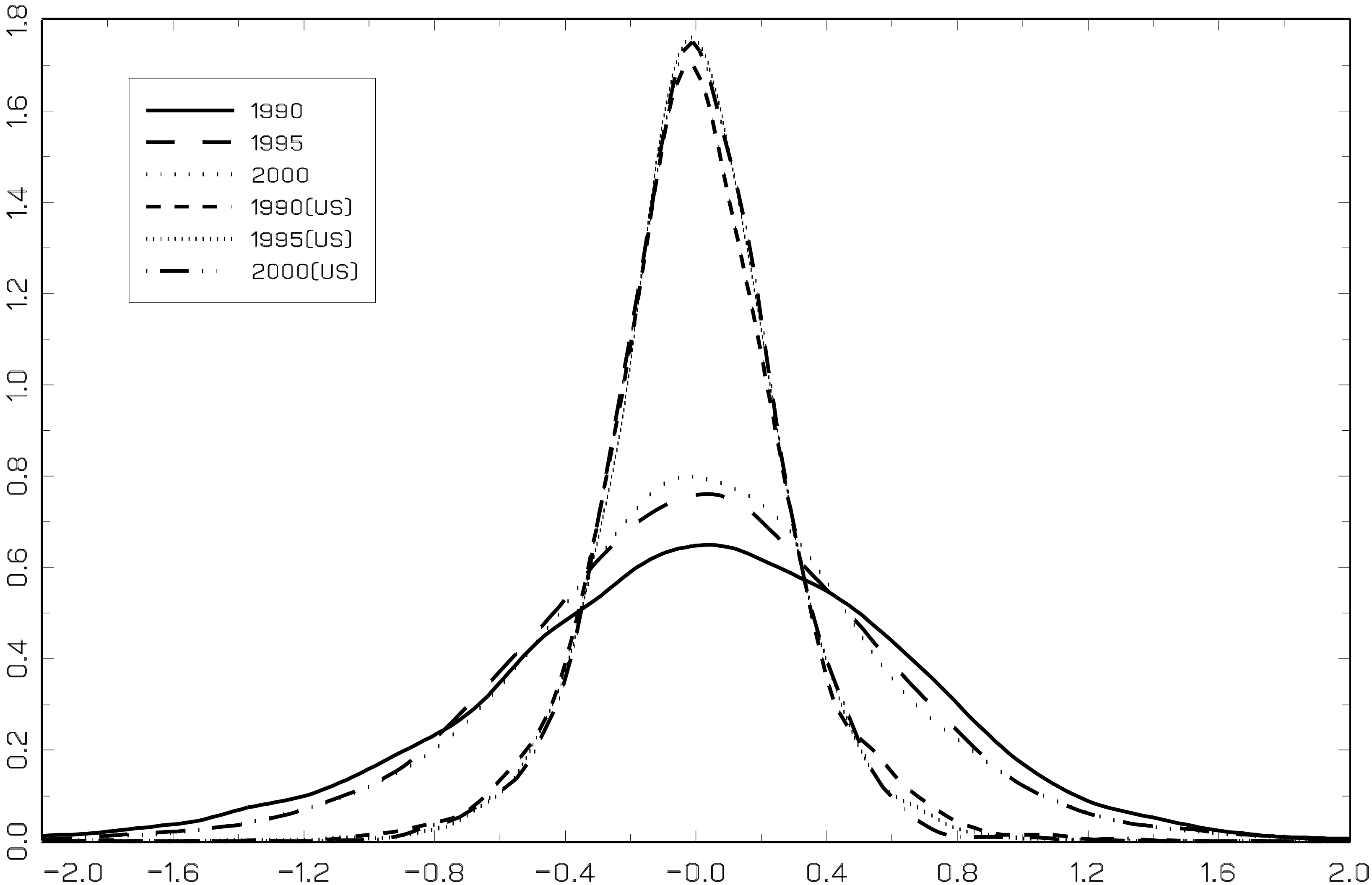

Figure 2

International and Intranational Price (Index)

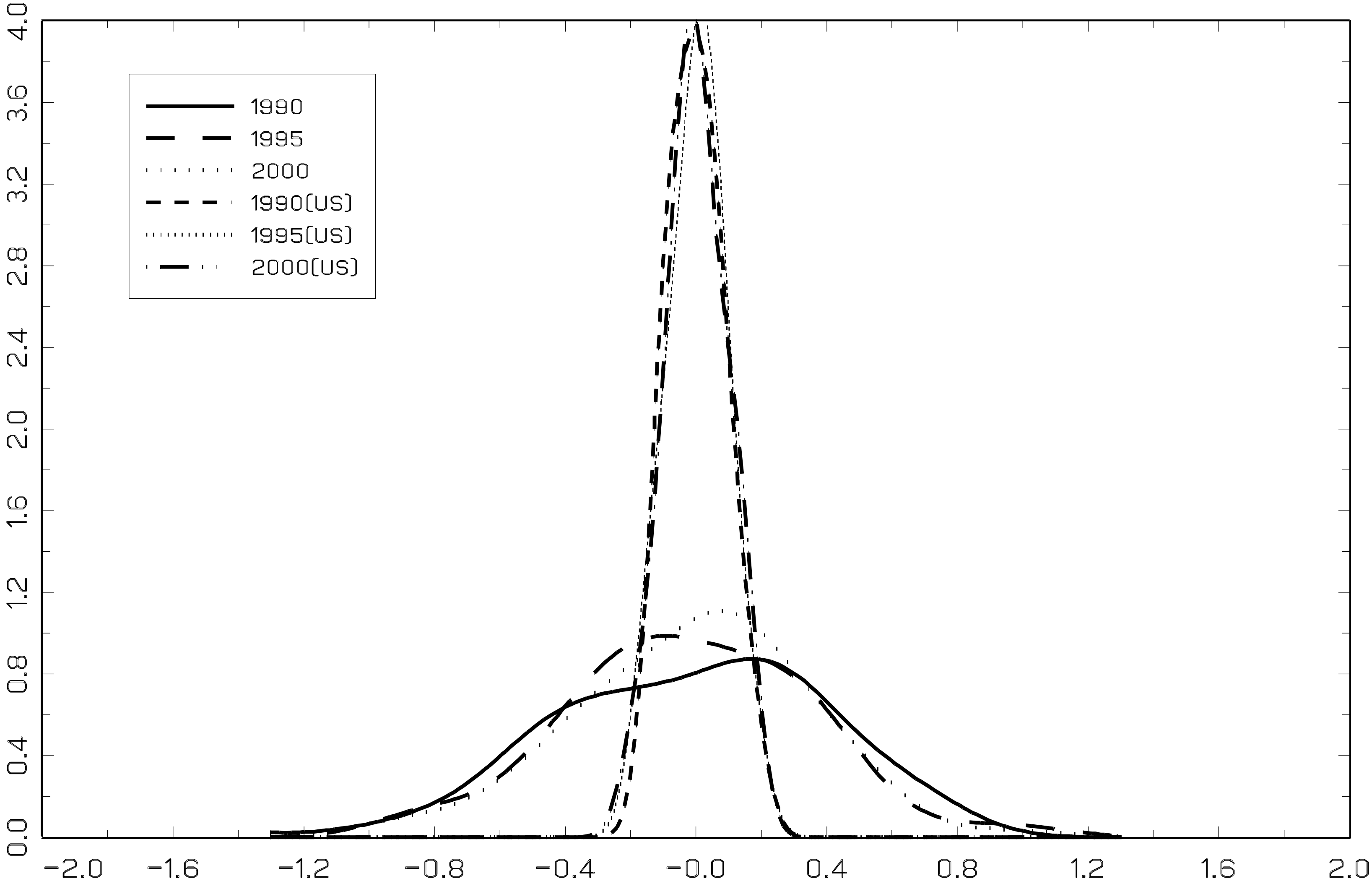

Figure 3
International and Intranational Price in 2000

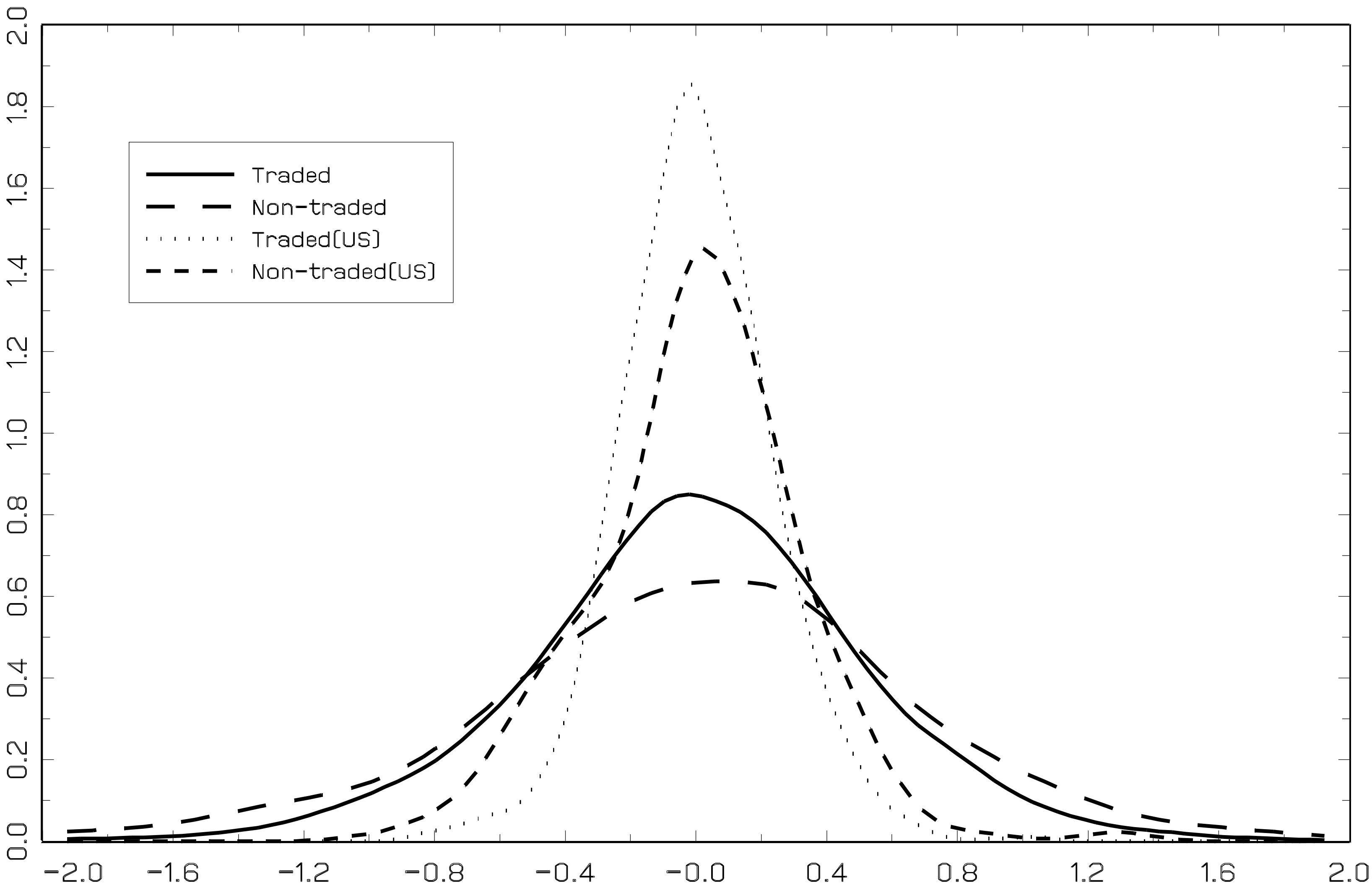

FIGURE 4. BIAS FUNCTION OF TRUE PERSISTENCE AND RELATIVE IMPORTANCE OF INDIVIDUAL EFFECTS FOR N=23