PPP Strikes Back: Aggregation and the Real Exchange Rate*

JEAN IMBS[a,b] HAROON MUMTAZ[c*] MORTEN O. RAVN[a,b] HELENE REY[b,d,e]
[a]*London Business School*, [b]*CEPR*, [c]*Bank of England*,
[d]*Princeton*, [e]*NBER*

This version: April 2004

Abstract

We show the importance of a dynamic aggregation bias in accounting for the PPP puzzle. We prove that the aggregate real exchange rate is persistent because its components have heterogeneous dynamics, that established time series and panel methods fail to control for. When heterogeneity is taken into account, the estimated persistence of real exchange rates falls dramatically. Its half-life, for instance, falls to around one year, significantly below Rogoff's 'consensus view' of three to five years. We show that corrected estimates are consistent with plausible nominal rigidities, thus, arguably, solving the PPP puzzle.

Keywords: Real Exchange Rate Persistence, Purchasing Power Parity, Parameter Heterogeneity.
JEL Classification: F36, F41, C43

---

*We thank Pól Antras, Mick Devereux, Charles Engel, Martin Evans, Lutz Kilian, Hashem Pesaran, Roberto Rigobon, Ken Rogoff, Barbara Rossi, Chris Sims, Ron Smith, Antonio Spilimbergo, Shang-Jin Wei, Mike Woodford, Charles Wyplosz, and seminar participants at ESEM 2003, ESSIM 2002, NBER IFM March 2003, Harvard, Penn, NYU, Stanford, Duke, Berlin, Bristol, LBS, Cambridge, NY Fed, the ECB, the IMF and the ESRC Money-Macro Finance conference in Warwick for helpful comments. Parts of this paper were completed while Ravn was a visitor at the University of Pennsylvania, and Imbs and Rey were visitors at the IMF. The authors are grateful to these institutions for their hospitality. The paper is part of the project "Exchange Rates, International Relative Prices, and Macroeconomic Models", funded by the ESRC (grant no.L138 25 1043), and of the RTN programme "The Analysis of International Capital Markets: Understanding Europe's Role in the Global Economy", funded by the European Commission (contract no. HPRN-CT-1999-00067). *This paper does not represent the views of the Bank of England or of Monetary Policy Committee members. The work was done while the author was at London Business School.

1 Introduction

The study of real exchange rates, defined as the international relative price of a basket of goods expressed in a common currency, is perhaps the most intensely researched area in international macroeconomics. In its absolute version purchasing power parity (PPP) states that real exchange rates should be constant and equal to one or, expressed in relative terms, that there should be no changes in the real exchange rate. Yet the consensus emerging from an extensive literature appears to be that, although real exchange rates may converge to parity in the long run, the rate at which this happens is very slow. The speed of mean reversion is usually summarized by one statistic, the half-life, or the time necessary for half the effect of a given shock to dissipate. According to Rogoff's (1996) authoritative survey, standard estimates of the real exchange rate half-life lay in the range of three to five years. While the high volatility of real exchange rates could potentially be explained by monetary (or financial) shocks, the rate of mean reversion is far too slow to be compatible with plausible nominal rigidities. Hence, Rogoff argues, the "PPP puzzle".[1] Kilian and Zha (2002) conduct a survey among international economists regarding their views on real exchange rate half-lives and find responses that display a single peak around four years, which thus seems to emerge as the 'consensus half-life' in the economics profession. Evidence on the law of one price (LOP) is hardly more encouraging, as it suggests persistent international differences in goods prices as well.[2]

If this interpretation of the empirical evidence is right, these large estimates have three important consequences. First, PPP is at best of little practical relevance over horizons of concern to policymakers or practitioners, who are typically interested in the short- to medium-run fluctuations of the economy. Second, economic models based on the PPP assumption are unlikely to provide an adequate description of reality at any relatively short horizon. Third, the slow convergence of international prices towards parity makes it quantitatively difficult to ascribe the failure of PPP to temporary arbitrage impediments or sticky prices. This has spawned a literature taking up the challenge of constructing calibrated macroeconomic models with nominal rigidities able to replicate the observed persistence of the real exchange rate.[3]

[1] See Froot and Rogoff (1995) for another survey. The "consensus view" is based on panel data estimates, as in Frankel and Rose (1996), Oh (1996), Wu (1996) or Lothian (1997), or on estimates using long spans of data, as in Mark (1995), Lothian and Taylor (1996), Abuaf and Jorion (1990), Glen (1992) or Taylor and Sarno (1998). Diebold, Husted, and Rush (1991) look at the Gold Standard and find a similar result, while Coakley et al (2002) survey the evidence on long-run PPP. The debate is not over however. Recently, studies emphasizing non-linearities argue that the true half life is in fact smaller than this consensus estimate (Taylor, Peel and Sarno (2001), Taylor and Peel (2000) or Obstfeld and Taylor (1997)). Others argue instead that it could in fact be much bigger (Murray-Papell (2002c), Engel (2000a)) or that the confidence intervals are far too wide to tell (Murray-Papell (2002a), Rossi (2003), Kilian and Zha (2002)).

[2] Classic studies include Giovannini (1988), Isard (1977), Knetter (1989, 1993), and Richardson (1978). Goldberg and Knetter (1997) surveyed the LOP literature. More recent references include Rogers and Jenkins (1995), Crucini, Telmer and Zachariadis (2001), Haskel and Wolf (2001) or Parsley and Wei (2003).

[3] Examples include Kollmann (2001), Chari, Kehoe and McGrattan (2002) and Bergin and Feenstra (2001).

This paper takes issue with the "consensus view" and shows that slow mean reversion in the aggregate real exchange rate is consistent with - on average - much faster adjustment of disaggregated relative prices. Existing estimates of real exchange rate persistence are based upon the (implicit) assumption that all relative prices composing the real exchange rate converge to parity at the same speed. This is a maintained hypothesis in both pure time series and panel studies of the real exchange rate, including those using sectoral relative prices. But there is little (if any) theoretical justification for the assumption that the relative prices embedded in real exchange rates adjust at identical rates over time. Indeed, it is hard to think of reasons why clothes and vegetables, say, should revert to parity at the same speed. Rather, the speed of mean reversion should depend on good and market characteristics, insofar as they affect the price-setting behavior of firms across borders.

We show how the failure to allow for heterogeneity in price adjustment dynamics at the good level induces a *positive* bias in persistence estimates, irrespective whether the estimation is performed using pure time series, a panel of aggregate real exchange rates, or even a panel of sectoral relative prices. We stress the importance of correcting for heterogeneity when estimating persistence in the real exchange rate. When heterogeneity in adjustment dynamics is allowed for, panel data estimates point to an average speed of mean reversion much faster than the consensus view. The persistence of disaggregated relative prices is on average much smaller than the persistence of the aggregate real exchange rate itself.[4]

The possibility that heterogeneity should generate biases in panel estimation was first highlighted by Robertson and Symons (1992), followed and generalized by Pesaran and Smith (1995). But, to the best of our knowledge, it has never been applied to panel estimates of relative price adjustment, nor to standard estimates of real exchange rate persistence.[5] Furthermore, we show that under general conditions (which will turn out to be borne out in the data), the bias is systematically *positive,* and its magnitude increases with the extent of heterogeneity. *Importantly, our result does not require nor imply that persistence be systematically lower at the disaggregated*

[4]Going back to Isard (1977), there has been a long tradition of research concerned with the behavior of disaggregated relative prices, with focus on estimates of the exchange rate pass-through at the good or sector level. Perhaps most prominently, Engel (2000b) shows that most relative prices adjust only slowly to changes in the nominal exchange rate, which suggests persistence is high in general at the sectoral level. Engel does not provide direct measures of persistence, however, which makes direct comparison with our estimates difficult. Even so, our contention is not that disaggregated relative prices should all revert quickly to parity, but rather that they do so to an heterogeneous extent. See also Campa and Goldberg (2002), or Cheung, Chinn and Fujii (2001) for recent persistence estimates at the sectoral level. It is this very heterogeneity that we argue obscures aggregate estimates of persistence.

[5]A search on citations for Pesaran and Smith (1995) yielded 99 entries, only one of which pertains to the real exchange rate. Boyd and Smith (1999) focus on heterogeneity across countries, and find it to be only mildly relevant for the real exchange rate. We show that the heterogeneity is large across *sectors*, and this affects standard international panel estimates.

level.[6] If persistence were homogeneously low across all disaggregated relative prices, there would be no PPP puzzle in the first place because the aggregate persistence would also appear low.[7,8]

Our contention is not that aggregate estimates of real exchange rate persistence are "wrong", as the notion of a bias might suggest. The real exchange rate is a well-defined object, and one can study its properties using standard techniques. Rather, we take issue with the interpretation of the standard results. The PPP puzzle arises because the estimated real exchange rate persistence is construed to be excessive in reference to theories where differences in prices are sustained by limits to arbitrage or nominal rigidities. Chari, Kehoe and McGrattan (2002) go one step further and show that even if one assumes implausibly large impediments to price adjustment one cannot account for the observed real exchange rate persistence. Here we argue that impediments to arbitrage or nominal rigidities have every reason to vary with each good's characteristics, and thus so do the dynamics of relative prices.[9] Hence some sectoral real exchange rates revert to the mean slowly and others do quickly. This heterogeneity is precisely the reason why the aggregate properties of the real exchange rate are *not* puzzling, since it is heterogeneity that gives rise to highly persistent *aggregate* series while relative price persistence is low on average at a *disaggregated* level. To make this precise, we show that an estimate of aggregate persistence that does account for dynamic heterogeneity is low, and in line with the model of Chari, Kehoe and McGrattan (2002), calibrated with plausible nominal rigidities. In this sense, there is no PPP puzzle.

We quantify the bias using an international sectoral price database issued by Eurostat. We find it to be substantial. Our estimate of the half life is eleven months with a confidence interval ranging from seven to twelve months. This is far below standard estimates, and it is not due to any specificities in our data, since we reproduce the 'consensus view' when we do not correct for heterogeneity. The intuition for this dramatic result is straightforward. If the persistence of relative prices varies across sectors, but aggregate estimates are calculated under the premise of a common aggregate autoregressive coefficient, the heterogeneity in the relative price dynamics gives rise to correlation between the regressors and the residuals. Under conditions which hold in our sectoral price data, this induces a systematically positive bias in persistence estimates.

---

[6] In that sense, our results are different from - but not in contradiction with- Parsley and Wei (1996), who examine the rate of convergence of relative goods prices across U.S. cities and find faster mean-reversion than in the aggregate.

[7] And one would need to resort to other - yet not competing - arguments, for instance along the lines argued in Taylor (2001), who studies *temporal* aggregation issues. These two types of aggregation biases (temporal and sectoral) are quite distinct conceptually and may well both be present at the same time.

[8] Our results are therefore also consistent with Crucini and Shintani (2002), who find *homogeneously* rapid reversion to parity amongst the relative prices sampled internationally by the Economist Intelligence Unit. Crucini and Shintani find little difference between aggregate and disaggregate mean reversion precisely because there is little heterogeneity in their specific sample of goods prices.

[9] For instance, Blanchard (1987) or more recently Bils and Klenow (2002) discuss the relevance of heterogeneity in price adjustments at the disaggregated level.

Our results are robust. In particular, we consider two alternative explanations. First, measurement error may be a serious problem in sectoral data since it may give rise to an attenuating bias in the sectoral autoregressive parameters. If this were the case, however, persistence estimates would be systematically lower at the sectoral level - which they are not. We nevertheless run formal tests for the presence of errors-in-variables but find practically no evidence for the presence of measurement error. We also confirm our results in other versions of the Eurostat dataset.[10] Second, a recent strand of literature contends that real exchange rate half-lives may be larger than our - and other - estimates suggest. When the underlying data generating process is highly persistent, as is the case for real exchange rates, standard least squares persistence estimates tend to be biased downwards, unless the sample is sufficiently long. This bias, in turn may translate into abnormally low half-life estimates. Based on this insight, Murray and Papell (2002a) implement a bias reduction method to a single real exchange rate series. Their corrected confidence intervals for the real exchange rate half-life are so wide as to bring the whole 'consensus view' into question.[11]

We implement a bias reduction method on our disaggregated data. The approach is adapted so that it is well-suited to a dynamic heterogeneous panel of sectoral relative prices. Our persistence estimates corrected in this way rise only from thirteen to eighteen months, and they are estimated precisely with a confidence interval that excludes Rogoff's 'consensus view'. The substantial heterogeneity in the data induces a (positive) bias considerably larger in magnitude than the (attenuating) small sample bias.[12]

The rest of the paper is structured as follows. We next describe in detail the bias that plagues dynamic panel and time series estimates when there is sectoral heterogeneity. We derive a general analytical expression for the bias and present conditions for this bias to be positive. Section 3 introduces the data, performs basic tests and shows that the conditions for the positivity of the bias are borne out in the data. In sections 4 and 5, we review the various existing procedures used to estimate half-lives, and we reproduce standard results with our data. We next test for heterogeneity and find strong support for heterogeneous dynamics across sectors in the data. We use accordingly estimators that allow for the heterogeneity. Results change dramatically. Section 6 examines alternative explanations for our findings and performs robustness checks. Section 7 concludes.

[10] We use the version advocated by Charles Engel on his website. For more details, see also Imbs, Mumtaz, Ravn and Rey (2004).

[11] The conclusion that there is not enough information in the aggregate data to pin down reliably the value of the half-life is also consistent with results in Kilian and Zha (2002) obtained in a Bayesian framework. More recently, Murray and Papell (2002b) have argued that Rogoff's consensus view may be rehabilitated on the basis of real exchange rates panel evidence.

[12] In Imbs, Mumtaz, Ravn and Rey (2004), we detail the reasons why the small sample bias is limited in our dataset. Most studies are based on panels of *aggregate* real exchange rates, and therefore draw conclusions about the importance of small sample bias that do not apply here.

2 Heterogeneous Adjustment Dynamics in Theory

This section explains how the failure to account for heterogeneity in relative price dynamics gives rise to a positive bias whose magnitude increases in the degree of heterogeneity. We do this in three steps. First, we focus on a panel of sectoral relative prices, follow Pesaran and Smith (1995) and show the conditions for an upward bias that rises with heterogeneity. Second, we show that the problem subsists in (time-series) estimations using real exchange rates, i.e. an aggregate of sectoral relative prices. Third, we extend the result to panels of real exchange rates. The latter two sub-sections show the bias first analyzed by Pesaran and Smith (1995) prevails in most of the existing studies of real exchange rate persistence. The average persistence in relative prices is systematically smaller than the persistence of the aggregate real exchange rate, because of heterogeneity in the dynamics of relative prices.

2.1 Bias in Dynamic Heterogeneous Panels

We build on the classic work of Pesaran and Smith (1995) who generalized the insights from Robertson and Symons (1992) on the econometric issues arising in panels with heterogeneous dynamics. We show the conditions under which standard estimators will be biased upwards in panels of sectoral real exchange rates.

Consider estimating the mean persistence of sectoral real exchange rates in a panel consisting of N cross-sectional units. To simplify, but without loss of generality, assume that each of the panel elements is given by a first-order autoregressive process:

$$q_{it} = c_i + \rho_i q_{it-1} + \varepsilon_{it},\ i = 1, .., N \tag{1}$$

with $c_i = c + \eta_i^c$ and $\rho_i = \rho + \eta_i^\rho$.[13] We assume that η_i^c and η_i^ρ have zero mean and constant covariance, and that the set of random coefficients ρ_i has support within the interval $]-1, 1[$. Furthermore, we assume for now that ε_{it} is independently distributed with mean 0 and variance σ_i^2.[14] We seek to estimate ρ, the average persistence of the relative prices. We can, without loss of generality order the N sectors so that for all i, $\rho_{i+1} \geq \rho_i$, $\rho_i \in]0, 1[$, where we impose, realistically, that relative prices are positively serially correlated.

An estimation where the persistence parameters are constrained to be homogeneous across sectors would have the following form:

$$\begin{aligned} q_{it} &= c_i + \rho q_{it-1} + e_{it} && (2) \\ e_{it} &= \eta_i^\rho q_{it-1} + \varepsilon_{it},\ i = 1, .., N. && (3) \end{aligned}$$

[13] Throughout, we consider drawing η_i^ρ from a discrete set of values.

[14] This section follows closely Pesaran and Smith (1995) including their distributional assumptions, and is used to build the intuition for our results. We relax the assumption of zero cross sectoral correlation when we turn to the properties of estimators run on aggregate indices in the next section.

It follows immediately from equation (3) that as soon as the dynamics of the panel units are constrained to be homogeneous, the lagged dependent variable enters the error term and estimates of ρ are inconsistent.

Now let ρ^Q denote the fixed effect estimator of the first order autoregressive coefficient. Pesaran and Smith (1995) show that its asymptotic expectation is given by

$$\rho^Q = \rho + \Delta$$

where

$$\Delta = \frac{\frac{1}{N}\sum\limits_{i=1}^{N}\left(\frac{\eta_i^{\rho}\sigma_i^2}{1-\rho_i^2}\right)}{\frac{1}{N}\sum\limits_{i=1}^{N}\left(\frac{\sigma_i^2}{1-\rho_i^2}\right)} = \sum_{i=1}^{N}\left(\rho_i - \rho\right)\frac{\frac{\sigma_i^2}{1-\rho_i^2}}{\sum\limits_{i=1}^{N}\left(\frac{\sigma_i^2}{1-\rho_i^2}\right)}$$

Hence the expression of the bias is given by

$$\Delta = \sum_{i=1}^{N}\left(\rho_i - \rho\right)\alpha_i$$

where

$$\alpha_i = \frac{\frac{\sigma_i^2}{1-\rho_i^2}}{\sum\limits_{i=1}^{N}\left(\frac{\sigma_i^2}{1-\rho_i^2}\right)}$$

.

Proposition 1 The sign of the bias arising from the failure to account for dynamic heterogeneity across panel units is given by the sign of $\Delta = \sum_{i=1}^{N}\left(\rho_i - \rho\right)\alpha_i$. For a large N, the bias is therefore positive *if and only if* $cov\left(\widetilde{\rho}, \widetilde{\alpha}\right) > 0$, i.e. the covariance between the vector of persistence parameters $\widetilde{\rho} = \{\rho_i\}_{i=1}^{N}$ and the vector of coefficients $\widetilde{\alpha} = \{\alpha_i\}_{i=1}^{N}$ is positive.

Proof. From the above derivation, it is immediate that the sign of the bias is the same as the sign of Δ. By definition, $cov\left(\widetilde{\rho}, \widetilde{\alpha}\right) = \lim\limits_{N\to+\infty}\frac{1}{N(N-1)}\sum_{i=1}^{N}\left(\rho_i - \rho\right)\left(\alpha_i - \alpha\right)$ where α is the mean of the coefficients $\{\alpha_i\}_{i=1}^{N}$. Hence $cov\left(\widetilde{\rho}, \widetilde{\alpha}\right) = \lim\limits_{N\to+\infty}\frac{1}{N(N-1)}\sum_{i=1}^{N}\left(\rho_i - \rho\right)\alpha_i - \lim\limits_{N\to+\infty}\frac{1}{N(N-1)}\sum_{i=1}^{N}\left(\rho_i - \rho\right)\alpha$.

By definition of ρ and the ρ_is $\lim\limits_{N\to+\infty}\frac{1}{N(N-1)}\sum_{i=1}^{N}\left(\rho_i - \rho\right)\alpha = 0$. The sign of $cov\left(\widetilde{\rho}, \widetilde{\alpha}\right)$ is therefore the same as the sign of $\frac{1}{N(N-1)}\sum_{i=1}^{N}\left(\rho_i - \rho\right)\alpha_i$, which is the sign of Δ. ■

Corollary 1.1 A sufficient condition for the dynamic heterogeneity bias to be positive is that $0 \leq \alpha_i \leq \alpha_{i+1}$ for all i.

Proof. These inequalities constitute a strong sufficient condition that ensures the positivity of Δ. Observe that $\sum_{i=1}^{N} \alpha_i = 1$ so that the α_i constitute *convex weights.* Our assumption on the ordering of the ρ_is implies that $0 \leq \alpha_i \leq \alpha_{i+1}$. Therefore we have

$$\sum_{i=1}^{N} \rho_i \alpha_i \geq \frac{1}{N} \sum_{i=1}^{N} \rho_i$$

since two sets of *convex* combinations of the ρ_is are compared, one with increasing weights and one with equal weights. The left hand side convex combination, involving the α_is, gives higher weights to the largest ρ_is, while it gives equal weights $\frac{1}{N}$ to the ρ_is on the right hand side of the inequality. Hence the inequality holds. Then $\frac{1}{N} \sum_{i=1}^{N} \rho_i = \rho = \sum_{i=1}^{N} \rho \alpha_i$ implies:

$$\sum_{i=1}^{N} \rho_i \alpha_i \geq \sum_{i=1}^{N} \rho \alpha_i$$

which is equivalent to:

$$\sum_{i=1}^{N} \alpha_i \left(\rho_i - \rho\right) \geq 0$$

Therefore, $\Delta = \sum_{i=1}^{N} \left(\rho_i - \rho\right) \alpha_i \geq 0$ and the bias is positive. ∎

Section 3 verifies that the condition spelled out in Proposition 1 for positivity of the bias holds in our data, i.e. the covariance between the estimation weights $\{\alpha_i\}_{i=1}^{N}$ and the persistence parameters $\{\rho_i\}_{i=1}^{N})$ is positive. The intuition for the sign of the bias can be understood straightforwardly with the help of the sufficient condition described in Corollary 1. If $0 \leq \alpha_i \leq \alpha_{i+1}$, then α_i is higher for large realizations of $\rho_i - \rho$, and the fixed effects estimates of ρ^Q are dominated by the components of the relative prices that revert to parity the slowest. Hence $\sum_{i=1}^{N} \alpha_i \left(\rho_i - \rho\right) \geq 0$ and the bias is unambiguously positive. In practice it is of course enough, when N is large, that α_i *tend to be higher* for large realizations of $\rho_i - \rho$. This is equivalent to checking that $cov\left(\widetilde{\rho}, \widetilde{\alpha}\right) > 0$ is verified in the data.

Corollary 1.2 A (positive) bias tends to increases, *ceteris paribus*, with the cross-sectoral dispersion in persistence.

Proof. This comes directly from the expression of the bias, $\Delta = \frac{\frac{1}{N} \sum_{i=1}^{N} (\rho_i - \rho) \frac{\sigma_i^2}{1-\rho_i^2}}{\frac{1}{N} \sum_{i=1}^{N} \left(\frac{\sigma_i^2}{1-\rho_i^2}\right)}$ whose magnitude increases on average with $(\rho_i - \rho)$, the distance between sectoral persistence to the cross-sectional average, holding $\{\sigma_i^2\}$ constant. ∎

It is also evident that any type of instrumentation will not solve the problem since any instrument that is highly correlated with the dependent variable will unavoidably also be correlated with the error term. In the presence of a lagged dependent variable, a common approach to handling the presence of fixed effects is to first-difference the data and use the IV or GMM estimators suggested in Anderson and Hsiao (1982) and Arellano and Bond (1991). But under dynamic heterogeneity, this will still lead to inconsistent estimates since

$$\begin{aligned}\triangle q_{it} &= \rho \triangle q_{it-1} + \triangle e_{it} \\ \triangle e_{it} &= \eta_i^{\rho} \triangle q_{it-1} + \triangle \varepsilon_{it}\end{aligned}$$

Standard panel data estimators suffer from inconsistency when there is dynamic heterogeneity across panel units and, under plausible conditions, they will overestimate the average persistence of relative prices.

The vast majority of papers dealing with the PPP puzzle base their estimates of relative prices persistence not on sectoral data but on time series (or panels) of aggregate real exchange rates. Zellner (1969) showed that in the context of *static* panels, aggregate estimators averaging the data over groups (as aggregate real exchange rates average relative prices across sectors) provide consistent and unbiased estimates of the coefficient means. However, in *dynamic* panels the failure to control for heterogeneous dynamics will still result in inconsistent aggregate estimates. We next show why aggregation fails to solve the problem created by heterogeneity.

2.2 Aggregation Bias: Time Series

This section describes how the heterogeneous dynamics at the sectoral level translates into biased aggregate estimates. To repeat, the "aggregation bias" in this paper is nothing but a manifestation of dynamic heterogeneity in the aggregate estimates, that stems from heterogeneity in the components of the aggregate index. We first focus on the case where the panel consists of the relative prices of goods for a single country pair. Consider an economy with N sectors indexed by i and maintain the same assumption as earlier

$$q_{it} = c_i + \rho_i \; q_{it-1} + \varepsilon_{it}, \;\; i = 1, .., N$$

with $c_i = c + \eta_i^c$ and $\rho_i = \rho + \eta_i^{\rho}$. We now allow for non-zero cross-sectoral covariances of ε_{it}, with $E(\varepsilon_{it} \; \varepsilon_{jt}) = \sigma_{ij}$ for $i \neq j$. These correlations could arise, for example, from common shocks across goods or from omitted (unobservable) global influences. We again order the sectors so that ρ_i is non-decreasing in i, positive and strictly less than unity. It follows that:

$$\begin{aligned}\sigma_{q_i}^2 &= \frac{\sigma_i^2}{1-\rho_i^2} \\ \sigma_{q_i,q_j} &= \frac{\sigma_{ij}}{1-\rho_i \rho_j}\end{aligned}$$

The bilateral real exchange rate q_t can be approximated by a linear aggregation of the different sectors with weights ω_j associated with the j^{th} good[15]:

[15] This is a log-linear approximation to the CPI based real exchange rate when CPI weights are equal across countries (see appendix A1). For a large number of sectors, the discrepancy between the logarithm of a sum and

$$q_t = \sum_{j=1}^{N} \omega_j q_{jt}, \ \ \sum_{j=1}^{N} \omega_j = 1$$

In general, q_t can be written as:

$$\left[\prod_{i=1}^{N} (1 - \rho_i L)\right] q_t = \sum_{i=1}^{N} \omega_i \left(\prod_{j \neq i} (1 - \rho_j L)\right) \varepsilon_{it}$$

where L denotes the lag operator. As is well-known, cross-sectional aggregation of N $AR(1)$ processes produces an $ARMA(N, N-1)$.[16] If the dynamics of the cross-sectional units were homogeneous, $\rho_i = \rho$ for all $i = 1, .., N$, the roots would cancel out, and this ARMA process would simplify into an autoregressive process of order one.[17] In econometric terms, heterogeneity modifies directly the underlying process driving the real exchange rate, under some assumptions on the processes prevalent at the disaggregated level. Allowing for heterogeneity by simply estimating the real exchange rate as a large order ARMA process is a theoretical possibility. But pursuing this route will be impossible in most cases, for lack of degrees of freedom unless the sample period is long enough.[18] Heterogeneous estimators are better-suited to tackling the issue than estimating processes with infinite (or even high order) ARMA terms.

In fact, the vast majority of PPP studies estimate the persistence of the real exchange assuming that its dynamics are best described by an AR(p) process. Many studies actually use an

---

the sum of logarithms is negligible, and well-approximated by the third and higher moments of the cross-sectoral distribution of relative prices. Furthermore, using the sum of (log) relative prices to approximate PPP exchange rates is common practice.

[16]Granger and Morris (1976) showed that aggregation of N $ARMA\,(p_i, q_i)$ processes leads to an $ARMA\,(P, Q)$ process where $P \leq \sum_i p_i$ and $Q \leq \max_j\,(P - p_j + q_j)$. In general, the size of P and Q will depend on the degree of heterogeneity of the dynamics of the underlying data. Granger (1980) shows further that if N $AR(1)$ series are aggregated and the autoregressive parameters can take on *any* value (as opposed to, say, M discrete values) in a given interval, the aggregated data will correspond to no $ARMA$ process with a finite number of parameters. He derives examples in which the aggregated data displays long memory.

[17]In this case the process can be written as $(1 - \rho L)^N q_t = (1 - \rho L)^{N-1} \sum_i \omega_i \varepsilon_{it}$ which implies that $q_t = \rho q_{t-1} + \sum_i \omega_i \varepsilon_{it}$. Actually, any sub-set of autoregressive coefficients without heterogeneity will turn out to cancel out in the aggregate.

[18]If sectoral relative prices followed autoregressive processes of order I, the real exchange rate would follow an $ARMA[N \cdot I, (N-1) \cdot I]$. With $I = 2$ and 200 cross-sections, for instance, this implies properly accounting for heterogeneity requires estimating an $ARMA(400, 398)$. This is the reason why Pesaran and Smith (1995) argue heterogeneity complicates considerably the nature of the error structure in aggregate estimations imposing homogeneity, and actually recommend considering an *infinite* distributed lag specification. Few datasets (and certainly not ours) can afford this kind of degrees of freedom.

AR(1) as their standard specification.[19]. So will we to simplify the derivations. We have

$$\begin{aligned} q_t &= c + \rho q_{t-1} + \varepsilon_t \\ c &= \sum_{j=1}^{N} \omega_j c_j, \ \varepsilon_t = \sum_{j=1}^{N} \omega_j \varepsilon_{jt} + \sum_{j=1}^{N} \omega_j \eta_j^{\rho} q_{jt-1} \end{aligned}$$

Thus, the lagged dependent variables are present in the error term and we can show as in the previous section that this 'aggregation bias' is positive under plausible conditions and increasing in the degree of heterogeneity. To economise on notations but without loss of generality, we now assume that $c = 0$. Consider the (asymptotic properties of the) least squares estimate of the first-order autoregressive coefficient of q_t, given by $\rho^Q = E\left(q_t q_{t-1}\right)/E\left(q_t^2\right)$. We can derive (see appendix A2) that

$$\rho^Q = \rho + \Delta$$

with

$$\Delta = \left[\sum_{i=1}^{N} \frac{\omega_i^2}{1-\rho_i^2}\sigma_i^2 + \sum_{i\neq j}^{N} \frac{\omega_i\omega_j\sigma_{ij}}{1-\rho_i\rho_j}\right] \frac{(\rho_i - \rho)}{\sum_{i=1}^{N}\left(\frac{\omega_i^2}{1-\rho_i^2}\sigma_i^2 + \sum_{i\neq j}^{N}\frac{\omega_i\omega_j\sigma_{ij}}{1-\rho_i\rho_j}\right)}$$

It is useful to rewrite the bias as

$$\Delta = \sum_{i=1}^{N} (\rho_i - \rho)\,\delta_i$$

with

$$\delta_i = \frac{\left[\frac{\omega_i^2\sigma_i^2}{1-\rho_i^2} + \sum_{i\neq j}^{N}\frac{\omega_i\omega_j\sigma_{ij}}{1-\rho_i\rho_j}\right]}{\sum_{i=1}^{N}\left(\frac{\omega_i^2}{1-\rho_i^2}\sigma_i^2 + \sum_{i\neq j}^{N}\frac{\omega_i\omega_j\sigma_{ij}}{1-\rho_i\rho_j}\right)}$$

We can now spell out the conditions for the bias to be positive.

Proposition 2 The sign of the bias arising from the failure to account for dynamic heterogeneity across real exchange rate components is given by the sign of $\Delta = \sum_{i=1}^{N}(\rho_i - \rho)\,\delta_i$. For a large N, the bias is therefore positive *if and only if* $cov\left(\widetilde{\rho}, \widetilde{\delta}\right) > 0$, i.e. the covariance between the vector of persistence parameters $\widetilde{\rho} = \{\rho_i\}_{i=1}^{N}$ and the vector of coefficients $\widetilde{\delta} = \{\delta_i\}_{i=1}^{N}$ is positive.

Proof. The proof is the same as the proof of proposition 1, replacing α_i by δ_i. ■

Corollary 2.1 A sufficient condition for the dynamic heterogeneity bias to be positive is that $0 \leq \delta_i \leq \delta_{i+1}$ for all i.

[19] For instance, in their discussion of the literature on PPP, Choi, Mark and Sul (2003) assume all along an AR(1), a choice they label "conventional", and that is followed among many others by Taylor (2001) or Murray and Papell (2002b).

Proof. Observe that $\sum_{i=1}^{N}\delta_i = 1$. For the same reasons as in corollary 1.1, $0 \leq \delta_i \leq \delta_{i+1}$ for all i constitute a strong sufficient condition that ensures the positivity of Δ. ■

Corollary 2.2 A (positive) bias tends to increase *ceteris paribus* with the cross-sectoral dispersion in persistence.

Proof. The proof is identical to that of corollary 1.2. since again the magnitude of Δ increases with the distance between sectoral persistence and its cross-sectional average. ■

We note the similarity between these results and the results for disaggregated panel estimators that we derived in the previous section. The same intuition carries through. Again we verify in section 3 that the condition spelled out in Proposition 2 holds in our data, i.e. that the covariance between the estimation weights $\{\delta_i\}_{i=1}^{N}$ and the persistence parameters $\{\rho_i\}_{i=1}^{N}$ be positive.

Three points are worth stressing at this stage. First, Proposition 2 suggests the covariances between sectoral price residuals affect the magnitude and sign of the bias. If for instance price residuals in high persistence sectors tend to correlate negatively (positively) with other prices, high ρ_i sectors will have low (high) δ_i, and the conditions for Proposition 2 might not be met in the data (or on the contrary reinforced). By extension, even if we have $cov\left(\widetilde{\rho},\widetilde{\delta}\right) > 0$, it remains possible that the covariances σ_{ij} should affect the magnitude of the (then positive) bias. Hence, controlling for cross-sectoral correlations will be important in our empirical application. Second, Proposition 2 also clarifies the role of the weights used in aggregating sectoral prices into the Consumer Price Index. While δ_i depends on ω_i, the sign or the magnitude of the relation is by no means straightforward. As will become clear, our heterogeneous estimators aggregate autoregressive coefficients estimates, not price series. Ultimately, whether δ_i increases in ρ_i is an empirical question, which we address in the next section, and the weight each good receives in the Consumer Price Index is only one of the determinant of the relation between δ_i.and ρ_i. Third, our result does *not* imply that the persistence of the real exchange rate is not informative in itself. We merely stress that, in the face of heterogeneity, the persistence of the real exchange rate will not be a consistent estimate of the mean persistence of relative prices, ρ.

Consider the following simple example, illustrative of a class of sectoral prices for which the above (strong) sufficient condition spelled out in corollary 2.1 is verified.

Example 1. Suppose that for all i and j $\sigma_i^2 \backsimeq \sigma^2$, $\omega_i \simeq \omega_j$ and $\sigma_{ij} \backsimeq \chi \geq 0$.

Since $\rho_1 \leq \rho_2 \leq ... \leq \rho_N$, we have

$$\frac{\sigma^2}{1-\rho_{i+1}^2} \geq \frac{\sigma^2}{1-\rho_i^2} \Longrightarrow \sum_{\substack{j=1\\ j\neq i+1}}^{N} \frac{\chi}{1-\rho_{i+1}\rho_j} \geq \sum_{\substack{j=1\\ j\neq i}}^{N} \frac{\chi}{1-\rho_i\rho_j}$$

as the inequality holds true term by term. Hence, for all $i \in [1, ..., N-1]$

$$\frac{\sigma^2}{1-\rho_{i+1}^2} + \sum_{\substack{j=1\\ j\neq i+1}}^{N} \frac{\chi}{1-\rho_{i+1}\rho_j} \geq \frac{\sigma^2}{1-\rho_i^2} + \sum_{\substack{j=1\\ j\neq i}}^{N} \frac{\chi}{1-\rho_i\rho_j}$$

which is equivalent to

$$0 \leq \delta_i \leq \delta_{i+1}$$

■

The simplifying restrictions in this example ensure that the innovation variances at the sectoral level are similar across sectors, that cross-sectoral covariances be positive and not too different from each other, and that CPI weights be roughly homogeneous.[20] We note that such families of restrictions (broadly defined) are plausible for sectoral price data. And in section 3 we show unambiguously that the coefficients δ_i and the persistence parameters $(\rho_i - \rho)$ *covary positively in the data.*

2.3 Aggregation Bias: Panels of Real Exchange Rates

Most recent papers have studied panels of aggregate real exchange rates, and not only pure time series exercises focused on one single exchange rate. Here we show that the insights developed for time series results apply also in such panels. As is standard, we control for country fixed effects. We let the autoregressive coefficient vary *across sectors*, while the intercept is allowed to vary *across countries*.[21] The proof can easily be generalized to allow for heterogeneity in the autoregressive coefficients across countries, but results in Boyd and Smith (1999) suggest this dimension of heterogeneity is mildly relevant at best. Our sectoral relative prices can be written as

$$q_{ict} = \gamma_c + \rho_i \; q_{ict-1} + \varepsilon_{ict}$$

where c denotes a given country. The fixed effect estimate of the first order autoregressive coefficient for the aggregate real exchange rate is given by

$$\rho^Q_{FE} = E\left[\left(Q_{ct} - \bar{Q}_c\right)\left(Q_{ct-1} - \bar{Q}_c\right)\right] / E\left(Q_{ct} - \bar{Q}_c\right)^2$$

where $\bar{Q}_c = E(Q_{ct}) = \frac{1}{N}\sum_i \frac{\gamma_c}{1-\rho_i}$ denotes a country-specific average level of the real exchange rate. Let $\tilde{q}_{ict} = q_{ict} - \frac{\gamma_c}{1-\rho_i}$. It is immediate that $Q_{ct} - \bar{Q}_c = \frac{1}{N}\sum_i \tilde{q}_{ict}$. Thus, since $\tilde{q}_{ict} = \rho_i \; \tilde{q}_{ict-1} + \varepsilon_{ict}$ by definition, we have

$$\rho^Q_{FE} = \frac{\sum_{i=1}^N \rho_i \sigma^2_{\tilde{q}_{ic}} + \sum_{i<j}^N \left(\rho_i + \rho_j\right)\sigma_{\tilde{q}_{ic},\tilde{q}_{jc}}}{\sum_{i=1}^N \sigma^2_{\tilde{q}_{ic}} + 2\sum_{i<j}^N \sigma_{\tilde{q}_{ic},\tilde{q}_{jc}}}$$

with obvious notations. Thus, the rest of the proof in the previous section carries through almost identically. A fixed effect estimator with country-specific intercepts continues to suffer from a positive bias under similar conditions.[22]

[20] That covariances be positive rules out degenerate cases where some sectoral processes exactly cancel out.

[21] Country fixed effects are well-known to be important in real exchange rate estimations. See Frankel and Rose (1996). Crucini and Shintani (2002) illustrate the importance of controlling for permanent differences in prices at the goods level.

[22] The only difference with the univariate case pertains to the innovation variances, which are here allowed to be country specific, so that $\sigma^2_{\tilde{q}_{ic}} = \frac{\sigma^2_{ic}}{1-\rho_i^2}$ and $\sigma_{\tilde{q}_{ic},\tilde{q}_{jc}} = \frac{\sigma_{ijc}}{1-\rho_i\rho_j}$.

The intuition for the sign of the bias is again straightforward, and identical irrespective whether the estimator is univariate or based on panel estimators. Imposing homogeneity results in the estimator systematically ascribing a higher weight to those sectors where persistence is larger, which overestimates the extent of aggregate persistence. The bias increases in magnitude with the extent of cross-sectional heterogeneity. This is the main reason why we find a large bias when focusing on sectoral aggregation, while Boyd and Smith (1999) find little evidence that the autoregressive properties of the real exchange rate vary across countries. We note that a test may reject heterogeneity in panels of *aggregate* real exchange rates if differences in relative prices dynamics are mostly relevant at the *goods level*. This is indeed what we find in our empirical application. After all, heuristically there are many more reasons for differences to prevail when comparing prices dynamics in cars and bread, say, than in comparisons between different car makes or different bread types.

3 Data

In this section, we first describe our data, including a discussion of their accuracy and representativeness. We then check that they verify the basic conditions for our bias to be positive.

3.1 Description

We study relative prices of goods at the sectoral level. We use price data obtained from Eurostat, the statistical agency of the European Union. We focus on (non-harmonized) price indices for consumption goods and services since harmonized price indices are available for only very short samples. The data correspond to monthly observations and cover at most the period 1960:1 to 2000:12. However, many observations are missing in the early and late part of the period, so we choose to focus on a [1981,1995] sample. This leaves us with a maximum of 180 time series observations. We report results based on checking the data in painstaking detail. In particular, we correct for obvious repetitions or outliers, and, whenever possible, we use primary data sources obtained from national statistical agencies to correct the Eurostat data. Engel (2000b) uses the same data. On his website, Engel reports the corrections he chose to implement on the Eurostat official data.[23] Our corrections include all those reported by Engel but in some cases we were able to obtain the correct data rather than using subjective corrections (or eliminating the data).[24] Arguably, our final dataset, described in Appendix B, is therefore more accurate than Engel's.[25]

[23]See http://www.ssc.wisc.edu/~cengel

[24]We have additional observations for Finland and Greece, and in a few cases, chose not to dismiss data points that Charles Engel excluded because we were able to compare them with primary data sources.

[25]Our data can be downloaded at http://faculty.london.edu/jimbs, http://www.princeton.edu/~hrey or http://faculty.london.edu/mravn.

However, for completeness we also computed the results of our estimations based on Charles Engel's exact dataset, which we report in the text as well.

Eurostat contains two-digit sectoral price indices for nineteen goods categories and thirteen countries. The goods categories are a mixture of low and high unit costs goods (e.g. bread and cereals versus vehicles), highly tradeable goods (e.g. clothing), goods commonly construed as non-tradeable in nature (public transport or hotels), and goods for which there is wide variation in the degree of product differentiation (fuel versus sound and photographic equipment).[26] Our sample thus constitutes an interesting cross-section with some variation along the dimensions commonly advanced to explain variations in relative prices. The cross-sectional variation is key to our analysis since it allows us to identify the heterogeneity in relative price dynamics.

Our real exchange rates are CPI-based and defined against the US dollar.[27] Since our purpose is to investigate the effects of heterogeneity and aggregation, our sample of countries and the time coverage are identical for the two levels of aggregation. Furthermore, our measure of real exchange rates is based on the aggregation of the same exact sample of goods for which we have disaggregated information.[28] In particular, sectoral relative prices write

$$q_{ict} = \ln\left(\frac{S_{ct}\ P_{ict}}{P_{i,US,t}}\right)$$

where S_{ct} denotes the nominal bilateral exchange rate between country c and the US dollar at date t, P_{ict} is the price of good i in country c at date t, and $P_{i,US,t}$ is the corresponding US price. Real exchange rates, in turn, write

$$q_{ct} = \ln\left(\frac{S_{ct}\ P_{ct}}{P_{US,t}}\right)$$

We test for unit roots both for aggregate real exchange rates and for sectoral real exchange rates. We used two panel data tests: Levin and Lin (1993) and Im, Pesaran and Shin (1997). The Levin and Lin procedure (henceforth LL) tests the hypothesis that *all* the cross-sectional units are stationary against the hypothesis that they are *all* non-stationary. The Im, Pesaran and Shin test (henceforth IPS) is more general in that it allows for *some*, but not all, of the series to be stationary under the alternative hypothesis. Table 1 reports the results for the two tests. The first column reports the outcome of several unit root tests for the panel of aggregate real exchange rates. The third column concerns the panel of sectoral real exchange rates, and, for completeness, the second column pertains to the nominal exchange rate. In each case we

[26] See Appendix B for details.

[27] In unreported results, we use the British Pound as the anchor currency, and all our conclusions stand.

[28] We also used real exchange rate measures based on the International Financial Statistics database released by the IMF. The results, based on standard full CPI baskets, were almost identical.

report the IPS test and two variations of the LL test, allowing or not for individual effects. Each estimation is performed both with or without a trend term. Table 1 shows that the evidence tilts in favor of stationarity in both the real exchange rates and sectoral relative prices, with four out of five test statistics supporting stationarity in each case.[29] This is consistent with the findings of an enormous literature.[30]

3.2 Conditions for a Positive Bias

In section 2.1 and 2.2, we derived conditions for the positivity of the bias. Whenever the covariance between the coefficients α_i or δ_i and the persistence parameters ρ_i is positive, the bias is positive, since more persistent sectors get higher weights on average in the estimation. We now verify whether these conditions hold in our data.

For the nineteen sectors of our database we retrieve estimates of the autoregressive parameters, as well as the variance covariance matrix of the sectoral innovations. We use these estimates to compute the coefficients α_i and δ_i on the basis of the formulas derived in sections 2.1 and 2.2. In Figure 1 we plot the coefficients α_i and δ_i against estimates of the autoregressive parameters ρ_i. It is clear from the graphs than in both cases there is a strong positive correlation between the weights and the persistence parameters. The covariance between α_i and ρ_i is equal to 0.166. It is equal to 0.229 for δ_i.[31] We can also compute directly the value of Δ which is in all cases unambiguously positive. Thus, in our application to relative prices and real exchange rates, the sign of the aggregation bias is unambiguously positive. Aggregating sectoral prices without paying attention to heterogeneity in their dynamics leads to an overestimation of the persistence of the real exchange rate. We next discuss how to account for heterogeneity in the *empirical estimates*.

---

[29] The IPS test fails to reject non-stationarity of disaggregate relative prices when a trend is included, and of the real exchange rate when no trend is included. But standard tests also reject the presence of a trend in relative prices, and suggest there may be one in the real exchange rate.

[30] Frankel and Rose (1996) rejects the unit root hypothesis in a panel of 150 countries over 45 years. So does Oh (1996) in a similar study. Wu (1996) rejects non-stationarity in a panel test for monthly and quarterly data from the IFS database. Lothian (1997) focuses upon the post Bretton-Woods period and rejects non-stationarity of the real exchange rate in a panel that includes 23 OECD countries. On the basis of Monte-Carlo evidence, Engel (2000a) argues unit root tests may be unable to detect unit roots in real exchange rates in the presence of a stationary, but noisy, component. He concludes PPP may not hold even in the long run despite formal rejection of non-stationarity. Ng and Perron (2002) question his conclusions on grounds of conceptual issues in his interpretation of the mean square forecast error ratio, both in terms of test size and the importance of the non-stationary component. They estimate a half-life for real exchange rate shocks between nine and fifteen quarters, right back in the consensus view.

[31] The weights used are the Eurostat harmonized indices of consumer prices weights (obtained from direct communication with Eurostat). When equal weights are used the covariance is 0.223.

4 Econometric Methods

In an attempt to make the results as robust as possible we measure persistence using three alternative statistics. In the PPP literature, the most commonly used measure is by far the 'half-life', denoted $T_{1/2}$ here, and defined as the number of periods it takes until half the effect of a given shock dissipates. In the case of an autoregressive process of order one, as in the great majority of the studies on relative prices, $T_{1/2}$ can be computed as $\ln(0.5) / \ln(\widehat{\rho})$ where $\widehat{\rho}$ is the estimated first-order autoregressive coefficient. For higher order autoregressive models, we use the estimated impulse response function. In particular, we follow Kilian and Zha (2002) - among others - and define the half-life as the largest value of $T_{1/2}$ such that $\widehat{IR}\left(T_{1/2}-1\right) \geq 0.5$ and $\widehat{IR}\left(T_{1/2}\right) < 0.5$ where $\widehat{IR}(j)$ denotes the estimated impulse response function at horizon j to a unit innovation at time 0. We then compute confidence intervals using a non-parametric bootstrap procedure with 500 replications.[32] The half-life is appealing in that it has immediate intuition. But for completeness (and to ensure robustness), we also report the largest autoregressive root in the processes we estimate, as well as the cumulated impulse response (CIR), which measures the total cumulative effect of a unit shock to relative prices.[33] In all cases, bootstrapping procedures are implemented to derive standard errors bands around our persistence estimates.

Below, we first present the panel estimators usually implemented in the real exchange rate literature, namely the fixed effects, Anderson-Hsiao and Arellano-Bond estimators. We show in section 5 that they reproduce standard results on our data. However, Section 2 showed none of these estimators are appropriate when there is sectoral heterogeneity in the dynamic parameters. So we next present two models allowing for heterogeneity, the Mean Group (MG) and the Random Coefficient (RC) estimators. They are simple generalizations of the Fixed and the Random Effects estimators, respectively, and allow for cross-sectional variation not only in the intercept, but also in the actual coefficients on the regressors.[34]

4.1 Standard Panel Estimators

We study both panels of disaggregated relative prices and panels of real exchange rates. We specify the latter as follows:

$$q_{ct} = \gamma_c + \sum_{k=1}^{K} \rho_k \, q_{ct-k} + \varepsilon_{ct} \tag{4}$$

These are computed as the aggregate of the exact same series as the ones used in the relative price panel.[35]

[32] For the Arellano-Bond estimator, the bootstrap was performed using the Brown and Newey (2002) method.

[33] See Andrews (1993) for a discussion.

[34] Just as is the case with standard panel estimators, a type of Hausman test is required to decide which of the MG or the RC models should be implemented on the data. We clarify this below.

[35] We also used official real exchange rate series from the IMF. Our results were unchanged.

The possible presence of fixed effects through γ_c in equation (4) requires that the specification be estimated in first- or mean-differences. As is now well-known, the presence of a lagged dependent variable makes it necessary to use instrumental variables when estimating equation (4). Anderson and Hsiao (1982) proposed to instrument the differenced lagged dependent variable with its lagged level to alleviate the bias. The resulting instrumentation is often weak, which is why Arellano and Bond (1991) introduced a GMM procedure using all available lags as instruments of the differenced lagged dependent variable.[36]

Correspondingly, we use a standard specification to investigate the speed of mean reversion in disaggregated relative prices:

$$q_{ict} = \gamma_{ic} + \sum_{k=1}^{K} \rho_k \; q_{ict-k} + \varepsilon_{ict} \tag{5}$$

which we estimate allowing for fixed effects.

4.2 Heterogeneous Models

Our next step is to explore the cross-sectional heterogeneity of our panel. We now introduce estimators which can tackle the aggregation bias described above. Both the Mean Group and the Random Coefficient models allow for heterogeneous coefficients, and are therefore generalizations of standard Fixed or Random Effects estimators. We allow for the possibility that

$$q_{ict} = \gamma_{ic} + \sum_{k=1}^{K} \rho_{ick} \; q_{ict-k} + e_{ict} \tag{6}$$

where slopes and intercepts are allowed to vary across the panel units.

The MG and RC models differ in their assumptions on the nature of heterogeneity. While MG assumes 'deterministic' heterogeneity, RC generalizes to random individual specific components in all estimated coefficients. In particular, it assumes $\gamma_{ic} = \gamma + \eta_c^1$ and $\rho_{ick} = \rho_k + \eta_{ick}^2$, where η^1 and η^2 are assumed to have zero means and constant covariances. In words, heterogeneity enters as random deviations from a common intercept and common autoregressive coefficients. RC is a generalization of Random Effects, just as MG is one of Fixed Effects. The RC model entails a Generalized Least Squares (GLS) procedure that optimally accounts for the heterogeneous stochastic nature of the residuals. In particular, GLS uses the variance-covariance matrix for η^1 and η^2 to weigh optimally the individual sector-specific slopes when aggregating them. The MG

[36] Goldberg and Verboven (2001) present results for an estimation similar to ours, but they focus on the relative prices of automobiles only. Their equation (2) is different from our equation (4) in that they include q_{ict-1}, a lagged level of the relative price, in the set of independent variables of the differenced version of equation (4). Our maintained assumption of stationarity enables us to specify equation (4) in levels, eliminate fixed effects and correct for the bias implied by the presence of lagged dependent variables.

model introduced in Pesaran and Smith (1995), instead simply performs an arithmetic average of sector-specific slopes, with equal weights. The MG estimator will only be efficient if the optimal weights happen to be insignificantly different from the arithmetic ones. However, asymptotically, the two estimators are equivalent. This suggests a test procedure choosing between the two estimators, which is akin to the Hausman test used for standard panel estimates.

The standard MG model estimates $\{\rho_k\}_{k=1}^{K}$, the mean autoregressive coefficients, by a simple arithmetic average of the least squares estimates for sector specific coefficients. We next describe the more general RC model, since MG is but a special case where heterogeneity is deterministic. Rewrite equation (8) as

$$q_{ict} = \gamma + \sum_{k=1}^{K} \rho_k \; q_{ict-k} + \varepsilon_{ict}$$

with $\varepsilon_{ict} = e_{ict} + \eta_{ic}^1 + \sum_{k=1}^{K} \eta_{ick}^2 q_{ict-k}$. Consistent GLS estimates of the coefficients of interest in equation (8) are given by an optimally weighted average of sector-specific point estimates. The analogy with the Random Effects estimator can best be seen by rewriting the model as

$$Q_{st} = Q_{sK} \; B_s + e_s$$

where $Q_{st} = [q_{11t}, ..., q_{NCt}]'$, $Q_{sK} = [1, Q_{st-1}, ..., Q_{st-K}]$, $B_s = B + \eta_s$, with $B = [\gamma, \rho]'$, $\eta_s = [(\eta_s^1)', (\eta_s^2)']'$, and $\sigma_s^2 = E(e_s' \; e_s)$.[37] We have assumed that $E(\eta_s) = 0$. Further define $E(\eta_s \; \eta_s') = \Gamma$. The random coefficient estimator of B is given by

$$\begin{aligned} \widehat{B} &= \sum_s W_s \; B_{OLS}^s, \; W_s = \left[\sum_s (\Gamma + V_s)^{-1}\right]^{-1} (\Gamma + V_s)^{-1} \\ V_s &= \sigma_s^2 \left(Q_{sK}' \; Q_{sK}\right)^{-1} \end{aligned}$$

where B_{OLS}^s denotes the sector specific OLS estimates of the slopes. Thus, RC applies the information in Γ efficiently when averaging the sector specific slopes. As shown in Pesaran (2003), RC and MG estimators are asymptotically equivalent.

The RC model is consistent and if the data are best described by random heterogeneity in persistence coefficients, the RC model is also efficient, since it accounts for heterogeneity optimally. But if slope heterogeneity is best described as deterministic in the data, then the MG estimator should be preferred, just as Fixed Effect estimators would be in the simpler case of intercept heterogeneity. Deciding between the RC and the MG models is an empirical question, analogous to performing a Hausman test to choose between Fixed and Random Effects. In section 5, we perform the relevant tests and let the data decide which estimator is the most appropriate.

In Section 2, we allowed for non-zero cross-sectoral correlations in the residuals, with $E(\varepsilon_{it} \; \varepsilon_{jt}) = \sigma_{ij}$ for $i \neq j$. We shall want our estimator to allow for this possibility, too, since our proof showed

[37] This follows Hildreth and Houck (1968) and Swamy (1970, 1971).

that the magnitude of the bias depends on σ_{ij}. It is standard to implement the traditional Seemingly Unrelated Regression (SURE-GLS) remedy to correct for cross-sectional correlations in error terms. It can be applied to either one of our heterogeneous estimators, but, since it estimates the covariance matrix of the residuals in the panel, it requires that the cross-sectional dimension be smaller than the time dimension of the data. This is unfortunately not the case in our data, where $N = 204$ and $T = 180$ in the full sample. We need to truncate our data. Unavoidably, this is arbitrary, and so we choose to focus on the exact cross-section posted on Engel's website.

Alternatively, Pesaran (2002) introduces a common correlated effects (CCE) estimator, which is well-tailored for large panels with both cross-sectional interdependence and heterogeneity. The estimator provides a correction to the MG estimator that accounts for unobserved common factors potentially correlated with individual-specific regressors. CCE allows for common effects in the residuals, that can have a different impact on individual units, and that can be arbitrarily correlated amongst themselves. It is likely to improve on the SURE approach, as the estimation of the covariance matrix of the residuals has lower dimensionality, thanks to the structure imposed through the common effects. In CCE, we can include all cross-sections in our data, while keeping identification parsimonious. CCE may also yield more accurate estimates than SURE, since the latter is unable to capture, for instance, the effects of a persistent common factor on the residuals covariance matrix. Furthermore, the CCE estimator is straightforward to implement since the common effects correction of the MG estimator, for instance, simply amounts to including lagged cross-sectional averages in the least squares regressions performed by MG. We check the properties of this estimator on the basis of Monte-Carlo simulations to ensure its consistency in the context of our dynamic panel.

5 Aggregation Bias in Practice : PPP Strikes Back

In this section we investigate empirically the importance of heterogeneity by comparing results derived from standard methods (Fixed Effects, Anderson-Hsiao, Arellano-Bond) to those obtained from estimators allowing for dynamic heterogeneity. We first review aggregate results, and confirm our data are not particular in any way, as we are able to reproduce consensus estimates. We then implement heterogeneous estimators, and find substantially faster mean reversion in relative prices. Finally, we ask whether our estimates are consistent with plausible nominal rigidities.

5.1 Results for panels of Aggregate Real Exchange Rates

We estimate equation (4) using real exchange rates vis-a-vis the US dollar. This corresponds directly to standard estimates of real exchange rate persistence based on panels of real exchange rates. The results are reported in Table 2. Lag lengths were identified using a general-to-specific technique starting from a maximum of twenty lags.[38] We report two tests for parameter homogeneity: a Hausman type test and that proposed by Swamy (1971). Neither of them can reject the null hypothesis that the dynamics of the panel units are homogeneous (across countries) at any conventional levels of confidence. This suggests cross-country (dynamic) heterogeneity is not a source of bias in aggregate estimates.[39]

The first row in Table 2 reports the results based on the OLS fixed effects panel estimator. The estimates imply a half-life roughly at the center of the consensus view, with a point estimate of three years and ten months. The bootstrapped 95 percent confidence interval ranges from around two and a half years to just below five years. The alternative measures we report also imply high persistence. The largest autoregressive root for example has a point estimate of 0.97. These are entirely in line with existing results. The Table also reports that the presence of fixed effects cannot be rejected, and a Hausman test favors the fixed effects specification over a random effects model.[40]

The presence of fixed effects demands that the model be estimated in first differences. But as mentioned earlier, in dynamic panels this produces correlation between errors and the regressors, which requires instrumenting the lagged dependent variable. We report the results of two estimators traditionally used to solve this problem, the Anderson-Hsiao IV-type estimator and the Arellano-Bond GMM estimator. Both lead to a significant upward revision in the estimate of real exchange rate persistence. As far as the Anderson-Hsiao estimator is concerned, the implied half-life is six years. The 95 percent confidence interval ranges from just below three years to infinity, but this is probably due to poor small-sample properties (and a large root mean square error).[41] The Arellano-Bond estimator on the other hand, has both better small-sample properties and a lower root mean square error. Estimates imply a half-life of four and a half years, with a 95 percent confidence interval between just below four years and just above six years. Consistent with this, both the largest autoregressive root and the cumulated impulse response indicates high persistence as well. Our aggregate results fall exactly within the consensus range, and are in agreement with the existing literature. The aggregate dimension of our data generates perfectly standard results.

[38] Twelve lags in the case of the Anderson-Hsiao estimator.

[39] It does not imply that heterogeneous dynamics are not an issue. There may be heterogeneity at the goods level, which will not be detected by Hausman or Swamy tests applied to panels of real exchange rates.

[40] Unsurprisingly, the half life implied by pooled data, without any country-specific intercept, is absurdly high, around 450 years.

[41] Lagged relative prices make for weak instruments, and hence a poor first-stage fit.

5.2 Results for Sectoral Real Exchange Rates

Now we instead investigate the results based on the panel of disaggregate prices. We first present the results of our estimations when sectoral heterogeneity is accounted for. Then we provide further insights with several Monte-Carlo simulations.

5.2.1 Estimations

We work with exactly the same panel of sectoral prices that compose the aggregate CPI used in the aggregate analysis in the previous section. We use six alternative estimates. First, simple fixed effects estimates, which would be valid under homogeneity. Second, we extend the fixed effects estimator to allow for cross-sectional dependence. We use either a SURE approach, or the adjustment for common effects introduced in Pesaran (2002). Second, we check which heterogeneous estimator is applicable to our data, RC or MG. We then present results for the same three variations of the preferred heterogeneous estimator.

Table 3 summarizes all results. The fixed effects estimator, which does not allow for sectoral dynamic heterogeneity, implies a half-life of three years. The estimates are relatively precise, as the 95 percent confidence interval roughly ranges from two and a half to three and half years. This lies at the lower end of the consensus view, but does not differ markedly from previous results in the literature. We implement both the SURE and the CCE corrections to the fixed effects estimator, but the estimates remain largely unaffected, or if anything tend to increase. This suggests correction for correlated residuals does not bring down persistence estimates based on homogeneous estimators.

But both the Hausman and Swamy tests indicate clear rejection of the hypothesis of homogeneity of the slope coefficients across sectors, at any level of confidence.[42] This immediately implies that the fixed effects estimator is inconsistent, as discussed in section 2. A more subtle implication concerns the results reported in Table 2, which *also* suffer from the 'aggregation bias' we document, even though there is no heterogeneity across countries. Cross-sector heterogeneity induces a bias in cross-country standard panel estimates, as Section 2 made clear. This happens because, as soon as the basic conditions we list in Section 2 are met, any standard panel estimates will artificially assign a larger weight to the components of the real exchange rate that display slower mean reversion.[43] Thus, aggregate persistence estimates will be biased, and biased upwards given the specific properties of the price data that we documented in Section 3. It still remains to be seen however whether the bias is important quantitatively, too.

To that end, we now turn to estimators designed to account for dynamic heterogeneity. Table 3 reports the test introduced in Pudney (1978), meant to assess if the data support a Random

[42] The Hausman test allows for correlated residuals, and is based on the CCE correction. An alternative based on the SURE correction has value 57.68, with a P-value equal to zero.

[43] This can happen regardless whether CPI weights themselves increase or decrease with ρ_i.

Coefficient or a Mean Group model. Random Coefficient will only be consistent if stochastic heterogeneity is uncorrelated with the regressors. It is clear from Table 3 that the null hypothesis does not pass the test, and so the data resoundingly reject the Random Coefficient model. In what follows we therefore use the MG estimator.

The mean group estimator (MG) produces a half-life just above two years, and a 95 percent confidence interval ranging from fourteen to twenty eight months.[44] This is already significantly below the 'consensus view', an interesting outcome in itself for it suggests the aggregation bias is large and prevalent in our data.[45] Further, the 95 percent confidence interval for the largest autoregressive root now ranges from 0.903 to 0.973. The MG estimates are all almost significantly distinct from the intervals obtained using (homogeneous) fixed effects, whether on aggregated or on disaggregated data. This is true of both half-life measures and estimates of the largest root.

This leaves open the question of the role of correlated residuals. Table 3 reports a Breusch-Pagan test checking the diagonality of the covariance matrix of the residuals, as implied by the MG regressions. The null-hypothesis of diagonality is overwhelmingly rejected. As mentioned in Section 2, correlated residuals can affect the magnitude of the bias. The Mean Group estimator modified to allow for correlated residuals (either in the SURE or the CCE model) may generate higher (lower) half-lives than standard MG, if the residuals e_{ict} tend to be negatively (positively) correlated across sectors. A common prior is that price movements tend to synchronize across sectors, which suggest our corrected MG estimates should yield even lower measures of persistence.[46] This is confirmed by both our SURE and CCE estimates.[47]

Allowing for cross-sectional dependence through the use of the MG SURE estimator lowers the point estimate of the half-life of relative prices to below two years (22 months), estimated with precision since our confidence interval ranges from 17 to 27 months. This is significantly below the "consensus view".[48] However, implementing the SURE estimate requires (arbitrary) truncation of our dataset, which otherwise contains too many cross-sections for its time dimension. As mentioned earlier, we chose to truncate our data so that they exactly reproduce the version reported on Charles Engel's website. On the other hand, the MG CCE estimator can be

[44] In Imbs et al. (2002), we used the Eurostat data corrected for some obvious mistakes, and a slightly shorter time series. On these data, the Random Coefficient model was to be preferred, and implied a half-life of fourteen months.

[45] To compute confidence intervals for our heterogeneous estimators, we use the mean coefficients to draw the residuals, and then perform sampling from the residuals themselves. This was suggested to us by Ron Smith.

[46] Actually, in our data, there is not a single instance of non-positive covariances between sectoral price residuals.

[47] As a robustness check, we confirm that all our results hold in Charles Engel's data set: the MG gives a half life of 25 months (confidence interval 9-31); the MG CCE a half life of 13 (confidence interval 9-24).

[48] Our SURE estimates are based on a sample where N is only marginally smaller than T. Given this dimensionality, that we should find such low and precise estimates suggests common effects are strongly present in our data.

implemented on our preferred dataset, and it implies a half-life point estimate of eleven months. The 95 percent confidence interval ranges from seven to twelve months. According to this estimator, an upper bound for the real exchange rate half-life is one year. This is strikingly lower than consensus results.

Figure 2 plots the frequency distribution of sector-specific persistence as measured by the largest auto-regressive root in each sector of our sample.[49] The Figure substantiates the empirical importance of heterogeneity. Some sectors do indeed display high persistence, confirming once again that our results do not require nor imply that all goods prices should revert to parity rapidly. What matters is the span of sectoral persistence estimates, which in our data is wide enough to induce a large positive bias in aggregate persistence estimates. The Figure illustrates graphically the roots implied by the Mean Group and the Fixed Effects estimates, corrected in both cases for correlated residuals. The heterogeneous estimator implies a root of 0.95, indeed where the mass of sectoral persistence is centered. The standard fixed effect estimator, in contrast, implies a root closer to 0.99. This discrepancy is the key to our result.

5.2.2 Monte Carlo Experiments

This section explore robustness along two dimensions. First, we compare the abilities various estimators have in capturing the heterogeneity bias. This is particularly relevant, for it enables us to compare standard panels to heterogeneous estimators on the one hand, and standard MG to its SURE and CCE variants on the other. Second, we let the extent of heterogeneity and persistence vary between plausible bounds and ask how the magnitude of the bias responds.

In Figure 3, we ask how the standard panel estimators would have performed if the data generating process were the one we estimated using the MG-CCE. Our focus is on the bias affecting the first autoregressive parameter. The figure makes it clear that the MG CCE estimator has satisfactory properties. The only other estimator that appears to be consistent is the MG SURE estimator. All other estimators have a large positive bias. In the cases of Fixed Effects and the Anderson-Hsiao variants, this can be traced back to the failure to account for dynamic heterogeneity. Allowing for cross-sectional error correlation does not appear to improve the FE estimator's properties either. Thus, any homogeneous estimator will tend to induce a bias in persistence estimates, in the presence of heterogeneity in the dynamics of the panel units - whether correlated residuals are accounted for or not. The uncorrected MG estimator also appears positively biased, though marginally less than FE. This bias stems from positively correlated residuals, *i.e.* positive realizations of σ_{ij} in Δ. Finally, the Figure shows the presence of a large positive bias in the aggregate estimator as well, confirming that *sectoral* heterogeneity translates into an *aggregate* bias.

[49] The estimates correspond to the AR(12) specification implied by the data in Table 3.

In Figure 4, we use a wide range of alternative data generating processes to illustrate the magnitude of the aggregation bias along two dimensions: the underlying persistence of the data generating process, and the underlying heterogeneity. As predicted by theory, the aggregation bias increases with the extent of heterogeneity, as measured by the cross-sectional variance of η_i, irrespective of the estimator implemented. Further, the aggregation bias continues to dominate even at high levels of persistence[50]. We come back to this point at length in Section 6.

5.2.3 The PPP puzzle

The point estimates for all our heterogeneous estimators are significantly below consensus results in the literature, most strikingly when correlations in residuals are taken into account. In all cases, our confidence intervals for real exchange rates half-life are entirely below the consensus of three to five years. Our estimates break the consensus, but, rather surprisingly, at the lower end of its range. We now explore whether our corrected estimates are actually consistent with theory.

Chari, Kehoe and McGrattan (2002) investigate whether a calibrated international business cycle model with plausible nominal rigidities can reproduce the persistence and volatility of the real exchange rate. With price stickiness assumed to be one-year long, their model implies a first-order autocorrelation for the HP-filtered real exchange rate equal to 0.62 (with a standard deviation of 0.08). This is significantly below the first order autocorrelation coefficient of 0.83 observed in (HP-filtered) US data. Even under alternative preference specifications, further shocks, or alternative monetary policy rules, their simulated estimate never exceeds 0.77. Chari, Kehoe and McGrattan label the inability of standard models to reproduce the observed persistence in real exchange rates the *persistence anomaly*. Our estimates resolve this anomaly.

We simulate our estimated process, and use the HP filter to detrend the resulting series, just as Chari, Kehoe and McGrattan HP-filter the series implied by their model. Using our MG-CCE estimates, we find an autoregressive coefficient of 0.597 (with a standard deviation of 0.073) assuming that the data is sampled at a point in time.[51] The same exercise using the MG SURE estimates implies a first order autoregressive coefficient of 0.589 with a standard error of 0.160.[52] The model in Chari, Kehoe and McGrattan is constructed with one sector only, so logic would have it that the empirical persistence estimates used to assess their model should be purged from

[50] We stress that none of the stochastic processes entering our simulations have autoregressive coefficients above one.

[51] Our MG-CCE estimates correspond to monthly data, so we transformed the simulated data into quarterly data by point-in-time sampling - just like the actual data. These data were then HP-filtered with a smoothing parameter of 1600 as in Chari, Kehoe and McGrattan (2002). We repeated this 1000 times for samples with 160 quarterly observations.

[52] The MG estimates imply a first order autoregressive coefficient of 0.766 with a standard error of 0.075. Even this includes the model moments in a 95 percent confidence interval.

sectoral heterogeneity. Indeed, the model has no trouble reproducing the data, once dynamic heterogeneity is accounted for. Hence there is a sense in which the PPP puzzle is no more.

Our explanation does not preclude the existence of other, different, aggregation biases. For instance, the half-life is defined as the number of periods necessary for half the initial effect of a given shock to disappear. This may be misleading in the presence of heterogeneity in the initial effects of the shock. In particular, if some sectoral relative prices rise while others decrease upon impact, but rise subsequently, the initial change in the aggregate relative price will tend to be small. Aggregate half-life is then bound to be large given the small size of the initial effect, irrespective of the persistence properties in the goods-specific relative prices. Likewise, temporal aggregation biases may also be at play, as described in Taylor (2001).[53,54]

6 Robustness Checks

In this section we investigate the robustness of our findings. We first assess the importance of measurement error in sectoral data. We then evaluate the importance of another (attenuating) bias, recently emphasized in the empirical exchange rate literature.[55]

6.1 Errors in Variables

There is a presumption that measurement error is more prevalent in sectoral data than in the aggregate. Indeed, if errors are uncorrelated across sectors, they tend to average away in the aggregate, and the resulting attenuating bias that may arise from examining disaggregated data might explain the discrepancy we just documented. However, as we illustrated in Section 3.2 and confirmed in Figure 2, there are more than a few measures of sectoral persistence estimated

[53] Several recent papers have also studied the impact of non linearities on the estimation of half-lives. For instance, Michael, Nobay and Peel (1997), Taylor and Peel (2000), Taylor, Peel and Sarno (2001), Kilian and Taylor (2003) and Imbs, Mumtaz, Ravn and Rey (2003).

[54] Lewbel and Ng (2003) develop an application where heterogeneity may induce persistence in aggregate demand estimations. The residuals in systems of aggregate demand equations are often found to be highly persistence, sometimes leading to the rejection of Almost Ideal Demand Systems. Lewbel and Ng show that aggregation over heterogeneous households in a slowly moving population may lead to near non-stationarity of the aggregate fixed effects, and thus creates persistence.

[55] We checked robustness along several other dimensions, but do not report all results. For instance, we used GBP as our anchor currency. Then, our Mean Group based half-life point estimate was seventeen months, down to fourteen months when corrected using CCE. As regards the choice of lag length, it is efficient to let the data decide. We did nevertheless verify that our results obtain even if we constrain the lag length to be the same across estimators. For all lag lengths between one and twenty, the aggregate fixed effect half-life estimates are significantly larger than Mean Group estimates, which in turn are significantly larger than MG corrected for CCE. For instance, with *one* lag, the FE half-life equals 36 months, MG's is 27 months and MG-CCE's is 16 months.

above the aggregate. We do not observe systematically lower half-lives at the sectoral level, and this casts doubt on the alternative explanation right at the outset. We do however also address the issue in a classic econometric manner. In the absence of measurement error, the OLS estimator of persistence, ρ^{OLS} and an instrumental variable estimator ρ^{TSLS} are both consistent, and the OLS estimator is efficient. However, in the presence of measurement error, the OLS estimator is inconsistent. Therefore, $plim\ (\rho^{OLS} - \rho^{TSLS})$ should be non zero in the presence of measurement errors. We perform a Hausman test along those lines, but take into account parameter heterogeneity. In particular, we carry out these tests at the sectoral level for each of the cross-sectional units.

Let q_{it}^{*} denote the observed value of the sectoral real exchange rate and q_{it} its true value. u_{it} denotes measurement error. The model is given by

$$q_{ict}^{*} = \gamma_c + \sum_{k=1}^{K} \rho_{ik}\ q_{ict-k}^{*} + \nu_{ict}$$

where $q_{ict}^{*} = q_{ict} + u_{ict}$ and $\nu_{ict} = -u_{ict} + \sum_{k=1}^{K} \rho_{ik}\ u_{ict-k} + \varepsilon_{ict}$. The lag structure of the model implies that $\{q_{ict}, .., q_{ict-K}\}$ are correlated with the error term $\nu_{ict.}$ Appropriate instruments for the TSLS estimate are therefore $\{q_{it-K-1}, ..., q_{it-2K}\}$. Hausman tests (available upon request) indicate the null hypothesis that OLS is consistent is rejected for only one of the 204 series. This makes it doubtful an errors-in-variables bias be relevant in our data.

6.2 Downward Bias in OLS for Highly Persistent Processes

For highly persistent autoregressive processes, it is well known that least squares estimators may be biased downward in small samples.[56] Since this is an attenuating bias, it could counter our claim that the PPP puzzle is due to a heterogeneity bias. Furthermore, the least squares bias may persist even as the cross-sectional dimension of our panel, N, rises to infinity.[57] The idea that the least squares bias in small samples may be quantitatively important has recently received considerable attention in the PPP literature. Murray and Papell (2002a) apply the approximately median bias-corrected estimator proposed by Andrews and Chen (1994) to study the persistence of real exchange rates. Their estimates of the (approximately median unbiased) confidence intervals are so wide that they conclude the data are basically uninformative about the half-life of the real exchange rate.[58] Murray and Papell (2002b) apply similar methods to a panel of real exchange rates. Their best corrected half-lives estimates lie exactly in the range of

[56] See for instance Hurvicz (1950) or Orcutt (1948).

[57] See Nickell (1981).

[58] Qualitatively similar results have also been obtained by Kilian and Zha (2002) using a different methodology. See also Elliott and Stock (2001).

three to five years, right back in the consensus view.[59] Their conclusion is that 'panels do not help solve the purchasing power parity puzzle'. Finally Murray and Papell (2002c) apply again the Andrews-Chen correction to a single exchange rate series (dollar-sterling), and argue that previous results were misguided. Once corrected, their estimates range from four and a half to eleven years depending on the number of lags included. Their 95 percent confidence intervals range between three to five years and infinity. They conclude that the purchasing power parity puzzle is even worse than usually believed. On the other hand, Rossi (2003) applies local-to-unity asymptotic theory to construct confidence intervals for the half life. These are robust to highly persistent data in small samples. Interestingly, although the upper bounds on the bias reduced real exchange rate persistence are still high, she finds that the lower bounds are low (around four to eight months), and her confidence interval does not exclude the consensus view.

The potential attenuating bias might be important in the present context, as it could contribute to explaining our surprisingly low estimates. Very little is known, however, about the joint effects of the heterogeneity bias we highlight, and the small sample bias stressed in the literature. In an interesting paper, Choi, Mark and Sul (2003) make an attempt at evaluating the relative importance of these two biases in the context of simulations.[60] They implement a Monte Carlo experiment with heterogeneous dynamics, and find a tendency for the overall bias in fixed effects estimators to be negative. They conclude that, under their parametrization, the small sample bias dominates. There are three reasons why our estimates do not fall directly victim to these *simulation* results. First, their data generating processes do not allow for the common correlated effects that we find are important in our data, and pertinent for our results. Second, the dominance of small sample bias appears sensitive to the parametrization of heterogeneity and the length of the artificial data. Third, the results of the Monte Carlo simulations are sensitive to initial conditions.[61] As far as *estimations* are concerned, we stress again that Choi, Mark and Sul (2003) never implement their correction methods directly on *heterogeneous* data. They carefully test for heterogeneity in their panel of aggregate exchange rates and fail to reject homogeneity. Their results are in this respect entirely similar to our Section 3, where we also consider a panel of *aggregate* real exchange rates. In a perfectly consistent way, they then implement small sample bias corrections on *homogeneous* estimators. In constrast, we have to implement bias correction techniques in the presence of heterogeneous dynamics in our panel of sectoral real exchange rates.

---

[59] This is also the conclusion in Cashin and McDermott (2003), who use the same type of near unit root correction method, but also allow for a moving average error structure. They show that real exchange rates half-lives remain firmly - and significantly - within the Rogoff consensus range, even after correcting for the bias.

[60] Most of Choi, Mark and Sul's (2003) analysis is centered around evaluating the relative importance of the small sample and temporal aggregation biases in actual data. Homogeneity cannot be rejected in their data, and they actually use pooled data to implement their bias correction method. Thus, the only conclusion they draw from actual data is that the small sample bias dominates the temporal aggregation one.

[61] See IMRR (2004) for more details.

In Imbs, Mumtaz, Ravn and Rey (2004) (henceforth IMRR (2004)) we examine the properties of various bias reduction techniques on the basis of Monte Carlo experiments. Importantly, we do allow for common correlated components in the data generating process. We show that ignoring these effects actually results in a serious *positive* bias in *corrected* half-life estimates based on the MG estimator.[62] In other words, the simulation results in Choi, Mark and Sul (2003) may overstate the importance of small sample bias *in our data*, which have strongly correlated residuals. Their simulation results cannot be taken as an indication that small sample bias always dominates the heterogeneity bias (and they certainly do not claim it does).

Our simulations apply two versions of the Kilian (1998) 'bootstrap-after-bootstrap' procedure to correct the half-life estimated on the basis of the MG estimator (as well as its SURE and CCE refinements). We first calculate an unbiased estimate of the half-life on the basis of the corrected autoregressive coefficients, an *indirect* approach followed Chen and Engel (2004). Alternatively, we correct the estimated half-life *directly* using the bootstrap algorithm. We find that the latter method outperforms the former, a result related to Pesaran and Zhao (1999) who argue the correction should be directly applied onto the object of interest in heterogeneous dynamic panels.[63]

In Table 4, we report half life estimates that were corrected for small sample bias using Kilian's method on our panel of disaggregated relative prices.[64] We implemented both the direct and indirect approaches. For the standard MG estimates, the corrected half-life is higher and has wider confidence intervals, especially if computed using the indirect approach. This is entirely consistent with the Monte Carlo experiments in IMRR (2004), who show that failing to control for common correlated effects biases the small sample bias corrected MG estimates upwards. When we turn to the proper estimator, the MG-CCE model, our corrected estimates point to a half-life of eighteen months. This constitutes a marginal increase relative to our original (uncorrected) estimate of eleven months.[65] The 95 percent confidence interval is narrow - spanning eleven to twenty-eight months - and still excludes the 'consensus view'. Small sample bias is potentially a relevant concern in our context, but our results remain largely unchanged when our estimates are corrected. In our data, the dynamic heterogeneity bias dominates.

[62] The same is true for the So and Shin (1999) method.

[63] For more details see IMRR (2004).

[64] Phillips and Sul (2003) discuss the properties of an alternative bias reduction method, the Panel Feasible Generalized Median Unbiased estimator, which is applicable to heterogeneous panels. This method relies upon applying a median unbiased correction to a SUR panel estimator. We chose to apply the Kilian's (1998) bootstrap-after-bootstrap procedure given the large cross-sectional dimension of our data.

[65] This result is based on the 'direct approach'. As far as the MG CCE estimator is concerned, the 'indirect approach' gives similar results, with a corrected half-life estimate of twenty months. The MG-SUR gives a half life of twenty seven months when the direct approach is used. We show in IMRR (2004) that this estimate is also biased upwards.

7 Summary and Conclusions

We have argued a simple and plausible mechanism can explain the difficulty in reconciling real exchange rate dynamics with the predictions of models with realistic impediments to price adjustment. Our argument rests on the possibility that relative price dynamics differ across the goods composing the real exchange rate. If for instance goods differ substantially in their tradability, the degree of competition or transportation costs, there is little reason to expect a priori that relative prices converge homogeneously. Under this premise, the paper shows that the persistence of the real exchange rate should be interpreted as a biased estimate of the average persistence in relative prices. Under empirically plausible and general conditions, the bias is positive. Our results do not imply nor require that disaggregated relative prices all converge faster than the aggregate real exchange rate. If relative prices all converged quickly, so would the aggregate and there would be no bias, nor, indeed, any PPP puzzle. In reality some prices converge slowly and others do quickly. The "aggregation bias" comes precisely from this heterogeneity in dynamics.

Our data reproduce consensual estimates for real exchange rate persistence when standard panel techniques are implemented. They do as well when price dynamics are constrained to be identical across different goods. But this constraint is actually rejected in our data, and with heterogeneous dynamics, our measure of average persistence falls dramatically. Our estimates point to a half-life for relative prices down to eleven months, with a tight confidence interval that excludes Rogoff's three to five year "consensus view".

Our results are robust. They withstand numerous alterations, truncations or variations to our dataset. They cannot be explained by the presence of measurement errors. They survive small sample bias corrections. Our corrected persistence estimates are only moderately larger than eleven months, up to eighteen months. And the confidence interval remains significantly below the "consensus view".

Our findings have potentially important implications. First, our estimates for the average persistence in relative prices can be reproduced in models with realistic price rigidities. Thus, in this sense, we solve the PPP puzzle. We bridge the gap between theory and evidence in this area of international macroeconomics and show how findings on persistence based on disaggregated price data do not necessarily translate into similar results in aggregate data. Second, we underline the importance of heterogeneity at the microeconomic level for understanding macroeconomic aggregate phenomena. When microeconomic heterogeneity is purged from the data used to evaluate them, macroeconomic models perform better, at least as far as the real exchange rate is concerned. By the same token, whether models with non-trivial sectoral heterogeneity are capable of mimicking aggregate data is in our opinion an exciting area for future research. Should such models prove unsuccessful at generating persistent real exchange rates, there would indeed be a PPP puzzle.

Appendix A1. Approximating the Real Exchange Rate

The CPI based real exchange rate is defined as:

$$Q_t = \frac{e_t P_{bt}}{P_{at}}$$

where e_t is the bilateral nominal exchange rate between country a and country b, P_{it} is the CPI of country i at date t. The CPI is a Laspeyres index defined as:

$$P_{it} = \frac{\sum_{s=1}^{M} P_{ist} \cdot q_{is0}}{\sum_{s=1}^{M} P_{ist} \cdot q_{is0}}$$

where P_{ist} is the price of good s in country i at date t, q_{ist} is the consumption of good s in country i at date t and "0" denotes the base year. This can be re-written as:

$$\begin{aligned}
P_{it} &= \frac{\sum_{s=1}^{M} P_{ist} \cdot q_{is0}}{\sum_{s=1}^{M} P_{is0} \cdot q_{is0}} = \frac{\sum_{s=1}^{M} \frac{P_{ist}}{P_{is0}} \cdot P_{is0} \cdot q_{is0}}{\sum_{s=1}^{M} P_{is0} \cdot q_{is0}} \\
&= \sum_{s=1}^{M} \frac{P_{ist}}{P_{is0}} \frac{P_{is0} \cdot q_{is0}}{\sum_{i=1}^{N} P_{is0} \cdot q_{is0}} \Rightarrow \\
P_{it} &= \sum_{s=1}^{M} \frac{P_{ist}}{P_{is0}} w_{is0} \\
w_{is0} &= \frac{P_{is0} \cdot q_{is0}}{\sum_{i=1}^{N} P_{is0} \cdot q_{is0}}
\end{aligned}$$

where w_{is0} is the associated CPI weight.

Thus, we can write the real exchange rate as:

$$Q_t = \frac{e_t \sum_s \left(\frac{p_{bst}}{p_{bs0}}\right) w_{bs0}}{\sum_s \left(\frac{p_{ast}}{p_{as0}}\right) w_{as0}}$$

Taking a log-linear approximation gives us that:

$$\begin{aligned}
q_t &= \ln Q_t \simeq \ln Q_0 + \widehat{q}_t \\
&= q_0 + \sum_s w_{bs0} q_{st} - \sum_s \left(w_{bs0} - w_{as0}\right) q_{st} \\
q_{st} &= \ln \frac{e_t p_{bst}}{p_{ast}}
\end{aligned}$$

where $\widehat{q}_t$ is the log differential of q_t evaluated at q_0.

Hence, if the expenditure weights are similar, apart from a constant term which corresponds to the initial value of the real exchange rate, the logarithm of the real exchange rate is approximately equal to a weighted sum of the log of relative prices.

Appendix A2. Bias: Analytics

We first derive the (asymptotic properties for the) least squares estimate of the first-order autoregressive coefficient of q_t, given by $\rho^Q = E(q_t q_{t-1}) / E(q_t^2)$.[66] Let us first derive the expression of the bias assuming constant weights to economize on notations ($\omega_i = \omega_j$ for all i). We have:

$$E\left(q_t^2\right)_{(\omega_i=\omega_j)} = \frac{1}{N^2}\left(\sum_{i=1}^{N} \sigma_{q_i}^2 + 2\sum_{i<j}^{N} \sigma_{q_i,q_j}\right)$$

$$E\left(q_t q_{t-1}\right)_{(\omega_i=\omega_j)} = \frac{1}{N^2}\left(\sum_{i=1}^{N}\left(\rho_i \sigma_{q_i}^2 + \sum_{i<j}^{N}\left(\rho_i+\rho_j\right)\sigma_{q_i,q_j}\right)\right)$$

$$\begin{aligned}
\rho^Q_{(\omega_i=\omega_j)} &= \frac{\sum_{i=1}^N \rho_i \sigma_{q_i}^2 + \sum_{i<j}^N \left(\rho_i+\rho_j\right)\sigma_{q_i,q_j}}{\sum_{i=1}^N\left(\sigma_{q_i}^2 + 2\sum_{i<j}^N \sigma_{q_i,q_j}\right)} \\
&= \rho - \rho + \frac{\sum_{i=1}^N \rho_i \sigma_{q_i}^2 + \sum_{i<j}^N \left(\rho_i+\rho_j\right)\sigma_{q_i,q_j}}{\sum_{i=1}^N\left(\sigma_{q_i}^2 + 2\sum_{i<j}^N \sigma_{q_i,q_j}\right)} \\
&= \rho + \frac{\sum_{i=1}^N \left(\rho_i-\rho\right)\sigma_{q_i}^2 + \sum_{i<j}^N \left[\left(\rho_i-\rho\right)\sigma_{q_i,q_j} + \left(\rho_j-\rho\right)\sigma_{q_i,q_j}\right]}{\sum_{i=1}^N\left(\sigma_{q_i}^2 + 2\sum_{i<j}^N \sigma_{q_i,q_j}\right)} \\
&= \rho + \frac{\sum_{i=1}^N \frac{\rho_i-\rho}{1-\rho_i^2}\sigma^2_{\ i} + \sum_{i<j}^N \left(\frac{\rho_i-\rho}{1-\rho_i\rho_j}\sigma_{\ ij} + \frac{\rho_j-\rho}{1-\rho_i\rho_j}\sigma_{\ ij}\right)}{\sum_{i=1}^N\left(\sigma_{q_i}^2 + 2\sum_{i<j}^N \sigma_{q_i,q_j}\right)}
\end{aligned}$$

Reintroducing non equal weights (which amounts to multiplying variance by ω_i^2 and covariances by $\omega_i\omega_j$), it follows trivially that

$$\rho^Q = \rho + \frac{\sum_{i=1}^N \frac{(\rho_i-\rho)\omega_i^2}{1-\rho_i^2}\sigma^2_{\ i} + \sum_{i<j}^N \left(\frac{(\rho_i-\rho)\omega_i\omega_j}{1-\rho_i\rho_j}\sigma_{\ ij} + \frac{\left(\rho_j-\rho\right)\omega_i\omega_j}{1-\rho_i\rho_j}\sigma_{\ ij}\right)}{\left(\sum_{i=1}^N\left(\omega_i^2\sigma_{q_i}^2 + 2\sum_{i<j}^N \omega_i\omega_j\sigma_{q_i,q_j}\right)\right)}$$

which we rewrite as[67]

$$\rho^Q = \rho + \Delta$$

with $\Delta = \left[\sum_{i=1}^N \frac{\omega_i^2}{1-\rho_i^2}\sigma^2_{\ i} + \sum_{i\neq j}^N \frac{\omega_i\omega_j\sigma_{ij}}{1-\rho_i\rho_j}\right] \frac{(\rho_i-\rho)}{\sum_{i=1}^N\left(\frac{\omega_i^2}{1-\rho_i^2}\sigma^2_{\ i} + \sum_{i\neq j}^N \frac{\omega_i\omega_j\sigma_{ij}}{1-\rho_i\rho_j}\right)}$. QED

[66] This derivation was first presented in Imbs, Mumtaz, Ravn and Rey (2003).

[67] This last expression is the same as the one presented in Chen and Engel (2004), who also generalize slightly our expression for the bias by including weights.

Appendix B. Data Coverage

Table B1

	BE	DE	DK	ES	IT	FR
Bread	81-1:95-12	81-1:95-7	81-1:95-12	81-1:95-12	81-1:95-12	81-1:95-12
Meat	81-1:95-12	81-1:95-7	81-1:95-12	81-1:95-12	81-1:95-12	81-1:95-12
Dairy	81-1:95-12	81-1:95-7	81-1:95-12	81-1:95-12	81-1:95-12	81-1:95-12
Fruits	81-1:95-12	81-1:95-7	81-1:95-12	81-1:95-12	81-1:95-12	81-1:95-12
Tobacco	81-1:95-12	81-1:95-7	81-1:95-12	81-1:95-12	81-1:9:1-12	81-1:95-12
Alcohol	81-1:95-12	81-1:95-7	81-1:95-12	81-1:95-12	81-1:95-12	81-1:95-12
Clothing	81-1:95-12	81-1:95-7	81-1:95-12	81-1:95-12	81-1:95-12	81-1:95-12
Footwear	81-1:95-12	81-1:95-7	81-1:95-12	81-1:95-12	81-1:95-12	81-1:95-12
Rents	81-1:95-12	81-1:95-7	81-1:95-12	81-1:95-12	81-1:95-12	81-1:95-12
Fuel	81-1:95-12	81-1:95-7	81-1:95-10	81-1:95-12	81-1:95-12	81-1:95-12
Furnit.	81-1:94-10	81-1:94-10	81-1:94-10	81-1:94-10	81-1:94-9	81-1:94-10
Dom. Appl.	81-1:94-10	81-1:94-10	81-1:94-10	81-1:94-10	81-1:94-9	81-1:94-10
Vehicles	81-1:95-12	81-1:95-7	81-1:95-12	81-1:95-12	81-1:95-12	81-1:95-12
Pub. Transp	81-1:95-12	81-1:95-7	81-1:95-12	81-1:95-12	81-1:95-12	81-1:95-12
Comm.	81-1:95-12	81-1:95-7	81-1:95-12	81-1:95-12	81-1:95-12	81-1:95-12
Sound	81-1:95-12	81-1:95-7	81-1:95-12	81-1:95-12	81-1:95-12	81-1:95-12
Leisure	81-1:95-9	81-1:95-7	81-1:95-9	81-1:95-9	81-1:95-9	81-1:95-9
Books	81-1:95-9	81-1:95-7	81-1:95-9	81-1:95-9	81-1:95-9	81-1:95-9
Hotels	81-1:95-9	81-1:95-7	81-1:95-9	81-1:95-9	81-1:95-9	81-1:95-9

Table B1: Continued

	GR	NL	PT	FI	UK	US
Bread	81-1:95-12	81-1:95-12	81-1:95-12	85-1:95-5	81-1:95-12	81-1:95-12
Meat	81-1:95-12	81-1:95-12	81-1:95-12	85-1:95-5	81-1:95-12	81-1:95-12
Dairy	81-1:95-12	81-1:95-12	81-1:95-12	85-1:95-5	81-1:95-12	81-1:95-12
Fruits	81-1:95-12	81-1:95-12	81-1:95-12	85-1:95-5	81-1:95-12	81-1:95-12
Tobacco	81-1:95-12	81-1:95-12	81-1:95-12	85-1:95-5	81-1:95-12	81-1:95-12
Alcohol	81-1:95-12	81-1:95-12	81-1:95-11	85-1:95-5	81-1:95-12	81-1:95-12
Clothing	81-1:95-12	81-1:95-12	81-1:95-12	85-1:95-5	81-1:95-12	81-1:95-12
Footwear	81-1:95-12	81-1:95-12	81-1:95-12	85-1:95-5	81-1:95-12	81-1:95-12
Rents	81-1:95-12	81-1:95-12	na	85-1:95-5	81-1:95-12	81-1:95-12
Fuel	81-1:95-12	81-1:95-12	81-1:95-12	na	81-1:95-12	81-1:95-12
Furnit.	81-1:94-10	81-1:94-10	81-1:94-10	na	81-1:94-10	81-1:95-12
Dom. Appl.	81-1:94-10	81-1:94-10	81-1:94-10	na	81-1:94-10	81-1:94-10
Vehicles	81-1:95-12	81-1:95-12	81-1:95-12	85-1:95-5	81-1:95-12	81-1:95-12
Pub. Transp	81-1:95-12	81-1:95-12	81-1:95-12	85-1:95-5	81-1:95-12	81-1:95-12
Comm.	81-1:95-12	81-1:95-12	81-1:95-12	85-1:95-5	81-1:95-12	81-1:95-12
Sound	81-1:95-12	81-1:95-12	na	85-1:95-5	81-1:95-12	81-1:95-12
Leisure	81-1:95-9	81-1:95-9	81-1:95-9	85-1:95-5	81-1:95-9	81-1:95-9
Books	81-1:95-9	81-1:95-9	81-1:95-9	85-1:95-5	81-1:95-9	81-1:95-9
Hotels	81-1:95-9	81-1:95-9	81-1:95-9	85-1:95-5	81-1:95-9	81-1:95-9

References

[1] Abuaf, Niso and Jorion, Philippe, 1990, "Purchasing Power Parity in the Long Run." *Journal of Finance*, 45(1), pp. 155-74.

[2] Anderson, T.W and Hsiao, C, 1982, "Formulation and Estimation of Dynamic Models using Panel Data." *Journal of Econometrics,* 18, pp. 47-82.

[3] Andrews, Donald W.K, 1993, "Exactly Median-unbiased Estimation of First Order Autoregressive/Unit Root Models." *Econometrica,* 61, pp. 139–165.

[4] Andrews, Donald W.K. and Chen, Hong-Yuan, 1994, "Approximately Median-Unbiased Estimation of Autoregressive Models." *Journal of Business and Economic Statistics*, Vol. 12 (2), pp. 187-204.

[5] Arellano, Manuel and Bond, Stephen, 1991, "Some Tests of Specification for Panel Data: Monte-Carlo Evidence and an Application to Employment Data". *Review of Economic Studies* 58, pp. 277-297.

[6] Bergin, Paul R. and Feenstra, Robert C., 2001, "Pricing-to-Market, Staggered Contracts, and Real Exchange Rate Persistence." *Journal of International Economics* 54(2), pp. 333-59.

[7] Bils, Mark and Peter Klenow, 2002, "Some Evidence on the Importance of Sticky Prices", *NBER Working Paper* No 9069.

[8] Blanchard, Olivier, 1987,. "Aggregate and Individual Price Adjustment", *Brookings Papers on Economic Activity* 1, pp.57-109.

[9] Boyd, Derick, and Ron Smith, 1999, "Testing for Purchasing Power Parity: Econometric Issues and an Application to Developing Countries", *Manchester School*, 67(3), pp.287-303.

[10] Brown, Bryan W. and Newey, Whitney K, 2002, "GMM, Efficient Bootstrapping and Improved Inference", *Journal of Business and Economics Statistics* 20, 507-17.

[11] Campa, Jose and Goldberg, Linda, 2002, "Exchange Rate Pass-Through into Import Prices: A Macro or Micro Phenomenon?" *NBER Working Paper 8934.*

[12] Cashin Paul and John McDermott, 2003, "An Unbiased Appraisal of Purchasing Power Parity", *IMF Staff Paper*, Vol. 50 (3).

[13] Chari, V.V, Kehoe, Patrick and McGrattan, Ellen, 2002,. "Can Sticky Price Models Generate Volatile and Persistent Real Exchange Rates?", *Review of Economic Studies* 69(3), 533–563.

[14] Chen, Shiu-Sheng, and Charles Engel, 2004, "Does 'Aggregation Bias' Explain the PPP Puzzle?", *NBER Working Paper* 10304.

[15] Cheung Yin-Wong , Menzie Chinn and Eiji Fujii, 2001, "Market Structure and the Persistence of Sectoral Real Exchange Rates", *International Journal of Finance and Economics* 6(2) : 95-114.

[16] Choi, Chi-Young, Nelson C. Mark, and Donggyu Sul, 2003, "The Dominance of Downward Bias in Half-Life Estimates of PPP Deviations", *Mimeo* Notre Dame.

[17] Coakley Jerry , Robert Flood and Mark Taylor, 2002,. "Long Run Purchasing Power Parity: the First Tests", *Mimeo* IMF.

[18] Crucini, Mario J., Telmer, Chris J. and Zachariadis, Marios, 2001,. "Understanding European Real Exchange Rates." *Mimeo* Vanderbilt University.

[19] Crucini, Mario and Mototsugu Shintani, 2002, "Persistence in Law-of-One-Price Deviations: Evidence from Micro-data" *Mimeo* Vanderbilt University.

[20] Diebold, Francis X., Husted, Steve and Rush, Mark, 1991,. "Real Exchange Rates Under the Gold Standard." *Journal of Political Economy* 99(6), pp. 1252-71.

[21] Elliott, Graham, and Stock, James, 2001, "Confidence Intervals for Autoregressive Coefficients Near One." *Journal of Econometrics* 103, pp. 155-181.

[22] Engel, Charles, 2000a, "Long-Run PPP May Not Hold After All." *Journal of International Economics*, 51, pp.243-73.

[23] Engel, Charles, 2000b, "Local Currency Pricing and the Choice of Exchange-Rate Regime." *European Economic Review*, 44, no 8, pp.1449-72.

[24] Frankel, Jeffrey A. and Rose, Andrew K., 1996, "A Panel Project on Purchasing Power Parity: Mean Reversion Within and Between Countries." *Journal of International Economics* 40, pp. 209-25.

[25] Froot, Kenneth and Rogoff, Kenneth, 1995, "Perspectives on PPP and Long-Run Exchange Rates." *Handbook of International Economics*,. Eds. G. Grossman and K. Rogoff, Vol.3, pp 1647-1688, North-Holland, New York.

[26] Giovannini, Alberto, 1988, "Exchange Rates and Traded Goods Prices." *Journal of International Economics* 24(1/2), pp. 45-68.

[27] Glen, Jack D., 1992, "Real Exchange Rates in the Short, Medium, and Long Run." *Journal of International Economics* 33(1/2), pp. 147-66.

[28] Goldberg, Pinelopi K. and Knetter, Michael M., 1997, "Goods Prices and Exchange Rates: What Have We Learned?" *Journal of Economic Literature* 35, pp. 1243-72.

[29] Goldberg, Pinelopi and Verboven, Frank, 2001, "The Evolution of Price Dispersion in European Car Markets". *Review of Economic Studies* October, pp. 811-848.

[30] Granger, Clive, 1980, "Long Memory Relationships and the Aggregation of Dynamic Models." *Journal of Econometrics*, 14, pp. 227-238.

[31] Granger, Clive and Morris, M., 1976, "Time Series Modelling and Interpretation." *Journal of the Royal Statistical Society*, A, 139, pp. 246-257.

[32] Haskel, Jonathan and Wolf, Holger, 2001, "Why Does the Law of One Price Fail? A Case Study." *Scandinavian Journal of Economics*, 103, pp. 545-558.

[33] Hildreth, Clifford and Houck, James P., 1968, "Some Estimators for a Linear Model with Random Coefficients." *Journal of the American Statistical Association* 63, pp. 584-95.

[34] Hurvicz, L., 1950, "Least Squares Bias in Time Series", in T. Koopmans (ed.): *Statistical Inference on Dynamic Economic Models*, New York: Wiley.

[35] Im, Kyung So, Pesaran, M. Hashem and Shin, Yongcheol, 1997, "Testing for Unit Roots in Heterogeneous Panels." *DAE Working Paper* No. 9526, University of Cambridge.

[36] Imbs, Jean, Mumtaz, Haroon, Ravn, Morten O. and Rey, Hélène, 2002, "PPP Strikes Back: Aggregation and the Real Exchange Rate.", December, *NBER Working Paper* 9372 .

[37] Imbs, Jean, Mumtaz, Haroon, Ravn, Morten O. and Rey, Hélène, 2003, "Non-Linearities and Real Exchange Rate Dynamics." *Journal of the European Economic Association* 1 (2-3), 639-49.

[38] Imbs, Jean, Mumtaz, Haroon, Ravn, Morten O. and Rey, Hélène, 2004, "Aggregation Bias Does Explain the PPP Puzzle", available at *http://faculty.london.edu/jimbs/, http://www.princeton.edu/~hrey/ and http://faculty.london.edu/mravn/.*

[39] Isard, Peter, 1977, "How Far Can We Push the Law of One Price?" *American Economic Review* 67(5), pp. 942-48.

[40] Kilian, Lutz, 1998, "Small-Sample Confidence Intervals for Impulse Response Functions", *Review of Economics and Statistics* 80(2), 218-30.

[41] Kilian, Lutz and Taylor, Mark, 2003, "Why is it so Difficult to Beat the Random Walk Forecast of Exchange Rates?" *Journal of International Economics,* 60(1), May, pp. 85-107.

[42] Kilian, Lutz and Zha, Tao, 2002, "Quantifying the Uncertainty about the Half-Life of Deviations from PPP." *Journal of Applied Econometrics*, Vol. 17, 2, pp. 107-125.

[43] Knetter, Michael M., 1989, "Price Discrimination by U.S. and German Exporters." *American Economic Review*, 79(1), pp. 198-210.

[44] Knetter, Michael M., 1993, "International Comparisons of Pricing-to-Market Behavior." *American Economic Review*, 83(3), pp. 473-86.

[45] Kollmann, Robert, 2001, “The Exchange Rate in a Dynamic-Optimizing Business Cycle Model with Nominal Rigidities: a Quantitative Investigation.” *Journal of International Economics*, 55, pp. 243-62

[46] Levin, Andrew and Lin, Chien-Fu., 1993, “Unit Root Tests in Heterogeneous Panels.” *UCSD Working Paper* No. 93-56.

[47] Lewbel, Arthur and Serena Ng, 2003, “Non-Stationary Aggregate Demand Systems and Heterogeneous Consumers", *Mimeo* University of Michigan.

[48] Lothian, James R., 1997, “Multi-Country Evidence on the Behavior of Purchasing Power Parity under the Current Float.” *Journal of International Money and Finance*, 16(1), pp. 19-35.

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

[50] Mark, Nelson, 1995, “Exchange Rate Fundamentals: Evidence on Long-Horizon Predictability.” *American Economic Review*, 85 (1), pp. 201-218.

[51] Michael, Panos, Nobay, A. Robert and Peel, David A., 1997, “Transactions Costs and Nonlinear Adjustment in Real Exchange Rates: An Empirical Investigation.” *Journal of Political Economy*, 105(4), pp. 862-79.

[52] Murray, Christian J. and Papell, David H., 2002a, “The Purchasing Power Parity Persistence Paradigm.” *Journal of International Economics*, 56, pp. 1-19.

[53] Murray, Christian J. and Papell, David H., 2002b, “Do Panels Help Solve the Purchasing Power Puzzle?” *Mimeo* University of Houston.

[54] Murray, Christian J. and Papell, David H., 2002c, “The Purchasing Power Parity Puzzle is Worse Than You Think: A Note on Long-Run Real Exchange Rate Behavior.” *Mimeo* University of Houston.

[55] Ng, Serena and Perron, Pierre, 2002, “PPP May not Hold Afterall: A Further Investigation”, *Annals of Economics and Finance*, 3, 41-64.

[56] Nickell, Stephen, 1981, “Biases in Dynamic Models with Fixed Effects”, *Econometrica*, 49, pp. 1417-46.

[57] Oh, Keun-Yeob, 1996, “Purchasing Power Parity and Unit Root Tests Using Panel Data.” *Journal of International Money and Finance*, 15(3), pp. 405-18.

[58] Obstfeld, Maurice, and Taylor, Alan M., 1997, "Nonlinear Aspects of Goods-Market Arbitrage and Adjustment: Heckscher's Commodity Points Revisited." *Journal of the Japanese and International Economies*, 11, pp. 441-79.

[59] Orcutt, Guy H., 1948, "A Study of the Autoregressive Nature of the Time Series Used for Tinbergen's Model of the Economic System of the United States", *Journal of the Royal Statistical Society* (B), 1-45.

[60] Parsley, David and Wei, Shang-Jin, 1996, "Convergence to the Law of One Price without Trade Barriers or Currency Fluctuations." *Quarterly Journal of Economics* (November), pp.1211-36.

[61] Parsley, David and Wei, Shang-Jin, 2003, "A Prism into the PPP Puzzles: The Microfoundations of Big Mac Real Exchange Rates", *Mimeo* IMF and Vanderbildt.

[62] Pesaran, M. Hashem, 2002, "Estimation and Inference in Large Heterogeneous Panels with Cross-Section Dependence." *University of Cambridge*, DAE Discussion Paper no. 0305.

[63] Pesaran, M. Hashem, 2003, "Lecture Notes", *University of Cambridge.*

[64] Pesaran, M. Hashem and Smith, Ron, 1995, "Estimating Long-Run Relationships From Dynamic Heterogeneous Panels." *Journal of Econometrics*, 68, pp. 79-113.

[65] Pesaran, M. Hashem and Z. Zhao, 1999, "Bias Reduction in Estimating Long-run Relationships from Dynamic Heterogeneous Panels" in C. Hsiao, K. Lahiri, L-F Lee and M.H. Pesaran (eds), *Analysis of Panels and Limited Dependent Variables*: A Volume in Honour of G S Maddala**,** Cambridge University Press, Cambridge, chapter 12, pp.297-321.

[66] Phillips, Peter C.B., and Donggyu Sul, 2003, "Dynamic Panel Estimation and Homogeneity Testing Under Cross Section Dependence", *Econometrics Journal* 6, 217-59.

[67] Pudney, Stephen E., 1978, "The Estimation and Testing of Some Error Components Models", *Mimeo* London School of Economics.

[68] Richardson, J. David, 1978, "Some Empirical Evidence on Commodity Arbitrage and the Law of One Price." *Journal of International Economics*, 8(2), pp. 341-51.

[69] Robertson, Donald and Symons, James, 1992, "Some Strange Properties of Panel Data Estimators." *Journal of Applied Econometrics*, 7(2), pp. 175-189.

[70] Rogers, John and Jenkins, Michael, 1995, "Haircuts or Hysteresis? Sources of Movements in Real Exchange Rates." *Journal of International Economics,* May, 38, pp. 339-60.

[71] Rogoff, Kenneth S., 1996, "The Purchasing Power Parity Puzzle." *Journal of Economic Literature*, 34, pp. 647-68.

[72] Rossi, Barbara, 2003, "Confidence Intervals for Half-life Deviations from Purchasing Power Parity." *Mimeo* Duke University.

[73] So, Beong Soo and Dong Wan Shin, 1999, "Recursive Mean Adjustment in Time-Series Inferences" *Statistics and Probability Letters*, vol. 43, pp. 65-73.

[74] Swamy, P.A.B.V., 1970, "Efficient Inference in a Random Coefficient Regression Model." *Econometrica*, 38, pp.311-323.

[75] Swamy, P.A.B.V., 1971, "Statistical Inference in Random Coefficient Regression Models." *Springer-Verlag*.

[76] Taylor, Alan 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, pp. 473–98.

[77] Taylor, Mark and Sarno, Lucio, 1998, "The Behavior of the Real Exchange Rate During the Bretton Woods Period". *Journal of International Economics*, 46, pp. 281-312.

[78] Taylor, Mark P. and Peel, David A., 2000, "Nonlinear Adjustment, Long-Run Equilibrium and Exchange Rate Fundamentals." *Journal of International Money and Finance*, 19, pp. 33-53.

[79] Taylor, Mark P., Peel, David A. and Sarno, Lucio, 2001, "Nonlinear Mean-Reversion in Real Exchange Rates: Towards a Solution to the Purchasing Power Parity." *International Economic Review*, 42(4), pp. 1015-42.

[80] Wu, Yangru, 1996, "Are Real Exchange Rates Nonstationary: Evidence from a Panel-Data Test." *Journal of Money, Credit, and Banking*, 28(1), pp. 54-63.

Unit Root tests

Table 1: Unit Root Tests

Test	Trend	$\ln\left(\frac{(P_i * e_{i,us})}{P_{us}}\right)$	$e_{i,us}$	$\ln\left(\frac{(P_{ij} * e_{i,us})}{P_{ujs}}\right)$
IPS ADF	no	-2.48050	-2.3638	5.41498
		[0.44723]	[0.009]	[0.000]
IPS ADF	yes	-2.85247	-4.1100	0.31558
		[0.00217]	[0.000]	[0.37616]
LL	no	-2.48050	-4.81754	-5.4429
		[0.0065]	[0.000]	[0.000]
LL	yes	-2.48041	-4.8138	-5.4339
		[0.00656]	[0.000]	[0.000]
LL^1	no	-40.29696	-6.09430	12.9538
		[0.0000]	[0.000]	[0.000]

Notes: P-values are in the parenthesis. . All test regressions include an intercept. IPS denotes Im, Pesaran and Shin and LL stands for Levin and Lin. LL^1 is Levin and Lin test that includes individual effects.The lag length for the IPS tests is set to 12

Aggregate Results

Table 2: Persistence Estimates using Aggregate Data

$q_{c,t} = \gamma_c + \sum_{p=1}^{P} \rho_p \ q_{c,t-p} + \varepsilon_t$					
Model	P	$\sum_{p=1}^{P} \rho_p$	Half-Life	LAR	CIR
Fixed Effects	18	0.98	46 $(31, 57)$	0.97 $(0.962, 0.981)$	64.38
Anderson-Hsiao	11	0.99	72 $(33, \infty)$	0.96 $(0.941, 1.05)$	109.68
Arrelano Bond	18	0.99	54 (46.75)	0.98 $(0.975, 0.989)$	75.57
$^{a}H0 : \rho_c = \rho$	-0.4046 (1.0000)		$^{c}H0 : E(\gamma_c, X) = 0$	25.856 (0.0021)	
$^{b}H0 : \rho_c = \rho$	70.96 (0.9999)		$^{d}H0 : \gamma_c = 0$	9.8714 (0.000)	

Notes: The estimates are based on real exchange rates from 11 countries over the period 1975:01-1996:12. The choice of P is based on general to specific lag selection procedure with a maximum lag of 20 for all models, except AH where it was restricted to 12. At each choice of P, the impulse response was examined and the specification was only selected if the IRF was continuous around 0.5. For the GMM estimator two lags of the levels of relative prices were used as instruments.The confidence intervals in the parenthesis were estimated using non-parametric bootstrap with 500 replications. Note that the bootstrap for the Arellano and Bond estimator was carried out using the methods described in Brown and Newey (2001). "LAR" denotes the largest autoregressive root. "CIR" denotes the cumulated impulse response. "a" is the Hausman test for homogeneity, while "b" denotes the Swami test for this hypothesis. "c" and "d" are the Hausman test for random effects and an F-test for fixed effects, respectively.

Disaggregated Results

Table 3: Persistence Estimates using Disaggregated Data

$q_{ict} = \gamma_c + \sum_{k=1}^{K} \rho_{ik}\ q_{ict-k} + e_{ict}$					
Model	P	$\sum_{k=1}^{K} \rho_{ik}$	Half-Life	LAR	CIR
Fixed Effects	12	0.98	36 $(21, 47)$	0.97 $(0.961, 0.981)$	46.71
Fixed Effects (SURE)	12	0.98	34 $(27, 43)$	0.97 $(0.958, 0.978)$	44.30
Fixed Effects (CCE)	12	0.99	58 $(10, 91)$	0.99 $(0.980, 0.995)$	104.20
Mean Group	19	0.97	26 $(14, 28)$	0.95 $(0.903, 0.973)$	33.15
Mean Group (SURE)	20	0.96	22 $(17, 27)$	0.96 $(0.945, 0.968)$	29.48
Mean Group (CCE)	12	0.95	11 $(7, 12)$	0.95 $(0.924, 0.963)$	20.51
$^{a}H0 : \rho_i = \rho$	98.15 (0.0000)		$^{d}H0 : E(\gamma_c, X) = 0$	14765 (0.000)	
$^{b}H0 : \rho_i = \rho$	4353.4 (0.0007)		$^{e}H0 : \gamma_c = 0$	2.1168 (0.000)	
$^{c}H0 : E(\eta_i, X) = 0$	485.02 (0.0022)		^{f}LM	2194698 (0.000)	

Notes: The estimates are based on relative prices on a maximum of 19 goods from 11 countries over the period 1981:01-1995:12. The choice of P is based on general to specific lag selection procedure with a maximum lag of 20 for all models, except AH where it was restricted to 12. At each choice of P, the impulse response was examined and the specification was only selected if the IRF was continuous around 0.5. The confidence intervals in the parenthesis were estimated using non-parametric bootstrap with 500 replications. "LAR" denotes the largest autoregressive root. "CIR" denotes the cumulated impulse response. "a" is the Hausman test for homogeneity (allowing for correlated residuals), while "b" denotes the Swami test for this hypothesis. "c" is the Pudney (1978) test for the null of no correlation between the random coefficients and the error term. "d" and "e" are the Hausman test for random effects and an F-test for fixed effects, respectively, while "f" is a Breusch-Pagan test for the diagonality of the covariance matrix.

Bias Correction

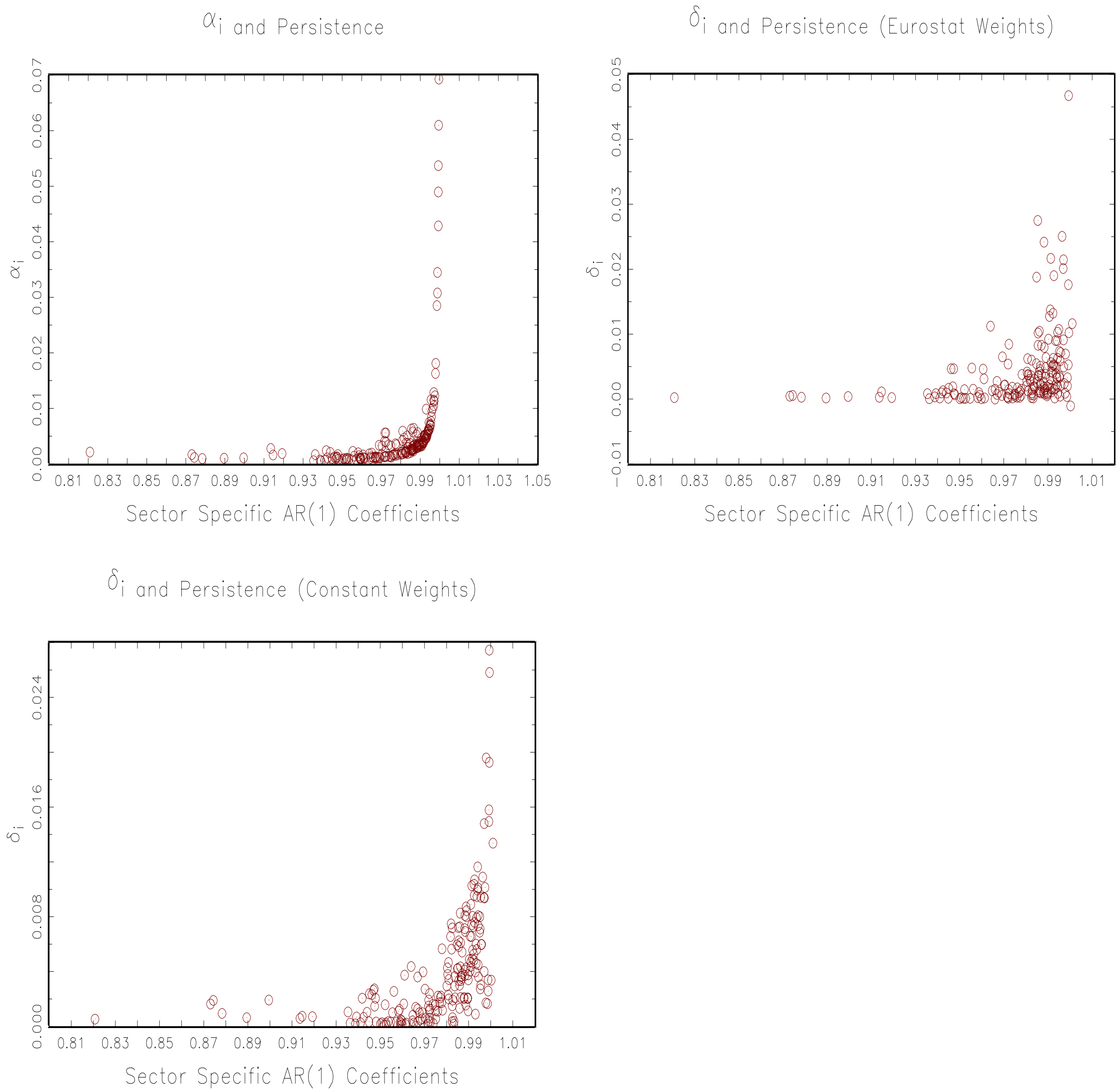
Table 4: Persistence Estimates using Disaggregated Data (Bias Corrected)

$q_{ict} = \gamma_c + \sum_{k=1}^{K} \rho_{ik}\ q_{ict-k} + e_{ict}$

Model	P	Half-Life (Indirect)	Half Life (Direct)
Mean Group	19	41	31
		$(17, 64)$	$(17, 57)$
Mean Group (SURE)	5	43	27
		$(18, 105)$	$(16, 65)$
Mean Group (CCE)	5	20	18
		$(11, 28)$	$(11, 28)$

Notes: The Bias Correction is carried out via the Kilian (1998) bootstrap method. "Indirect" refers to a method where ρ is corrected and the half life is estimated on the basis of ρ^*. In the "direct" case, the half-life is corrected directly. In each case, the bootstrap uses 500 replications. For the Mean group model N=204, and T=1981:01 to 1995:12. For the other two models the cross section in Chen and Engel (2004) is used to ensure non-singularity of covariance matrices. In addition, the time series is restricted to 1981:06 to 1994:09 in order to produce a balanced panel. This helps to decrease computation time and has little impact on the underlying (uncorrected) estimates. The confidence intervals are calculated via a double bootstrap procedure. That is, at each replication bootstrap samples are drawn using the mean estimates from the models in the table and the generated data is used to estimate the models via Kilian's bootstrap using 100 replications (50 for Mean Group (Sure)). This is repeated 100 times (50 for Mean Group (Sure)) and the 5% confidence intervals are calculated.

Figure 1



α_i and Persistence
α_i
Sector Specific AR(1) Coefficients
δ_i and Persistence (Eurostat Weights)
δ_i
Sector Specific AR(1) Coefficients
δ_i and Persistence (Constant Weights)
δ_i
Sector Specific AR(1) Coefficients

Figure 2: Sector Specific LAR's

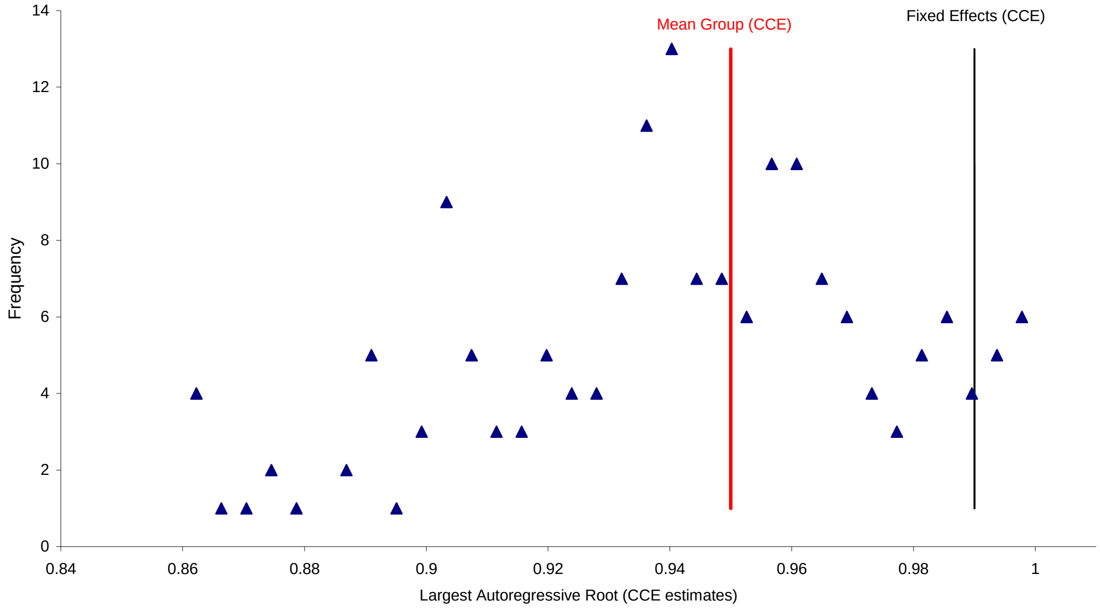
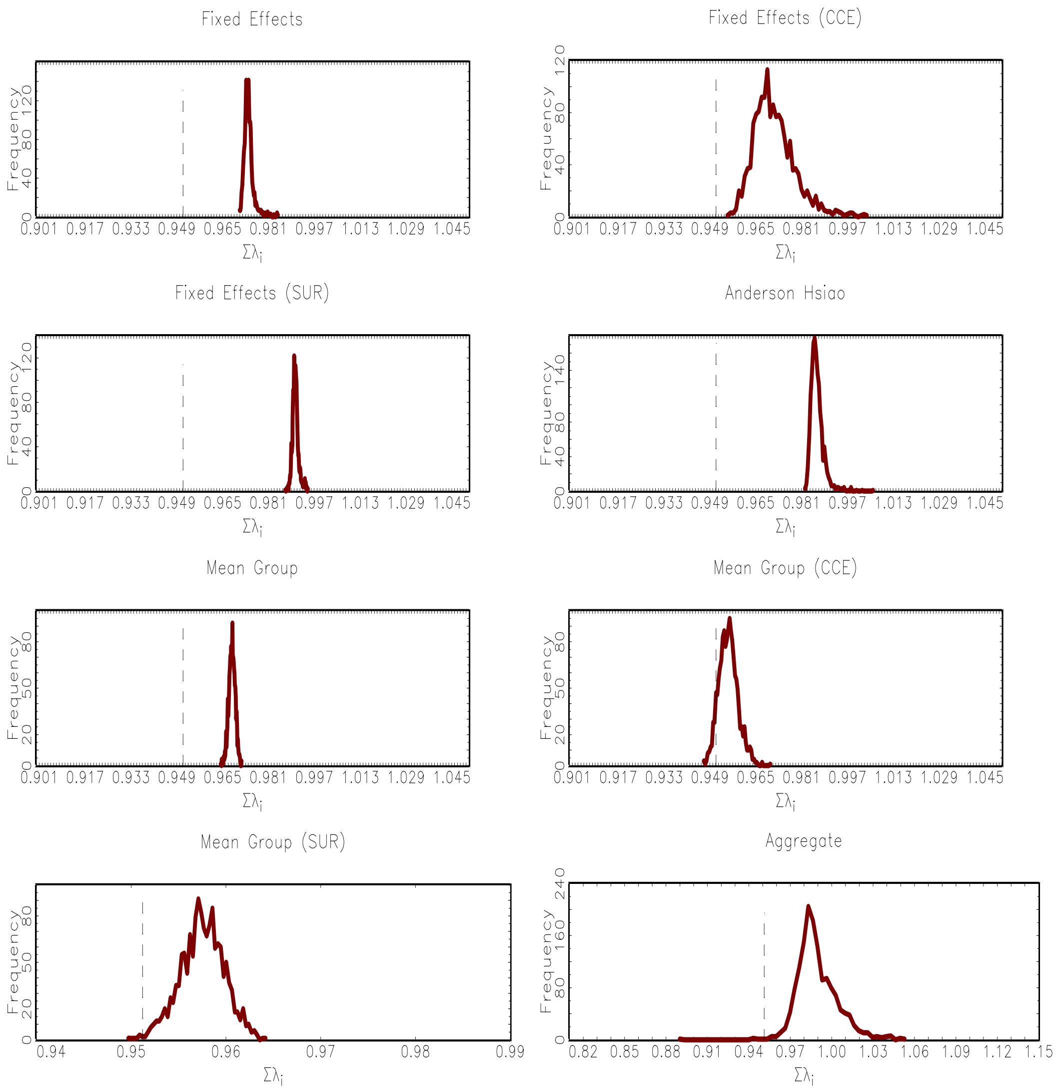


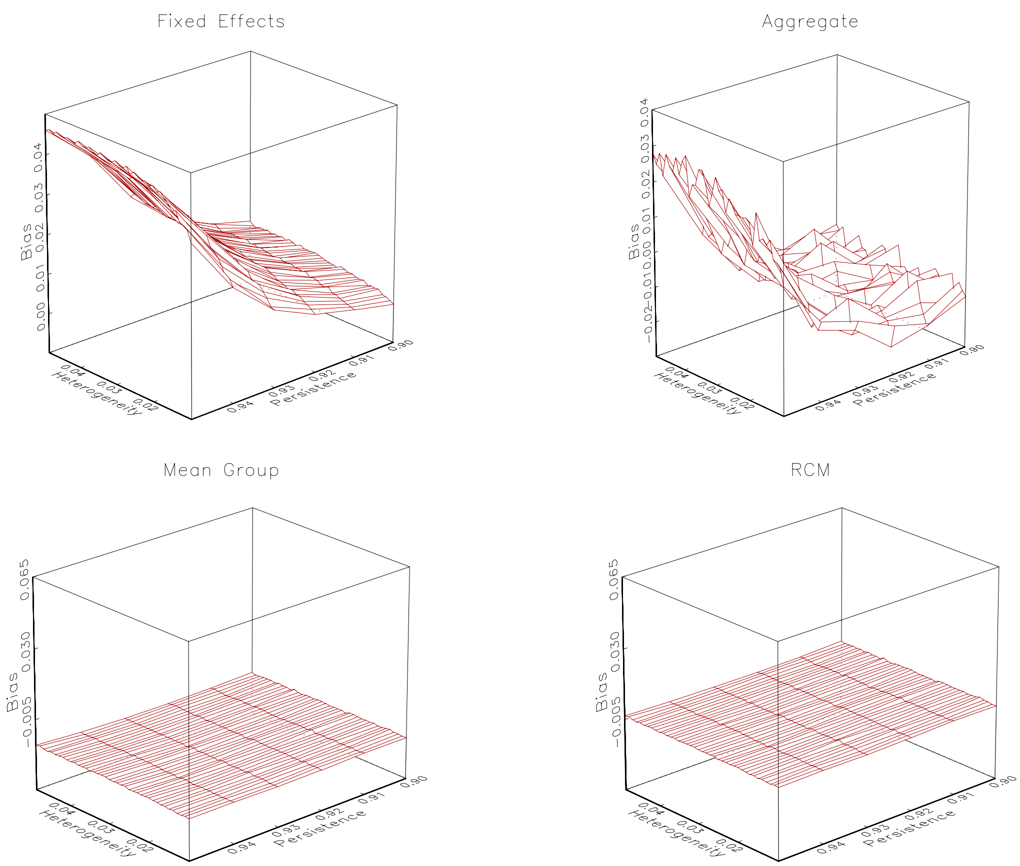
Figure 3



Notes:

The data was generated using the coefficient estimates and residuals from the Mean group model that allows for the CCE correction. At each replication, iid samples were drawn from each of the N=204 residuals and these (along with the mean $\frac{1}{N}\sum_{i=1}^{N} q_{i,t}$) were used to generate N AR(12) series, using actual observations as starting values and with the correlation structure implied by our coefficient estimates. The generated data was then used to estimate the various models shown above. Note that "aggregate" denotes estimation on time series data obtained by averaging over the cross sections. For the SUR models, the first 60 cross sections were dropped in order to make estimation feasible. The figure plots the histograms of the sum of coefficients obtained from 1500 replications and compares these with the true estimate shown as the dotted black line.

Figure 4



Notes:

The data was generated form the following AR(1) panel data model (with N=204, T=200): $y_{i,t} = \alpha_i + \lambda_i y_{i,t-1} + v_{i,t}$, where $\alpha_i, v_{i,t} \sim N(0,1)$. The heterogeneous AR coefficients are drawn from the following scheme: $\lambda_i = \lambda + \eta_i$, where $\lambda = \{0.9, 0.91, 0.92, 0.93, 0.94, 0.95\}$ and η_i is sampled from a truncated $N(0, \sigma^2)$ with bounds ± 0.05. We consider 40 values for σ^2 varying from 0.02 to 0.4 with increments of 0.001. This controls the underlying heterogeneity in λ_i and is shown on the "heterogeneity" axis in the figures. We use our data to generate y_0 and then discard the first 50 observations in each cross section to reduce influence of starting values.

The figures plot the mean bias in each estimator derived from 200 replications for each combination of λ and σ^2.