"Aggregation Bias" DOES Explain the PPP Puzzle[*]

JEAN IMBS[a,b,f], HAROON MUMTAZ[c], MORTEN O. RAVN[d,f], HELENE REY[e,f,g]

[a]HEC Lausanne, [b]FAME, [c]Bank of England,
[d]European University Institute, [e]Princeton University, [f]CEPR, [g]NBER

August 26, 2005

Abstract

This article summarizes our views on the role of an "aggregation bias" in explaining the PPP Puzzle, in response to the several papers recently written in reaction to our initial contribution. We discuss in particular the criticisms of Imbs, Mumtaz, Ravn and Rey (2002) presented in Chen and Engel (2005). We show that their contentions are based on: (i) analytical counter-examples which are not empirically relevant; (ii) simulation results minimizing the extent of "aggregation bias"; (iii) unfounded claims on the impact of measurement errors on our results; and (iv) problematic implementation of small-sample bias corrections. We conclude, as in our original paper, that "aggregation bias" goes a long way towards explaining the PPP puzzle.

1   Introduction

In "PPP Strikes Back: Aggregation and the Real Exchange Rate", we show that a dynamic aggregation bias is an important component of the PPP puzzle. Aggregate real exchange rates are persistent because their components have heterogeneous dynamics, for which established time series and panel methods fail to control[1]. When we use estimators allowing properly for heterogeneity in the data that we study, the persistence of the real exchange rate falls dramatically. Its half-life, for instance, falls to around one year. We show that the corrected estimates are consistent with existing dynamic general stochastic equilibrium models for plausible degrees of nominal rigidity. These models do not incorporate any heterogeneous dynamics and should hence be compared with our corrected estimates, which are purged from heterogeneity. Thus, arguably, we solve the PPP puzzle.[2]

Chen and Engel (2005) [henceforth CE] criticize our paper on four grounds: (i) they question the applicability of the "aggregation bias" to the PPP puzzle; (ii) they claim the size of the bias is "shown to be much smaller than the simulations in Imbs, Mumtaz, Ravn and Rey (2002) suggest"; (iii) they contend that measurement error contaminates our results and (iv) that small sample bias is the main reason behind them. In addition, they find that country-by-country estimates contradict our results. Here we respond to each of these and cover in detail what is often left as footnotes in our published paper, Imbs, Mumtaz, Ravn and Rey (2005) [henceforth IMRR (2005)].[3]

The next four sections all point towards the same direction: our results survive each of the criticisms raised by CE. In section 2, we discuss the general analytical proof presented in IMRR (2005) and explain why heterogeneous dynamics translate into an "aggregation bias" in our price data, thus answering CE's concern regarding the applicability of the bias to the PPP puzzle. The proof allows for sectoral correlations and varying expenditure weights. We show in particular that the counter-examples presented in CE are not empirically relevant. Even though it is theoretically possible that the bias be non-positive, it is certainly not the case in our data, nor in theirs. Section 3 makes clear the reason why the simulations in CE do not yield a large aggregation bias is their choice of parameters. The extent of heterogeneity they use in their simulations is smaller than in the data, and they choose a set of initial conditions that acts to minimize the bias.

---

[1] For a detailed exposition of the concept of dynamic aggregation bias see Pesaran and Smith (1995).

[2] We point out in the conclusion of IMRR (2005) that the question of whether dynamic general equilibrium models featuring non-trivial sectoral heterogeneity are capable of mimicking aggregate data is an exciting area for future research. Should such models prove unsuccessful at generating persistent aggregate real exchange rates, there would be a "new" PPP puzzle.

[3] Reidel and Szilagyi (2005) focus mainly on the interaction between small sample and aggregation biases. We also address their claims in this article. Most of them are largely similar to CE's.

In section 4, we deal with claims in CE that we use a dataset plagued with measurement errors, supposedly at the source of our results. It is doubtful right from the outset that short-lived measurement error could explain our conclusions. CE's argument is that measurement error acts to lower persistence estimates based on sectoral data, while estimates based on aggregate data are immune to the problem as it tends to be averaged away. But for this to be convincing, one would need to observe persistence estimates at the disaggregated level systematically lower than in the aggregate. We do not: there are more than a few sector-level persistence estimates in excess of the aggregate. Nevertheless, we perform extensive robustness checks to show our results are not sensitive to this argument. First, we use exactly the same dataset that Charles Engel advocates on his website. When we apply the appropriate estimator to his data, we find a half-life of 13 months (with a confidence interval ranging from 9 to 24 months), hardly different from our results of 14 months in IMRR (2002). CE find a different result simply because they use an estimator which is rejected by their data: they implemented on their sample the same estimator we used in our original dataset. Their data, however, call for another estimator. Second, we improved on the IMRR (2002) dataset by systematically checking Eurostat time series against national sources. We believe that the final dataset used in IMRR (2005) is of better quality than Engel's. It yields a half-life around 11 months. Finally we perform formal tests for errors-in-variables and remove any suspicious series. Again, our results stand.[4]

In section 5, we discuss the importance of the claim that our results are not robust when our panel estimates are decomposed country-by-country. Criticizing our results on the ground that they do not hold within countries is akin to criticizing results based on panel unit-root tests on the ground that unit roots cannot be rejected on a country-by-country basis. The section next details the small-sample bias corrections we implement, and shows that they still yield half-life estimates well below the "consensus view". CE find otherwise because they implement inappropriate estimators which, together with their bias correction technique, induce a *positive* bias in their corrected half-life estimates. Section 6 concludes.

2 Applicability of "Aggregation Bias" to the PPP puzzle

2.1 Some Theory

---

[4] The first version of our paper, IMRR (2002) already had a full section dedicated to the treatment of measurement errors. Our cleansed data has been available on the internet since September 2003.

CE claim the "aggregation bias" is not really applicable to the PPP puzzle since the bias is not necessarily positive. They argue cross-sectional correlation of the errors may give rise to a negative bias. Further, the sign and magnitude of the bias may depend on expenditure weights. To bolster their claim, they present counter-examples for which the bias is either zero or negative.

We focus here on the case where the panel consists of the relative prices of goods for a single country pair. IMRR (2005) generalizes to panels of exchange rates. Consider an economy with N sectors, indexed by i. For simplicity, suppose that (the log of) relative prices in each sector follows an autoregressive process of order one, defined by

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

with $c_i = c + \eta_i^c$ and $\rho_i = \rho + \eta_i^\rho$. We 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 $]-1, 1[$.[5] We also suppose that ε_{it} is independently distributed with mean 0 and variance σ_i^2 with $E(\varepsilon_{it}^2) = \sigma_i^2$.[6] We 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. Without loss of generality, we order the N sectors so that $0 < \rho_1 \leq \rho_2 \leq ... \leq \rho_N < 1$.

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 jth good.[7]

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

PPP studies estimate the persistence of the real exchange assuming that its dynamics are best described by an AR(p) process. Many use an AR(1) as their standard specification. So will we to simplify the derivations. An aggregate estimation ignoring the heterogeneity in the dynamics of the subcomponents would write

$$q_t = c + \rho q_{t-1} + \varepsilon_t$$

with

$$c = \sum_{j=1}^{N} \omega_j c_j, \; \varepsilon_t = \sum_{j=1}^{N} \omega_j \varepsilon_{jt} + \sum_{j=1}^{N} \omega_j \eta_j^\rho q_{jt-1}$$

It is immediately apparent that since lagged dependent variables are present in the error term, aggregation across sectors also leads to an inconsistent estimate of the mean persistence. In IMRR (2005), building on Pesaran and Smith

---

[5] We consider drawing from a discrete set of H values in the interval $]-1; 1[$.

[6] See IMRR(2005) for other technical assumptions.

[7] See IMRR (2005).

(1995), we show under what conditions this inconsistency gives rise to a positive "aggregation bias", that is increasing in the degree of heterogeneity. In particular, we show that the bias Δ writes

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

with $\delta_i = \dfrac{\left[\frac{\omega_i^2 \sigma_i^2}{1-\rho_i^2} + \sum_{i \neq j}^{N} \frac{\omega_i \omega_j \sigma_{ij}}{1-\rho_i \rho_j} \right]}{\sum_{i=1}^{N} \left(\frac{\omega_i^2}{1-\rho_i^2} \sigma_i^2 + \sum_{i \neq j}^{N} \frac{\omega_i \omega_j \sigma_{ij}}{1-\rho_i \rho_j} \right)}$. Hence the bias depends on cross sectoral correlations, persistence of the various sectors and expenditure weights. We prove that the aggregation bias is positive ($\Delta > 0$) whenever the coefficients δ_i are positively correlated with the persistence parameters ρ_i. This turns out to be unambiguously the case in our data, as well as in Engel's. Furthermore, in IMRR (2005), we show that the bias is not only positive but also quantitatively important.

Could the bias be negative or negligible *in theory*? It certainly could, as is obvious from the inspection of the above expression. And it is easy to come up with simple analytical examples in which the bias is either zero or negative. On page 53, CE develop an example where two price series are perfectly negatively correlated (and thus exactly cancel out). In that case, if $N = 3$, the aggregate persistence is that of the third, uncorrelated series. But none of these rather extreme assumptions hold in price data. Next CE choose to linearize the expression of the bias around a perfectly homogeneous case to argue the bias is small or inexistent whenever $\omega_i = \omega$ and $\sigma_{ij} = c$, or $\sigma_i = \sigma$ and $\sigma_{ij} = c$. We note the Taylor expansion is computed around the homogeneous case. This is important, as we showed in IMRR (2002) that the magnitude of the bias increases with heterogeneity, and indeed is zero under homogeneity. Since CE focus on almost homogeneous processes, and use an approximation imposing linear effects of parameter heterogeneity, it is to be expected the heterogeneity bias will be small. Irrespective of the expansion point chosen to perform the approximation, however, what matters is whether the restrictions imposed in all these experiments are plausible empirically or not.

CE made some other related points:
1) On page 52, CE claim "we can unambiguously state there is aggregation bias only when α_i (the weights in the price index) are equal for all i, σ_i^2 (the innovation variance for x_{it} as defined in equation(3)) is the same for all i; and, the cross-correlations of all series are equal". Inspecting expression (1) shows immediately that this is not correct. There are many different cases, with different weights, different variances and covariances, which give a positive bias Δ. Even the condition that we show in IMRR (2005) to be sufficient for the positivity of the bias (*i.e.*. $0 \leq \delta_i \leq \delta_{i+1}$ for all i) can be fulfilled under many possible data configurations, again with different weights, variance and covariances. And, in

our data as well as in Engel's, the bias is unambiguously positive, even though weights and innovation variances are *not* the same.

2) CE mention evidence in Rogers and Jenkins (1995) that unit roots can be rejected only for a few perishable items, which tend to have a low weight in the CPI. They infer that the aggregation bias should be small. Things actually go the exact opposite way. In reality, inspection of the expression for Δ shows that for the bias to be negative, one would need *highly* persistent relative prices to have (ceteris paribus) *low* CPI weights ω_i.[8] The fact that low persistence items tend to have low CPI weights would if anything reinforce the positivity of Δ, since it suggests ω_i (and therefore δ_i, ceteris paribus) tend to be low for low ρ_i. Of course, the only sure answer to this question is to compute δ_i and compare it with estimates of ρ_i. In our data, the verdict is unambiguous, and the bias is positive.[9]

3) CE criticize us on the ground that it is "well-known" that summing AR processes yields an ARMA process. Well-known or not, this *only* happens under heterogeneity. In the absence of any heterogeneity, the roots cancel out and the aggregate real exchange rate is an AR process, akin to the one driving sectoral prices. In other words, estimates of the persistence in real exchange rate that only include autoregressive terms (no matter how many) ignore heterogeneity. And since nearly all the papers of the PPP literature estimate AR(p) -and often AR(1)- they *de facto* ignore heterogeneity. Furthermore, pursuing the route of taking heterogeneity into account by allowing for ARMA terms in the real exchange rate may not even be feasible in practice, since with sufficient heterogeneity, one would quickly run out of degrees of freedom unless the sample period is long enough. Heterogeneous estimators are better-suited to tackling the issue than estimating processes with infinite (or even high order) ARMA terms.

We have now demonstrated CE's claim that "aggregation bias" is not applicable to the PPP puzzle because the bias could be negative or zero in theory, is irrelevant empirically. But it should also be clear from the expression of δ_i that covariances in prices across sectors will affect the magnitude of the bias.[10] In IMRR (2005) we introduce heterogeneous estimators that do account for correlated residuals. By contrast, CE only point to the possible importance of non-zero σ_{ij}, and give empirically implausible analytical counter-examples. The estimators they use, however, do not investigate which way non-zero σ_{ij} affects the aggregation bias, *in the data*. Our estimators do.

Accounting for cross-sectional dependence in as large a panel as ours, while continuing to allow for heterogeneity in the slope coefficients is by no means

---

[8] As well as low innovation variance σ_i^2, and/or low covariances with other components of the real exchange rate. We note that in our price data covariances are systematically positive.

[9] See IMRR (2005).

[10] These correlations are uniformly non-negative in our data.

straightforward. We implement two estimators that can handle both issues, the standard Seemingly Unrelated Regression (SURE-GLS) estimator and the common correlated effects (CCE) estimator introduced by Pesaran (2002).[11] We show correlated residuals to be an important characteristic of the data. We find using both estimators that half-lives are even lower once correlated residuals are accounted for. Thus, $\sigma_{ij} \neq 0$ is an important element of Δ, but not because it tends to decrease the magnitude of our bias, as CE claim. Indeed it is rather the opposite. We also note that when we use a fixed effect CCE estimator without allowing for heterogeneous dynamics our estimated half life is still high (see Table III of IMRR (2005)). Hence it is not the CCE estimator per se that drives down our estimates. It is indeed the heterogeneity in the dynamics of the sectors that is responsible for the high measured persistence at the aggregate level. In other words, there is evidence of aggregation bias whether a standard fixed effect or a fixed effect CCE estimator is used.

The main empirical result of Reidel and Szilagyi (2005) [henceforth RS] is the discrepancy between the sectoral half-lives obtained using OLS and CCE estimators. These authors suggest the difference is problematic. But if the panel is characterized by cross-sectional correlation and if this correlation is driven by a common factor, then the discrepancy between OLS and CCE estimates arises *precisely* because OLS regressions do not take the correlation structure into account. And in panels of relative prices, the presence of a common component in the residuals is very plausible given the high likelihood of common shocks and missing variables. We also performed simulations to evaluate the properties of the CCE estimator. Simulations suggest estimated half-lives are biased (upwards) if correlated residuals are not accounted for. The issue is important empirically and allowing for correlated residuals actually strengthens our conclusions (see Figure III of IMRR(2005)).

2.2 Some Intuition

It is important to understand aggregation is a problem in panels *because* of ignored heterogeneity across the components of the real exchange rate. In other words, estimates based on international panels of real exchange rates will generate high persistence because each real exchange rate is composed of many sectoral relative prices whose dynamic properties are heterogeneous. Separating "aggregation bias" and dynamic heterogeneity bias, as CE seem to suggest at the end of section 1, is impossible. They are one and the same issue. In particular, our main argument is that wrongly imposing an identical speed of adjustment for the *components* of the real exchange rate can lead to a bias in

---

[11]Implementing SURE 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, which we accomplish in IMRR (2005) by using Engel's version of the Eurostat dataset, which has fewer observations than ours.

persistence estimates, *both* in aggregate time series and in a panel setting. More precisely, if the speed of relative price adjustment differs across goods, the speed of adjustment of the real exchange rate may not be an unbiased estimate of the average speed of adjustment of relative prices. Thus, our main argument is *not* that "impos[ing] an identical speed of adjustment across all real exchange rates [...] can bias estimates of the half-life of real exchange rates" (page 51). While such a bias is possible, our concern is about imposing identical speeds of adjustment across *different types of goods* and how this may affect estimates of relative price persistence.

Our paper explains how heterogeneity in the dynamics at the good level translates into persistent aggregate real exchange rates. *Our result does not require nor imply that persistence be systematically smaller at the disaggregated level.* Hence results in Crucini and Shintani (2002) and Engel (2000) that CE invoke in their conclusion do not contradict ours in any way.[12] That Crucini and Shintani (2002) should find *homogeneously* fast mean reversion in a wide range of good prices immediately suggests they will not find much evidence in support of "aggregation bias" in their data. There is little heterogeneity in their estimates which all point to strikingly low half lives. If we are right, there should not be much of an aggregation bias in a dataset of goods which unanimously tend to revert to parity quickly. But surely, there *is* a PPP puzzle. Most aggregate relative prices do tend to revert to parity slowly. That they do not in Crucini and Shintani (2002) does not invalidate our contention.

As a matter of fact, the revised version of Crucini and Shintani (2004) now finds explicit support for an aggregation bias. Their introduction reads: "But how do we go from half-lives in the neighborhood of one year at the level of individual goods to half-lives of CPI-based PPP deviations in the range of 3 to 5 years? The answer we present in this paper is a combination of small sample bias and aggregation bias". They therefore now confirm the existence of an aggregation bias in a data set that includes emerging markets and is very different from ours.

It is also peculiar that CE assert that "there is already a large literature devoted to [...] biases in panel estimates" due to heterogeneity, with application to the real exchange rate (page 51). An EconLit search yielded one single paper dealing with dynamic heterogeneity in the real exchange rate, and the point there was heterogeneity across countries, not across sectors.[13] In fact, in their footnote 2, CE describe the Random Coefficient Model (RCM) erroneously.

---

[12] In their conclusion CE write: "Crucini and Shintani (2002), using a large worldwide cross-section of goods prices from the Economics Intelligence Unit, do find rapid convergence to deviations from the law of one price (half-lives of 9 to 12 months). But they explicitly find no aggregation bias. The rate of convergence of the average of their prices is very similar to the average of the rates of convergence of their individual prices. Aggregation bias does not seem to explain the PPP puzzle."

[13] See Boyd and Smith (1999), who conclude there is very little heterogeneity in real exchange rates dynamics between countries.

RCM does not compute an arithmetic mean of sectoral persistence estimates to obtain aggregate half-lives. There are weights, and they are optimally inferred from the observed heterogeneity in the data, using a Generalized Least Squares procedure akin to that implemented in Random Effects estimators.

3 Simulations

The second main criticism of CE - taken up as well by RS - concerns the Monte-Carlo simulations in IMRR (2002). The reason why we performed Monte-Carlo simulations in the original version of our paper was not to ensure there was a bias. Comparing estimates where heterogeneity is allowed for, to ones where it is not is indeed sufficient to prove the presence of a bias. Our Monte-Carlo simulations were meant to quantify how the bias responds to variations in the extent of heterogeneity and/or of persistence (and how various estimators perform at capturing it). CE propose to use Monte-Carlo simulations to prove there is no bias, but both our data and theirs (even cleansed of measurement error and small-sample bias) show it is there. The proof is in the pudding.

CE question the validity of the bias derived in our original simulations (IMRR, 2002) on grounds that we allow the possibility for explosive roots in our simulated sectoral prices. However, IMRR (2003) showed the simulated bias remained substantial - identical, almost - even after changing the distribution of sectoral persistence coefficients so that they do not include any value (weakly) above unity[14]. Indeed, the Monte-Carlo simulations in IMRR (2005) *exclude* de facto any explosive roots. Thus, the discrepancy has to come from somewhere else.

There are first some obvious reasons. In IMRR (2002), we show the bias increases with the extent of heterogeneity, so part of the reason for the differences in simulation results stems from the fact that CE use a range for the heterogeneity in sectoral persistence parameters that is smaller than ours (and indeed smaller than what our -cleansed- data imply). Equally problematic are the assumptions RS use in their simulations. They consider only AR(1) processes and assume that the mean persistence parameter is equal to 0.97. In contrast IMRR (2005) use higher order autoregressive processes and estimate the largest autoregressive root to be 0.95. Both aspects are important. First, there is good reason to expect the presence of higher order terms, especially in monthly data. Low order autoregressive processes may poorly fit the dynamics of the data. Second, assuming a larger value for the autoregressive parameter implies a higher relative importance of the small sample bias. Furthermore, allowing for cross-sectoral correlations present in the data actually increases the magnitude of the bias - CE never allow for cross-sectoral correlations in their simulations.

---

[14]CE (2005) does not mention the simulations of IMRR (2003).

There is another, more subtle yet important reason why the Monte Carlo results of CE and RS differ from ours. In their simulations, CE allow for fixed effects. The initial conditions are such that each cross-sectional unit starts at its asymptotic mean. This is at best a special case, which has important consequences for the results. Similarly RS draw initial conditions from their unconditional distribution. Instead, Arellano and Bond (1991) impose zeros for all initial conditions. In IMRR (2005), we use the actually observed initial conditions. It turns out that the heterogeneity bias is important under any of these two standard alternatives.

The intuition is as follows. As soon as initial conditions differ from the asymptotic mean of the cross-sectional units, there is an initial period of convergence towards the asymptotic mean even in the absence of any shocks. The MG and the RCM estimators allow for heterogeneous dynamics over this adjustment period, but FE and indeed any other aggregate estimators do not. Monte-Carlo simulations that ignore this initial adjustment period will tend to minimize the discrepancy between heterogeneous estimators and standard ones. Indeed, simulations that assume initial conditions that are specifically equal to their long run values or drawn from their unconditional distributions are the only ones that will tend to minimize the superiority of heterogeneous estimators in the presence of heterogeneity. It seems inappropriate to reject the possibility that there could be an aggregation bias, on grounds of simulation setups that minimize its impact.

In IMRR (2005) we confirm that the bias is large, using actually observed initial conditions.[15] We note in particular that RS's simulation result that the CCE estimator is severely biased downwards depends entirely on the authors assumptions about the starting values for the AR processes used in their experiment. Specifically, if the simulations in the paper are started using the actual Eurostat data, the CCE estimators perform very well (see Figure III of IMRR(2005)). We stress that simulations are *not* meant to establish the existence of the bias in the data, since the formal tests and the results already show this. Rather they are meant to investigate robustness across heterogeneity and persistence parameters. We provide analytical and direct empirical support for the presence of a bias in our data. Given the importance of initial conditions, Monte-Carlo simulations should be used neither to establish nor to disprove the existence of a bias. We do not propose to do the former, and one should not attempt the latter.

4   Data

---

[15]In our simulations, we actually truncate the first 50 observations of our Monte-Carlo simulations, in order that the importance of initial conditions be minimized. Even so, we continue to find a large bias.

The third claim in CE concerns the impact of measurement error on our estimates. First, if short-lived measurement error were to explain our low estimates for persistence, it should do so on the basis of systematically lower persistence at the sectoral level, which would naturally aggregate into low half-life for the real exchange rate. As we show in IMRR (2005), there are no few instances of sectoral persistence exceeding the aggregate measure, which would be impossible if measurement error were generating our results. Even so, CE argue otherwise. In this Section, we detail the reasons for this discrepancy.

On his website, Charles Engel lists revisions to the Eurostat dataset that he deems appropriate. We present below estimates based on these exact corrected data, which were performed as early as January 2003. They confirm the half-life drops dramatically when heterogeneous dynamics are taken into account. We also explain in detail why CE find otherwise: they implement an estimator which is rejected by their data. In IMRR (2002), we used the official Eurostat data, which we had already corrected for some obvious typos (and other mistakes) to obtain our initial estimates. Our section 6.1 was entirely devoted to tests of errors-in-variables, and re-ran all our estimations with suspicious observations replaced by (sufficiently) lagged values, as is customary. Our results were confirmed. Finally, in IMRR (2005) we go one step further, and use sources from national statistical agencies to verify the consistency of the Eurostat data, as well as the corrections suggested on Charles Engel's website. Our revised dataset is therefore arguably of better quality than Engel's.[16] Again, our results stand.[17]

4.1 Data Corrections, Part 1: On the Importance of Implementing the Appropriate Estimator.

Engel's data have 127 cross sections for the period 1981:01-1996:12. There are 9 countries and a maximum of 16 goods. The coverage is considerably lower than the data used in IMRR (2002). In particular, Greece, Finland and Ireland are excluded as are goods such as Rents and Tobacco. In fact, the number of cross sections in their data set is about half of those used in the original version of our paper.[18] There are two standard estimators which control for heterogeneous dynamics: the Random Coefficient Model (RCM) and the Mean Group (MG). One should perform an appropriate test to ascertain which one should be used in a given data sample. With a smaller cross-sectional dimension and a shorter sample, it is to be expected the RCM estimator will perform less well, but the

---

[16] And is available on our websites.

[17] Prior to CE, Di Giovanni (2003) also contended that data limitations were the source of our result, and found dissenting evidence on data with longer time series. What follows also addresses his concerns. In particular, as the dataset alters, so may -potentially- the heterogeneous estimator that should be implemented.

[18] Our dataset had 221 cross sections.

MG estimator still has good small sample properties. We now discuss this in detail.

Table 1 confirms that we closely reproduce CE's RCM and FE estimates when using their data. However, the MG estimator produces a much shorter half-life (25 months), with a precise confidence interval (9 to 31 months).[19] In their Table 5, CE find an upper bound of 142 months when using the MG estimator on their data. Since we find low upper bounds when we use their exact data, the discrepancy must stem from their bootstrapping technique, which they do not explain. We follow Ron Smith's suggestion to bootstrap, and use the mean coefficients to draw the residuals, before performing sampling from the residuals themselves.

Figure 1 shows that our findings remain true for all possible lag lengths. The implication is clear. *Allowing for heterogeneity makes a significant difference in the CE sample as well. The Hausman test strongly rejects parameter homogeneity.* CE do not find a strong effect of heterogeneity on the basis of the RCM estimator, but in this much smaller sample the RCM estimator is rejected in favor of the MG estimator. Why do RCM and MG not perform equally well? Both estimators allow for slope heterogeneity, but only the former imposes distributional assumptions on heterogeneity. Using CE's data a Hausman test for heterogeneity strongly rejects homogeneous slopes. However, the alternative hypothesis could be either heterogeneous and deterministic, or heterogeneous and random. In other words, the alternative hypothesis is consistent with both MG and RCM. In order to distinguish between the two we use a test devised by Pudney (1978). This is a test for the assumptions underpinning random coefficients. A rejection of the null hypothesis implies rejection of the random coefficient assumption. In an AR(5) model we obtain a test statistic of 126.72 (0.00). This implies that the Mean Group model is more appropriate in these data.[20] When implemented on exactly Engel's data, the MG estimator yields a half life of 25 months, as shown in Table 4. But none of these estimates allow for cross-sectoral correlations. And indeed, the half-life drops further to 21 and 13 months when we allow for correlations across sectoral prices via a MG SURE or a MG Common Correlated Effect estimator (henceforth MG CCE), respectively.[21] We show in IMRR (2005) that the common effects are an important characteristic of price data.

The fact that CE's panel is narrower than ours could also explain their problems with the RCM model, especially at high lags. This is particularly important as the heterogeneous estimators we propose are essentially averages,

---

[19] All half-lives in our paper are defined as the number of periods it takes for the impulse response to cross 0.5 *permanently*, as is customary.

[20] In our original data the statistic was 10.29 (0.90).

[21] See Pesaran (2002) for details on the MG-CCE estimator. Table 4 also reports all the alternative measures of persistence we use to bolster our argument in IMRR (2005). They all lead to the same conclusion.

and their consistency and efficiency depends on the cross-section of the panel. We conducted a simple experiment. We assumed a heterogeneous data generating process, with 220 cross sections. Then we estimated RCM models only on the first 100 cross sections. Figure 2 plots the distribution of the resulting estimates and contrasts it with estimates from the whole panel. It is clear that the estimator using fewer cross sections has a much more dispersed distribution, i.e. the estimates are less precise. This problem is likely to be more severe as the number of parameters increases.

Direct evidence on the importance of this point can be seen in Table 2, where we list estimates obtained when CE's dataset is expanded. We add the following: 1) Data for Greece, Finland and Ireland. 2) Data for Tobacco and Rents. In each case, the data for all countries were checked and any outliers were removed, in a way similar to the selection method described on Charles Engel's website. This gives us a panel with 191 cross sections, a number closer to our original data. The fixed effects half-life is close to CE's estimate. The heterogeneous estimators, however, now produce shorter half-lives. In particular the MG model gives a half-life of 20 months. Figure 3 plots the half-lives obtained from these estimators against the lag-length. The MG half-lives are consistently less than two years. The RCM model produces half-lives close to two years, whereas the Fixed Effects estimator is biased upwards. Note that RCM and MG converge at higher lag lengths, as they should. Note also that we do *not* observe the kind of impulse responses found by CE.[22] In IMRR (2005) we provide confidence intervals for these results based on corrected data. Thus, CE find different results -even though we use almost identical data- because they implement the very same estimator we used on our original sample. But this estimator is rejected in their data. Had they used the appropriate estimator, they would have obtained our results.

4.2 Data Corrections, Part 2: Formal Tests for Errors-in-Variables, and Even Better Data.

CE correct the data by removing outliers and parts of the series that appear inconsistent. We do a similar exercise when we expand their data. However, removing "suspicious" data may also be problematic since it introduces a degree of subjectivity. In other words, there is a chance that it remove shocks that are actually informative. In fact, it is possible that such a procedure may produce persistence. We can infer the impact of this from the following experiment: 10,000 AR(1) processes with an autoregressive coefficient of 0.95 were generated. Estimation was carried out on (i) the generated series without any changes (ii)

---

[22]We tried various other combinations of the data. For example, adding data for Greece and Ireland only, produces very similar results. In addition we considered removing every series that has repeated observations (as noted by Charles Engel on his web page). Again the MG estimator gives a half-life of 23 for an AR(5) model.

on series where "outliers" where replaced by an average over $t+1$ and $t-1$. The mean estimated half-life in case (i) is 13.52 months, which is very close to the true half-life of 13.51. In case (ii) this goes up to 22.5 months. Figure 4 plots the distribution of the estimates. It is easy to see that in case (ii) the distribution is much more dispersed around a higher mean.

This is perhaps not very surprising because in this case outliers are erroneously removed. In reality, many of the "corrected" data may of course be true measurement errors. Correcting for measurement errors on the basis of removing replacing outliers is problematic, however, since it would for instance not remove "small" measurement errors. For that reason, a more objective approach to the measurement error problem might be desirable.

In the original version of our paper, we reported RCM estimates based on GMM estimators with instruments chosen to account for errors in variables. We showed this did not affect the results. We now examine how this estimator performs. We generate data for AR(1) models using a coefficient of 0.95. Then we add a random error ν to these data where $\nu \tilde{} N(0, 0.3)$. A typical sample is shown in Figure 5. It can be seen that the series inclusive of the error has many possible outliers. Next we estimate models using OLS, which is expected to be biased, and GMM, which is consistent. The distribution of the resulting estimates is shown in Figure 6. There is a downward bias in OLS, but GMM performs much better and its mean is close to the true estimate. This does indicate that if errors in variables were a substantial problem we should have observed a large difference between RCM estimates based on OLS and GMM. In IMRR (2002), we found very similar results from using either estimator, indicating that measurement error did not account for our results.

The data we use in IMRR (2005), are the result of thorough checks of Eurostat sources against series published by national statistical agencies.[23] Our results, all presented in our main paper, are almost unchanged. Our best estimate for the half life is 11 months with a confidence interval ranging from 7 to 12 months. Hence, once more, our results are confirmed.

5   Country-by-Country Evidence and Small-Sample Bias

The last section of CE develops two points. First, it is argued our results do not hold on a country-by-country basis, and there is little evidence of cross-sectoral heterogeneity within countries. Second, it is claimed that our persistence estimates suffer from a small sample bias, affecting least squares estimators when the data are persistent.

---

[23]Our data are available from our websites.

That our evidence should weaken on a country-by-country basis is unsurprising. We have made up for the lack of detailed disaggregated data on relative prices by using the country dimension in our panel. Criticizing our results on grounds that they do not hold within countries is akin to criticizing results based on panel unit-root tests, on ground that unit roots cannot be rejected on a country-by-country basis. For instance in the related literature on real exchange rates persistence, should one dismiss the results in Frankel and Rose (1996) or Murray and Papell (2003), which rest on the improved performance of unit-root tests when a panel dimension is brought to the task? Should one counter their argument with claims that it does not hold on a country-by-country basis?

We think one should not. That is not to say we have not tried to increase the sectoral (or temporal) dimension of our panel as well, but for all its faults Eurostat provides to our knowledge the best coverage of disaggregated relative prices there is. As we underline in our paper, the number of (monthly) observations in our data is large relative to the literature, and enough to alleviate somewhat the weakness of (panel) unit root tests. If the purpose were to address the question of real exchange rate persistence within countries, one would need much more disaggregated sectoral data than those we have.

Second, CE claim that a small sample bias pervades our estimates.[24] At first, CE's results may appear consistent with the evidence presented in an interesting related paper by Choi, Mark and Sul (2003) [CMS henceforth]. CMS examine the relative importance of three influences on real exchange rate persistence estimates: temporal aggregation, small samples, and heterogeneity. Their empirical analysis centers on the relative magnitude of the first two biases, because they fail to reject slope homogeneity in their data (they use a panel of aggregate real exchange rates, not sectoral data). Indeed we report similar results when we test for slope heterogeneity in a panel of aggregate real exchange rates. We found heterogeneity to be key at the sectoral level not at the country level.

The bias correction procedure used by CMS - suggested by So and Shin (1999) and extended by Sul, Phillips and Choi (2002) to account for common effects in residuals - is only meant to assess the relative importance of the small sample and the temporal aggregation biases. In their panel of aggregate real exchange rates, CMS find the small sample bias dominates. Their quantification of the heterogeneity bias, on the other hand, is based on simple Monte Carlo experiments, in which the data generating process has slope heterogeneity. In their simulations, they examine whether the total bias of a simple OLS estimator is positive (in which case the heterogeneity bias dominates) or negative (in which case the small sample bias dominates). For artificial data with fewer than 200 observations, they find that the total bias under OLS is negative, but somewhat sensitive to the calibration of heterogeneity. For 200 observations, the bias of the

---

[24]What follows also largely applies to RS's criticism.

OLS estimator can turn positive. Importantly however, their data generating process does not allow for the common components that we find are important. Their simulations, therefore cannot be used to dismiss the heterogeneity bias *in our data*, where common components are crucial. Above all, CMS never propose to implement their bias correction procedures to data with slope heterogeneity. They never claim, with reason, that the methods they apply have desirable properties when applied to data with characteristics akin to ours.

By contrast, CE correct for the small sample least squares bias by implementing the standard So and Shin (1999) procedure, which relies on recursive demeaning of the data, and a bootstrap-after-bootstrap procedure suggested by Kilian (1998). They find that either method gives rise to a substantial increase in the corrected half-life of the data. The Kilian (1998) procedure implies an increase in the mean bias corrected MG estimate to 44 months as against least squares estimates of 26 months. The 95 percent confidence interval goes from 13 months to infinity. Application of the So and Shin correction instead raises the point estimate to 161 months with a 95 percent confidence interval spanning 112 months to infinity. Thus, their results imply a worsening of the PPP puzzle, and bring our results into doubt.

There are several problems with CE's procedure. First, and this is a major point, relatively little is known about the properties of the bias correction methods that CE use, when applied to heterogeneous panels with common correlated effects. We note again that CMS carefully tests for homogeneity (and fail to reject) before applying these corrections on their data. Second, CE consider estimators that do not allow for common effects while we show common effects to be an important characteristic of our price data. Third, there are reasons to question CE's application of the Kilian (1998) procedure. In particular, whenever the statistic of interest is a non-linear function of the estimated autoregressive parameters, Pesaran and Zhao (1999) have shown that in heterogeneous panels, bias corrections should be performed *directly* on the statistic of interest, and *not indirectly* on estimated parameters. Using corrected autoregressive coefficients to calculate a half-life, as CE do, may result in asymptotically biased corrections.[25]

We therefore first provide some Monte Carlo evidence on the properties of the bias reduction methods.[26] Unlike CE, our correction procedures do account

---

[25]Pesaran and Zhao (1999) discuss this issue in an application to a panel setting where the object of interest is a long-run multiplier of a change in an exogenous variable. But the same problem arises when estimating the half-life. Indeed, the issue will arise whenever one attempts to correct short-term point estimates in order to obtain an unbiased assessment of long-run phenomena. In general, the expression for long-run properties is highly non-linear in short-run estimates.

[26]Phillips and Sul (2003) suggest a Panel Feasible Generalized Median Unbiased estimator that can be applied in heterogeneous dynamic panels with common effects. Their method generalizes the median unbiased correction to a SUR estimator. As noted in IMRR (2005) the cross-sectional dimension of our data is large and the MG CCE estimator may outperform the SUR estimator. For this reason we consider alternative bias correction methods.

for correlated residuals. Following Pesaran and Zhao (1999), we perform our correction directly on the half-life, which is the variable of interest.[27] In Appendix B, we describe the steps in our *direct* bootstrapping approach. It stands in stark contrast with the indirect bias correction implemented in CE. We report the outcome of a simple Monte Carlo experiment meant to compare the direct and indirect approaches in the presence of correlated residuals. We assume a data generating process in which the panel units are generated by AR(1) processes with slope heterogeneity and common correlated effects. We assume that the time series dimension, T, is 200 and that the cross sectional dimension, N, is 180 so that the panel is close to ours. By assuming AR(1) processes for the panel units we minimize the defects inherent in the indirect approach because the non-linearity here is less severe than for higher order autoregressive processes. We then apply the direct and indirect versions of the Kilian (1998) procedure to the MG and MG CCE estimators and the So and Shin (1999) procedure to the MG estimator.[28]

Table 3 presents the results. We report least-squares results as well as bias-corrected estimates, using both the direct and indirect approaches to the bootstrap procedure and, for the simple MG estimation the So and Shin procedure. Several results are worth mentioning. First, the least squares MG estimates display a negative bias when the common correlated effect is relatively unimportant. However, as common components rise in importance, simple MG becomes increasingly inaccurate, as the bias due to the neglected common effects starts dominating the least squares bias. Finally, a *positive* bias arises. This suggests the simulations in CMS, which do not allow for common effects, are not well-tailored to *our data,* where common effects are important. The least squares MG CCE estimator has a negative - but very small - bias regardless of the importance of the common effect.

Second, using the indirect approach, or the So and Shin procedure to correct the MG-based estimates is only accurate when the common effect is relatively small. As common effects rise in importance, both induce a *positive* bias, which can be large. Our suggested direct approach works better but still implies a systematic positive bias. In contrast both bootstrap procedures (direct or indirect) are accurate for the MG-CCE estimator, and give relatively precise estimates that appear immune to the properties of the common effect.

In summary, the results in Table 3 suggest that, in the presence of correlated residuals, the preferred estimate should unsurprisingly be MG-CCE. Then, the

---

[27]We also use a balanced version of our dataset in order to diminish the (considerable) computational burden of the experiments. If anything, truncation works against us, since it makes it harder to obtain precise half-life estimates.

[28]In the Monte Carlo experiments we perform only the first bootstrap step of the Kilian procedure. It would be computationally infeasible to implement bootstrap-after-bootstrap in the panel setting because of time constraints and cycling. Cycling would imply that the simulated distribution would not emulate the asymptotic distribution. The empirical estimates, however, *do* apply bootstrap-after-bootstrap.

Kilian (1998) bias correction procedure is accurate, especially if computed using the direct approach. In contrast, applying bias reduction techniques to the MG estimator not allowing for common effects results in corrected estimates that are themselves biased *upwards*, indeed overestimating the true degree of persistence. This is especially true when the indirect approach is used.[29] This may well explain the results in CE, who apply the So and Shin (1999) method, and the indirect version of the Kilian (1998) procedure.

In IMRR (2005) we apply the bootstrap-after-bootstrap procedure to our panel of relative prices. We find a very modest increase in the estimated half-life. In particular, the corrected half-life estimate increases to eighteen months only, as opposed to our uncorrected estimate of eleven months. Our confidence interval is narrow and excludes the "consensus view". The indirect approach yields slightly higher but very comparable results, with a corrected half-life of twenty months. As suggested in our simulations, choosing the direct or indirect approach makes little difference when common effects are accounted for. It does however change the results dramatically (and push corrected estimates upwards) when the estimation does not allow for common effects, as in CE. This confirms the robustness of our results, and accounts for the discrepancy with CE's. They simply use estimators and bias correction procedures that are inappropriate in price data.

6   Conclusion

Existing dynamic stochastic general equilibrium models featuring plausible degrees of nominal rigidity cannot replicate the persistence of the *aggregate* real exchange rate. This should not be surprising. One sector models should only match the moments of data purged from any heterogeneous dynamics. In IMRR (2005), we estimate the half-life of the real exchange rate in Eurostat data and find it is roughly 11 months once one controls for heterogeneous sectoral dynamics. These estimates are entirely compatible with existing calibrated models. Thus, IMRR (2005) propose that the PPP puzzle, as defined by Rogoff (1996), is due to an aggregation bias. Our results were obtained for US and European real exchange rates, using two-digit sectoral data. Interestingly, Crucini and Shintani (2004) have subsequently confirmed the importance of an aggregation bias, using data disaggregated at a finer level and for a sample of countries that includes emerging markets.

Our paper opens several lines of research. First, it is undoubtedly of great interest to explore whether introducing heterogeneous sectoral dynamics in general equilibrium models with nominal rigidities will enable these models to replicate

---

[29] The RMSE of the indirect approach to the bias correction for the MG estimator is large, while both corrections have quite small RMSE when MG CCE estimator is implemented.

the persistence of the aggregate real exchange rate. Promising first steps in this direction have already been taken, for instance by Carvalho (2005). Second, we stress the importance of initial conditions in simulations aimed at quantifying small sample and aggregation biases. We also emphasize the relevance of the technique used to correct for small-sample bias: correcting the half-life is not the same as directly correcting an autoregressive parameter estimate. Taken together, these subtleties go a long way towards explaining some of the discrepancies between Imbs, Mumtaz, Ravn and Rey (2005), Chen and Engel (2005) and Reidel and Szilagyi (2005). A rigorous, analytical and systematic study of these technical refinements would be welcome and useful. Finally, the concept of aggregation bias may be relevant for other macroeconomic variables. This is an avenue we are currently exploring.

Appendix A: Asymptotic equivalence of RCM and MG estimators

Pesaran (2003) demonstrates the asymptotic equivalence between the RCM estimator and the MG estimator and here we summarize the key parts of Pesaran's analysis.

Consider a panel with two cross sections and t_i observations. Let the OLS coefficients be $\hat{b}_1$ and $\hat{b}_2$ and covariance matrices V_1 and V_2 where:

$$V_i = E\left[(\hat{b}_i - b_i)(\hat{b}_i - b_i)'\right]$$

where b_i is the true value of the coefficient. Consider the MG estimator in this panel:

$$\beta_{mg} = \frac{1}{2}(b_1 + b_2)$$

The RCM estimator is:

$$\beta_{rcm} = \left(\frac{1}{2}\frac{b_1^2 + b_2^2 - 2b_1b_2 + 2V_2}{b_1^2 + b_2^2 - 2b_1b_2 + V_2 + V_1}\right)b_1 + \left(\frac{1}{2}\frac{b_1^2 + b_2^2 - 2b_1b_2 + 2V_1}{b_1^2 + b_2^2 - 2b_1b_2 + V_2 + V_1}\right)b_2 = \Psi_1 b_1 + \Psi_2 b_2$$

As $t_i \to \infty$, $\left(\hat{b}_i - b_i\right) \longrightarrow^p 0$ and $\Psi_i \longrightarrow \frac{1}{2}$ as V_i gets smaller. In other words for large T, the variance of the estimators gets very small and the RCM weights approach the MG weights.

Appendix B. Direct Bootstrapping Method

The bias correction procedure can be summarized as follows:

Step 1 Use the appropriate estimator to get group specific slopes, intercepts and error variances. Denote the estimated mean slope coefficients by $\widehat{\rho}^S$ where S denotes the relevant estimator. Compute the implied half-life and denote this by $\widehat{T}^S_{\frac{1}{2}}$.

Step 2 Generate bootstrap samples of the innovations $\left[\widehat{\varepsilon}^j_{it}\right]^T_{t=1}$ for $i = 1, ..N$ where j denotes replications. We generate these using a non-parametric bootstrap. When we allow for cross-sectional dependence, we first pre-whiten the residuals and then re-color them after the non-parametric bootstrap. Generate artificial samples of relative prices.

Step 3 Use the artificial data to estimate the mean coefficients of the model: i.e. $\widetilde{\rho} = \frac{1}{N} \sum \widetilde{\rho}_i$. We use the method detailed in Kilian (1998) to generate the cross section specific coefficients $\widetilde{\rho}_i$. Use $\widetilde{\rho}$ to calculate the half-life $\widetilde{T}^S_{\frac{1}{2}}$.

Step 4 Repeat steps 2 and 3 R times and obtain the bootstrap average $\frac{1}{R} \sum \widetilde{T}^S_{\frac{1}{2}}$.

Step 5 The bias corrected half life is given by $2\widehat{T}^S_{\frac{1}{2}} - \frac{1}{R} \sum \widetilde{T}^S_{\frac{1}{2}}$

References

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

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

Carvalho, Carlos, 2005, ""Heterogeneity in Price Setting and the Real Effects of Monetary Shocks"", mimeo Princeton University.

Chen, Shiu-Sheng, and Charles Engel, 2005, "Does "Aggregation Bias" Explain the PPP Puzzle?", *Pacific Economic Review* 10, February 2005, 49-72.

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

Crucini, Mario J., and Mototsugu Shintani, 2002, "Persistence in Law-of-One-Price Deviations: Evidence from Micro-Data", Working Paper No. 02-W22, Vanderbilt University.

Crucini, Mario J., and Mototsugu Shintani, 2004, "Persistence in Law-of-One-Price Deviations: Evidence from Micro-Data", Working Paper No. 02-W22R, Vanderbilt University.

Di Giovanni, Julian, 2003, "The Return of the Half-Life", mimeo Berkeley

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

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

Imbs, Jean M., Haroon Mumtaz, Morten O. Ravn, and Hélène Rey, 2002, "PPP Strikes Back: Aggregation and the Real Exchange Rate", NBER WP no. 9372.

Imbs, Jean M., Haroon Mumtaz, Morten O. Ravn, and Hélène Rey, 2003, "The PPP Controversy - A Summary of the Debate surrounding 'PPP Strikes Back' ", posted at http://www.hec.unil.ch/jimbs/, http://www.iue.it/Personal/Ravn, and http://www.princeton.edu/~hrey/.

Imbs, Jean M., Haroon Mumtaz, Morten O. Ravn, and Hélène Rey, 2005, "PPP Strikes Back: Aggregation and the Real Exchange Rate", Quarterly Journal of Economics, February, Vol CXX, Issue 1, pp. 1-43.

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

Murray, Christian and David Papell, 2003, "Do Panels Help Solving the Purchasing Power Parity Puzzle?", mimeo, Houston University.

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

Pesaran, M. Hashem, 2003, "Cambridge University Lecture Notes", Cambridge University.

Pesaran, M. Hashem, and Ron P. Smith, 1995, "Estimating Long-Run Relationships from Dynamic Heterogeneous Panels", *Journal of Econometrics*, Vol.68, 79-113.

Pesaran, M. Hashem, and Zhongyun 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.

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

Pudney, Stephen E., 1978, "The Estimation and Testing of Some Error Components Models", mimeo, London School of Economics.

Reidel, Demian and Jan Szilagyi, 2005, "A Biased View of PPP", mimeo Harvard University.

Rogers, John H. and Michael Jenkins, 1995, "Haircuts or Hysteresis? Sources of Movements in Real Exchange Rates", *Journal of International Economics* 38 (3-4), 339-60.

So, B.S., and D.W. Shin, 1999, "Recursive Mean Adjustment in Time Series Inferences", *Statistics and Probability Letters* 43, 65-73.

Sul, Dongguy, Peter C.B. Phillips, and Chi-Young Choi, 2002, "Prewhitening Bias in HAC Estimation", mimeo.

Table 1			
Method	p	$\sum \rho$	Half-Life
FE	*12*	*0.97767*	*35*
RCM	*5*	*0.97951*	*35*
MG	**5**	**0.97063**	**25**
Hausman Test	80.804 (0.000)		

Table 2			
Method	p	$\sum \rho$	Half-Life
FE	*12*	*0.97757*	*33*
RCM	*5*	0.97247	*26*
MG	**5**	**0.96334**	**21**
Hausman Test	37.068 (0.000)		

Table 3. Monte Carlo Evidence on Bias Corrections

	True	Simple MG Estimator				MG CCE Estimator		
		Least Sq.	Indirect	Direct	So-Shin	Least Sq.	Indirect	Direct
$\lambda = 0$	14.54	10.78	14.96	13.06	14.81	10.11	16.45	12.82
$\lambda = 0.08$	14.54	11.14	15.74	13.57	15.45	10.16	16.58	12.88
$\lambda = 0.16$	14.53	11.47	16.43	14.04	16.25	10.23	16.76	12.97
$\lambda = 0.24$	14.53	12.02	17.73	14.85	17.17	10.32	16.83	13.08
$\lambda = 0.32$	14.54	12.69	19.45	15.81	18.70	10.41	16.93	13.17
$\lambda = 0.40$	14.52	13.34	21.24	16.80	20.24	10.54	17.06	13.33
$\lambda = 0.48$	14.54	14.35	24.23	18.32	22.89	10.77	17.14	13.56
$\lambda = 0.56$	14.53	16.08	31.24	21.01	26.60	11.16	17.26	13.93

Notes: The table reports the mean MG estimates of the half-life of relative prices in a Monte Carlo experiment where the data is generated by the process: $q_{it} = \alpha_i + \beta_i q_{it-1} + x_t + \varepsilon_{it}$, $x_t = \lambda x_{t-1} + \xi_t$, $\alpha \sim N(0,1)$, $\beta \sim U[0.93, 0.99]$, $\varepsilon_i \sim i.i.d(0,1)$, $\xi \sim N(0,1)$. We assume that $T = 200$, and that $N = 180$. We initially draw 250 observations but then drop the first 50 to lower the impact of the initial condition. The column "True" reports the true half-life based on the impulse response function. The columns "least squares" report the least squares estimates of the half-life based on either the MG or the MG CCE estimators. The column denoted "So-Shin" reports the results of implementing the So and Shin (1999) bias correction to the MG estimator. The columns "indirect" and "direct" report the results of implementing the indirect and the direct versions of the bias reduction methods based on the Kilian (1998) bootstrap procedure. The number of replications is 1000.

Estimates using Charles Engel's Data

Table 4: 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	5	0.98	35 $(26, 43)$	0.97 $(0.962, 0.977)$	48.72
Fixed Effects (SURE)	5	0.98	35 $(28, 45)$	0.98 $(0.972, 0.983)$	49.53
Generalised Fixed Effects	5	0.99	50 $(24, 152)$	0.99 $(0.970, 0.995)$	71.35
Mean Group	5	0.97	25 $(9, 31)$	0.97 $(0.861, 0.974)$	35.09
Mean Group CCE	5	0.94	13 $(9, 24)$	0.93 $(0.827, 0.959)$	17.18
Mean Group (Sure)	5	0.97	21 $(17, 25)$	0.96 $(0.954, 0.973)$	29.50
[a]$H0 : \rho_i = \rho$	80.804 (0.000)				

26

Figure 1

Figure 2

RCM

All Cross sections
100 Cross sections

Figure 3

Figure 4

Half-Lives: True=13.51, Not "Corrected"=13.52, "Corrected"=22.5

Figure 5

Example of measurement error

Figure 6

OLS

GMM