

Aggregation is not the solution: the PPP puzzle strikes back.*

BY M. DOLORES GADEA^a AND LAURA MAYORAL^b

^aUniversity of Zaragoza; ^bInstitute for Economic Analysis, (CSIC)

June 2007

Abstract

The Purchasing Power Parity (PPP) Puzzle refers to the difficulty of reconciling the enormous short-run variability of real exchange rates with their longer-than-expected deviations from equilibrium (Rogoff, 1996). Recently, Imbs, Mumtaz, Ravn and Rey (2005, hereinafter IMRR) have argued that much of the PPP puzzle is due to upwardly-biased estimates of persistence. According to their view, the source of the bias is the existence of heterogeneous price adjustment dynamics at the sectoral level. As they put it, “the aggregate real exchange rate is persistent because its components have heterogeneous dynamics”.

This paper re-examines this claim in two steps. Firstly, we demonstrate that IMRR’s measures of sectoral persistence are systematically downwardly-biased because they are based on an inaccurate definition of the “average” Impulse Response Function (IRF). We then show that standard estimates of shock persistence are recovered after this bias is corrected. Secondly, building on the results in Mayoral (2007), that prove that aggregate and micro models induce similar shock persistence behavior, we show that estimates based on aggregate and sectoral exchange rates are in fact highly compatible. Therefore, aggregation does not solve the puzzle and further research is still necessary.

JEL classification: C22, F31, O11.

Keywords: PPP puzzle, real exchange rates, aggregation bias, persistence, heterogeneous dynamics.

*We are very grateful to M. Arellano and A. Marcet for their valuable comments. Corresponding author: laura.mayoral@iae.csic.es.

1. INTRODUCTION

The so-called purchasing power parity puzzle is considered to be among the six major puzzles in international economics (Obstfeld and Rogoff, 2000). The puzzle refers to the difficulty of reconciling the high volatility of exchange rates with long-lasting deviations from their equilibrium levels, as defined by the theory of purchasing power parity (PPP). Rogoff (1996) highlighted this problem and, based on a reading of panel and long-span studies, noticed that the estimated half-lives (HLs) of real exchange rate adjustment tend to fall into the range of three to five years. On the one hand, explanations of short-term exchange rate volatility point to financial factors (asset price bubbles, monetary shocks, etc.). On the other, the slow adjustment to PPP can be easily justified in models where real shocks (such as shocks to tastes or to technology) are predominant. The puzzle arises because existing models based on real shocks cannot account for the large short-term exchange rate volatility.

The literature documenting and trying to find an explanation for the puzzle is very large. Some authors have noticed that Rogoff's consensus of 3 to 5 year half-lives of PPP deviations was based on univariate or panel studies using OLS estimates, which are known to be biased downwards. When the bias is corrected, it is generally found that HL point estimates are well above the "consensus view", implying that the size of puzzle is even larger than was originally believed (see Murray and Papell, 2002, 2005, Lopez et al., 2003, 2004).

In the opposite direction, there have been several attempts to solve the puzzle, most notably departures from linearity (such as nonlinear dynamics in real exchange rate adjustment or the existence of structural breaks)¹ and, in a linear setting, aggregation problems due to heterogeneity in the speed of price adjustment at the goods level, as advocated by Imbs, Mumtaz, Ravn and Rey (2005, hereinafter IMRR). The present paper looks at the latter potential solution to the PPP puzzle.

IMRR argue that estimated half lives are so large because the corresponding estimates are biased upwards. According to their view, the existence of heterogeneous dynamics at the sectoral level (that is neither taken into account explicitly nor handled in an appropriate manner in most

¹By introducing non-linearities into the real exchange rate adjustment, several authors have succeeded in enlarging the evidence of reversion, as in Michael et al. (1997) and Taylor et al. (2001). In the approach that considers structural breaks, Hegwood and Papell (1998) and Gadea et al. (2004), breaks have been able to reduce half-lives noticeably.

studies based on time series or panel data) gives rise to an ‘aggregation bias’ when aggregate data is used to draw inferences about the speed of price adjustment. By employing sectoral real exchange rates and explicitly allowing for heterogeneity, they report estimates of price adjustment that are completely in line with models of slow nominal price adjustment, with an ‘average’ half-life of price adjustment of about 1 year. Hence, they claim to have solved this long-debated puzzle and conclude that *“the aggregate real exchange rate is persistent because its components have heterogeneous dynamics”*.²

However, we argue that aggregation is not the solution to the puzzle. We build our argument in two steps. Firstly, we show that the IMRR measures of persistence computed with sectoral data systematically underestimate (average) persistence. IMRR’s conclusions are basically drawn from the analysis of the HL, which, in turn, is computed from an ‘average’ impulse response function. The source of the bias is precisely the definition of average impulse response function (IRF) used by these authors. Instead of computing the individual impulse responses and averaging them in order to produce an estimate of the average impulse response, they first estimate the mean value of the (heterogeneous) model coefficients in a panel of countries and, then, use this value to estimate their ‘average’ IRF, as if the model was one of homogeneous coefficients given by the mean value of the heterogeneous parameters. Since the IRF is a highly nonlinear function, averaging the IRFs or averaging the coefficients and then computing the IRF may yield very different results. In fact, Jensen’s inequality ensures that, for most empirically relevant cases, the former measure is always larger than the latter. The intuition of this result is clear: the IRF grows faster than linearly for highly persistent sectors. Hence, when averaging the individual responses, these highly persistent sectors increase the mean considerably. However, in the computations of IMRR, highly persistent sectors are eliminated in the first stage when averaging model coefficient’s estimates, so that their impact on average persistence is much smaller. This translates, not surprisingly, into lower persistence estimates. Using the same data set and the same estimation strategy as that employed in their paper, we have quantified the size of the bias that affects IMRR measures of persistence. It turns out that the bias is substantial and that, once it is corrected, persistence estimates increase considerably. Moreover, the classical results of 3-5 year half-lives of PPP deviations are recovered. It is important to emphasize that

²This paper has generated a considerable debate, see Engel and Chen, (2005) and Imbs et al. (2004). Nevertheless, our arguments are very different from those discussed in the above-mentioned articles.

the only difference between their results and the ones reported in this paper stems from the definitions of average IRF employed, since in all other aspects we have closely followed their estimation approach.

Secondly, we consider the question of whether an aggregation bias exists, that is, whether persistence, as measured by the IRF, differs across aggregation levels. We use the results in Mayoral (2007) that show that the standard IRF associated with the aggregate model is simply the expected value of the individual responses. This implies that aggregate persistence is directly determined by (average) sectoral persistence. In other words, the aggregate is persistent if the sectors are, on average, persistent, but not because they present heterogeneous dynamics, as argued by IMRR. We illustrate this theoretical result by showing that standard time series techniques allow one to obtain measures of persistence computed with aggregate data that are highly compatible with those obtained using sectoral data.

Thus, in our opinion the results in IMRR should be read as follows: if the economy was characterized by a representative sector (that they compute by averaging the coefficient estimates of the individual models), then the speed of price adjustment would be compatible with models based on nominal rigidities. But then, this analysis is far from explaining the PPP puzzle and it gives raise to two important questions: firstly, what is the interpretation of such an artificial representative sector, which is not constructed as an average of the sectors themselves but rather, by averaging only the coefficients of the individual models? And secondly, even if one considers that this could be a valid ‘average’ of the individual sectors, the fact that this ‘average’ is low-persistent does not solve the puzzle because it does not explain why some sectors are so persistent.

Summarizing, our results suggest that the different persistence behavior between aggregate and sectoral exchange rates reported by IMRR is not due to an upwards bias in the estimates based on aggregate data deriving from the existence of sectoral heterogeneity but rather, to a negative bias affecting their sectoral persistence estimates.

Hence, the bad news are that we do not yet have a convincing solution to the PPP puzzle since aggregation seems not to be the solution. The good news, however, are that applied macroeconomists can still rely on aggregate data for their studies since estimates derived from these data are trustable and should be very in much in line with micro estimates even when the assumption of individual homogeneity is violated.

The outline of the paper is as follows. Section 2 summarizes the main theoretical arguments needed to establish our results. Section 3 presents our estimates of persistence based on sectoral data and quantifies the magnitude of the negative bias affecting IMRR’s estimates of persistence computed with sectoral data. Section 4 reports measures of persistence computed with aggregate data and shows that they are highly compatible with the ones obtained in Section 3, illustrating the lack of “aggregation bias”. Section 5 concludes.

2. MEASURING PERSISTENCE AT DIFFERENT AGGREGATION LEVELS

In this section we present the theoretical background needed for developing the empirical results. We consider two aspects. Firstly, we analyze the issue of measuring (average) persistence with sectoral data and describe the bias affecting IMRR sectoral estimates. Secondly, we study the relation between the (population) IRFs associated with the aggregate and the sectoral models. It is shown that the IRF of the former is simply the expected value of sectoral impulse responses. This implies that the aggregate process is persistent if the sectors are themselves, on average, persistent, but not because they present heterogeneous dynamics. It also suggests that, provided the aggregate and the disaggregate models are properly specified, highly compatible measures of persistence should be found, irrespective of the level of aggregation. The latter results are fully developed in Mayoral (2007) and we only present a brief summary here.

2.1. Measuring persistence with sectoral data

IMRR consider a panel of sectoral exchange rates for several European countries defined against the U.S. dollar.³ In its simplest version, they assume that for each country c , each sector in the panel can be represented as (see equation (1) in IMRR),

$$q_{cit} = \gamma_{ci} + \rho_{ci}q_{cit-1} + u_{cit}, \quad i = 1, \dots, N, \quad c = 1, \dots, C; \quad t = 1, \dots, T, \quad (1)$$

where i , c and t denote sector, country and period, respectively, q_{cit} is the real exchange rate for country c , sector i at time t , $\gamma_{ci} = \bar{\gamma} + \eta_{ci}^\gamma$, $\rho_{ci} = \bar{\rho} + \eta_{ci}^\rho$, $\bar{\gamma}$ and $\bar{\rho}$ are constants, and ρ_{ci} has support on the interval $(0, 1)$. Furthermore η_{ci}^γ and η_{ci}^ρ are *i.i.d* zero-mean random variables,

³Sectoral exchange rates are defined as $q_{ict} = \log(S_{ct}P_{ict}/P_{i,US,t})$, where S_{ct} denotes the nominal bilateral exchange rate between the US and country c at date t , P_{ict} is the price of good i in country c at time t while $P_{i,US,t}$ is the corresponding US price.

mutually independent of u_{it} , which is independently distributed, with zero mean and variance σ_{ci}^2 . Contemporaneous correlation across sectors and countries is allowed.

As argued by IMRR, impediments to arbitrage or nominal rigidities vary considerably across goods. Since these factors are usually believed to be behind cross-country price differences, these impediments could bring about important heterogeneity in the speeds of reversion to parity across sectors and countries.⁴ Model (1) can account for different sources of heterogeneity: in addition to country and sector fixed effects (captured by the parameter γ_{ci}), it also allows for different speeds of shock adjustment by letting ρ_{ci} be heterogeneous.

How could one compute a measure that summarizes the persistence of a collection of sectoral real exchange rates? One of the most popular tools for shock persistence evaluation is the impulse response function, defined as the “effect of a change in the innovation by a unit quantity on the current and subsequent values of the variable of interest” (Andrews and Chen, 1994, p.189). For each unit in the panel, the response to a unitary shock, occurring at time t , h periods ahead, can be computed as the difference between two forecasts (see Koop et al., 1996),

$$IRF^{ci}(t, h) = \mathcal{E}(q_{cit+h}|u_{cit} = 1; z_{cit-1}) - \mathcal{E}(q_{cit+h}|u_{cit} = 0; z_{cit-1}), \quad (2)$$

where the operator $\mathcal{E}(\cdot)$ denotes the best mean squared error predictor and $z_{it-1} = \left(q_{cit-1}, q_{cit-2}, \dots \right)'$; Applied to the simple model in (1), it yields that the response of sector i in country c to a unitary shock in t , h periods ahead is

$$IRF^{ci}(t, h) = \rho_{ci}^h, \text{ for } h \geq 0. \quad (3)$$

If one is interested in the average response across sectors to a unitary shock, a natural measure of average persistence would be to consider the expected value of (3) over the distribution of units. The expected impulse response to a unitary shock h periods ahead is then given by

$$E(IRF(t, h)) = E\left(\rho^h\right), \text{ for } h \geq 0, \quad (4)$$

where $E(\cdot)$ denotes expectation across the distribution of units (sectors and countries). Then, the expected IRF associated with (1) is given by the h_{th} – *moment* of the distribution of ρ . From this expression, it is straightforward to define other popular measures of shock persistence, such as the half life (HL), defined as the number of periods it takes until half the effect of a shock

⁴See Cheung et al. (2001) and Bils and Klenow (2002) for some evidence.

dissipates, and the cumulated impulse response (*CIR*), which measures the total cumulative effect of a shock over time. Application of these definitions to the mean IRF defined in (4) allows us to compute the HL as the value of h that verifies

$$E(IRF(t, h = HL)) = 0.5, \quad (5)$$

whereas the CIR is

$$CIR = \sum_{i=0}^{\infty} E(IRF(t, i)). \quad (6)$$

Let us now revise how the calculations reported in IMRR relate to the measures defined in (4), (5) and (6). Their approach is to estimate the expected value of ρ_{ci} , $\bar{\rho}$ and, in a second step, to compute the IRF defined in (4), as if the true DGP was given by $q_{cit} = \bar{\gamma} + \bar{\rho}q_{cit-1} + a_{cit}$, for all $i = 1, \dots, N$, $c = 1, \dots, C$, that is, as if the DGP was a panel with a homogeneous autoregressive parameter given by $\bar{\rho}$. Therefore, they provide estimates of the function,

$$\overline{IRF}(t, h) = \bar{\rho}^h = E(\rho)^h, \text{ for } h \geq 0. \quad (7)$$

They estimate ρ according to different approaches and they plug these values into (7) to produce different IRF estimates, finding, in general, HL estimates considerably lower than those implied by the “consensus view”.

Clearly, under heterogeneity, (7) does not correspond to the average of the individual responses, which is defined in (4). Furthermore, it can be easily seen that, in most empirically relevant cases, (7) systematically underestimates the true average response. Whenever the support of ρ is positive, which is a very realistic assumption in this case, then ρ^h is strictly convex and application of Jensen’s inequality yields

$$E(IRF(t, h)) > \overline{IRF}(t, h), \text{ for all } h > 1, \quad (8)$$

or, in other words, (7) systematically underestimates average shock response.⁵ Since the HL and the CIR are directly computed from the IRFs above, the same inequality also holds for these measures.

The relation established in (8) does not only hold in the simple AR(1) case but also for more general *AR* dynamics. For instance, for heterogeneous AR(2) processes, whenever the support

⁵Although, in their empirical exercise, more general AR(p) dynamics are considered, the same procedure for obtaining the average response to shocks is employed and thus, similar criticisms apply.

of the first autoregressive coefficient ρ_1 is positive (which implies that the largest autoregressive root is greater than zero and greater in absolute value than the other root), a similar inequality holds. Since we are dealing with very persistent processes, this is a very realistic situation. More generally, in the $AR(p)$ case, the individual IRF can be written for large h as (see Rossi, 2005)

$$IRF^i = \alpha_{1i}^h b(1)^{-1},$$

where α_{1i} is the largest autoregressive root and $b(L) = (1 - \alpha_{2i}) \dots (1 - \alpha_{pi})$ is the polynomial containing the remaining autoregressive roots. Again, it can be seen that, provided the support of α_1 is positive, then IRF is a convex function and Jensen's inequality ensures the result above.

To illustrate the inequality in (8), Figure 1 presents the graphs of several IRFs corresponding to model (1), computed according to (7) or to (4). The heterogeneous autoregressive lag coefficient has been generated according to a Uniform as well as to a Beta distributions (top and bottom graphs, respectively), with mean values of ρ equal to 0.8 and 0.9, for each of the distributions.⁶ The solid line depicts the expected IRF as defined in (4) while the dashed line shows the IRF computed as in IMRR. Two main conclusions can be drawn. Firstly, the gap between the two lines is considerable, even at long horizons. Secondly, the denser the distribution is around 1, the higher the gap, as illustrated by the IRFs corresponding to AR coefficients generated with the beta distribution. Since real exchange rates are highly persistent, these plots suggest that the bias implied by (7) can be very large.

(Figure I about here)

Summarizing, the results above show that IMRR persistence estimates computed with sectoral data are likely to underestimate the true average shock response of the real exchange rates. But, so far, we have not said anything about their major claim, namely, that the existence of heterogeneous dynamics at the sectoral level introduces a positive bias into estimates of persistence computed with aggregate data. We analyze this argument in the following subsection.

⁶The Uniform distribution has been defined in the (0.6,0.99) and (0.8,0.99) intervals while the parameters for the Beta distribution are (p=2.5, q=0.5) and (p=4.5, q=0.5) corresponding to the cases $E(\rho) = \{0.8, 0.9\}$, respectively.

2.2. Comparing aggregate and sectoral persistence

We now turn to examine the relation between our preferred measure of sectoral persistence, the average of sectoral IRFs as defined in (4), and the standard IRF associated with the model obtained by aggregating (1) across sectors. This question has been addressed in Mayoral (2007) and we only summarize those results briefly here.

Under homogeneity, the relation between disaggregate and aggregate IRFs is trivial. The aggregate and individual model dynamics are the same and, therefore, the IRF also remains the same across aggregation levels. Nevertheless, under individual heterogeneity, aggregation of (1) yields a process with rather different dynamics than the micro units, as has been shown by many authors.⁷ Hence, before deriving the IRF of the aggregate model we explicitly consider aggregation of (1) across sectors. This issue has been considered by Lewbel (1994), who followed the approach introduced by Stoker (1984). The latter author defines an aggregate function as one given by the expected value across individuals of the disaggregate relations. Hence, the aggregate real exchange rate for country c could be obtained as

$$Q_{ct} = E(\gamma_c) + E_t(\rho_c Q_{ct-1}) + U_{ct}. \quad (9)$$

where $E_t(\cdot)$ denotes expectation over the cross-section distribution and provides the time path of the dependent variable mean, $Q_{ct} = E_t(q_{ct})$ is the aggregate real exchange rate for country c at time t and $U_{ct} = E_t(u_{ct})$ is an aggregate shock. Lewbel (1994) showed that, under certain assumptions, expression (9) is equivalent to,

$$Q_{ct} = \sum_{s=1}^{\infty} A_s Q_{ct-s} + U_{ct}, \quad (10)$$

for constants A_1, A_2, \dots that satisfy the equation

$$A_s = m_s - \sum_{r=1}^{s-1} m_{s-r} A_r. \quad (11)$$

where $m_s = E(\rho^s)$ is the moment of order s of ρ . It follows that, under heterogeneity, the aggregate model displays very complicated dynamics even when the behavior of the micro units

⁷See Granger (1980), Robinson (1978) and Zaffaroni, (2004).

is very simple, as it is in this case. The standard aggregate IRF associated with model (10) can be computed as the difference between the forecasts,

$$IRF_{AG}(t, h) = \mathcal{E}(q_{ct+h}|U_{ct} = 1; Z_{t-1}) - \mathcal{E}(q_{ct+h}|U_{ct} = 0; Z_{t-1}), \quad (12)$$

where $Z_{t-1} = (Y_{t-1}, Y_{t-2}, \dots)$. Application of this definition to (10) yields,

$$\begin{aligned} IRF_{AG}(t, 1) &= A_1; \quad IRF_{AG}(t, 2) = A_1^2 + A_2 \\ IRF_{AG}(t, 3) &= A_1(A_1^2 + A_2) + A_2A_1 + A_3, \end{aligned}$$

and in general,

$$IRF_{AG}(t, h) = \sum_{j=1}^h A_j IRF_{AG}(h-j).$$

Now, noticing that (11) can be rewritten as $m_s = \sum_{r=0}^{s-1} m_r A_{s-r}$, one can iterate the latter expression and check that

$$\begin{aligned} IRF_{AG}(t, 1) &= m_1 = E(\rho), \\ IRF_{AG}(t, 2) &= m_2 = E(\rho^2), \\ &\dots \\ IRF_{AG}(t, h) &= m_h = E(\rho^h). \end{aligned}$$

That is, the aggregate IRF equals the non-centered moments of the distribution of the AR coefficients which, in turn, coincide with the expected value of the individual IRFs, defined in (4). It turns out that our preferred measure of (average) sectoral persistence coincides with the aggregate IRF, implying a tight relation between sectoral and aggregate shock response. This result can be extended to more general micro *AR* dynamics, as shown in Mayoral (2007).

Two important considerations should be emphasized at this point. Firstly, since the population values of aggregate and (average) sectoral persistence are the same, under the usual hypothesis of correct specification, similar estimates should be obtained with either type of data, as in the case where no individual heterogeneity is present. There is, however, an important difference between the homogeneous and the heterogeneous case. In the latter situation, the dynamics of the aggregate process become very complex in such a way that Q_{ct} might not admit a representation with a finite number of parameters, as illustrated by (10). Notice, however, that, even in this situation, it is still possible to obtain consistent estimates of the autoregressive

lag parameters of the aggregate model provided a sufficiently long AR model is specified, where the number of lags grows at an appropriate rate with respect to the sample size. See Berk, (1974) and Ng and Perron, (2001).

To illustrate that these theoretical considerations hold for reasonable sample sizes and when very standard techniques are employed, we have run a small Monte Carlo experiment. We have generated 300 heterogeneous AR(1) processes, allowing for heteroskedasticity and contemporaneous correlation across individuals. Two distributions were employed to generate the autoregressive coefficients, namely, a Uniform and a Beta distribution, with mean equal to 0.8 in both cases.⁸ The aggregate process was computed as the arithmetic average of the micro units. The sample size was set at 180, to match the sample size of this paper's data set. Next, an AR(p) process was fitted, where p was chosen according to the AIC. The maximum number of lags was set at 20 (the maximum number of lags usually considered by IMRR in their calculations). Notice that from equation (11), A_1 , the first AR coefficient of the aggregate model, equals the mean of the AR(1) lag coefficients, $E(\rho)$. In this simple simulation, the mean value of A_1 was found to be equal to 0.81 and 0.79, with standard deviations of 0.080 and 0.086, for the uniform distribution and the beta distributions, respectively, which are very good estimates of the true value of $E(\rho) = 0.8$.

Secondly, the results above highlight the source of the bias reported in the analytical calculations of IMRR. In order to show that aggregate time series data overestimate persistence, they consider the same model specification for the aggregate and the for disaggregate data, namely, an AR(1) specification. They consider whether the estimate of the first AR coefficient of the aggregate model is a consistent estimator of $E(\rho)$. But, as is clear from (10), the aggregate model is misspecified and, not surprisingly, \hat{A}_1 is not a consistent estimator of $A_1 = E(\rho)$. Thus, it is important to emphasize that the source of the bias discussed in IMRR is the misspecification of the aggregate model rather than a bias arising as a consequence of the aggregation of heterogeneous processes. As illustrated above, consistent estimates of $E(\rho)$ can be obtained when aggregate data is employed provided a longer AR structure is specified.

⁸The uniform distribution was defined in the intervals (0.65, 0.95). The p, q parameters in the beta(p,q) distribution were set at 4 and 1, respectively. The error in the AR(1) units was defined as the sum of two Gaussian shocks, one idiosyncratic and the other common to all units. The number of replications was 1000.

3. RESULTS FOR SECTORAL DATA

This section quantifies the magnitude of the negative bias affecting IMRR measures of sectoral shock persistence. We employ the same data set as in their paper, that is, nineteen monthly exchange rates covering, at most, the period 1981:1 to 1995:12. (Non-harmonized) price indexes are provided by Eurostat and real exchange rates (RERs) are defined against the U.S dollar. See IMRR, Appendix 3, for more details.

IMRR consider the model

$$q_{ict} = \gamma_{ci} + \sum_{k=1}^K \rho_{cik} q_{cit-k} + u_{cit}, \quad (13)$$

where they assume that sectors are homogeneous across countries, so that $\gamma_{ci} = \gamma_i$, $\rho_{cik} = \rho_{ik}$ for all i, k . IMRR are interested in the average values across sectors of the autoregressive coefficients, $\bar{\rho}_k$, for $k = 1, \dots, K$. They apply the Mean-Group (MG) estimator (see Pesaran and Smith, 1995), with and without correction for cross-sectional correlation in the errors and with and without correction for downward bias in the OLS estimates. The procedure consists of applying the corresponding panel technique to estimate sector-specific coefficients and then the parameters $\bar{\rho}_k$ are estimated as a simple average of the corresponding sector-specific estimates. To correct for non-zero cross-sectoral correlations in the residuals, the Seemingly Unrelated Regression (MG-SURE) and the Common Correlated Effects estimator (MG-CCE) are implemented (see Pesaran, 2006, and IMRR for details). Finally, they re-compute the MG, MG-SURE and MG-CCE, correcting for the OLS small-sample bias using Kilian's (1998) bootstrap-after-bootstrap method. Next, IMRR use the averages of the original estimates to compute their estimates of sectoral IRF, as described in Section 2.

In order to gauge the magnitude of the bias of IMRR's measure of sectoral persistence, we have closely followed their estimation strategy to obtain sector-specific coefficients. The only difference between IMRR's approach and ours is that, instead of averaging the sectoral estimates and using the resulting averages to estimate an IRF (and the corresponding HL), we estimate an IRF for each of the sectors and then average these functions across sectors. In addition, we have also considered the possibility that sectors are heterogeneous across countries so the above-described calculations have been performed for each of the countries individually.

Tables I and II present our results. To facilitate the comparison, IMRR's notation for the different estimation approaches has been preserved and HLs computed according to their proce-

dure are also reported. Table I contains the HLs obtained by applying the MG estimator (with and without correcting for the OLS small-sample bias). HL_{IMRR} is the HL computed as in IMRR, that is, averaging the AR coefficients and then computing the IRF from these average values. HL_m and HL_w are the HLs associated with a simple arithmetic average and with a weighted average of the individual IRFs, respectively. Ideally, in order to facilitate the comparison with HL measures computed with aggregate data, the weights should be those employed to construct aggregate exchange rates. Unfortunately, Eurostat does not publish non-harmonized price weights so harmonized price weights (corresponding to 2006) have been employed instead. Thus, our preferred measure of sectoral persistence is HL_w , since it weighs sector-specific responses in a similar manner to which the aggregate function weighs sectors. As for the rows, the first one displays panel data estimates (calculated under the assumption that sectors are homogeneous across countries), while the remaining ones present time series estimates, obtained by allowing for cross-country sectoral heterogeneity. In all cases, $AR(p)$ processes were specified. The order of the autoregressive was 19 for the panel data case, to match IMRR's estimates. For the remaining rows, it was chosen according to a general-to-specific criterion with a maximum number of lags of 20, as in IMRR's paper.⁹ Confidence intervals have been calculated using bootstrap techniques.

In order to compute bias-corrected estimates, Kilian's (1998) method has been employed and we have followed an *indirect* approach, that amounts to computing bias-corrected autoregressive coefficients and, then, deriving the corresponding bias-corrected HL from these coefficients. Alternatively, IMRR advocate a *direct* approach, by which the estimated HL is corrected directly using the bootstrap. We have conducted a Monte Carlo analysis to determine which method performs best for our definition of HL. It turned out that the direct approach tends to underestimate the true HL whereas the indirect one performs reasonably well. The intuition of this result is clear. When the HL is computed according to (5) and the data is very persistent, the small-sample bias affecting the autoregressive estimates produces a very large bias in the IRF, since this function is highly sensitive to small changes of the coefficients when they are close to the non-stationarity boundary. As a consequence, the HL is also severely biased downwards and, when the *direct* approach is implemented, the bias is not entirely corrected by the bootstrap algorithm. This problem affects IMRR's calculations to a lesser extent because the impact

⁹The AIC was also employed and very similar results were obtained.

of highly persistent sectors on their measure of *average* persistence is less important than in our case (since sectoral coefficient estimates are averaged in the first stage) so, the resulting “*average*” IRF does not present such a large bias. However, when the IRF is computed using bias-corrected AR estimates, in accordance with the *indirect* approach, very accurate estimates of the HL are obtained.¹⁰

Several conclusions can be drawn from Table 1. Firstly, we are able to match IMRR’s panel estimates very closely and, as expected, HL_{IMRR} figures are always smaller than HL_m and HL_w . In addition, this table allows us to quantify the negative bias affecting IMRR’s estimates. When no small-sample bias correction is introduced (first three columns), the HL_{IMRR} panel estimate is below the “consensus view” (26 months). Nevertheless, the conclusions are reversed when HL_m and HL_w are considered, as they present values slightly above three years, (36 and 37 months, respectively) in line with the standard literature. Allowing for cross-country heterogeneity does not substantially modify the conclusions: the HL_{IMRR} estimates are, in general, below 36 months (with the only exception of Spain), whereas HL_m and HL_w are in general above the latter figure (only GR, NL, FI and UK present values of HL_w below 36 months).

When the OLS small-sample bias is corrected, the gap between HL_{IMRR} and HL_m - HL_w estimates becomes much larger (columns 4 to 6). All estimates increase considerably, suggesting that the negative bias affecting the OLS estimates is, in fact, quite large. This is not surprising since this type of bias is known to be large when OLS is applied to highly persistent data, in which case, the IRF is very sensitive to small changes in the parameters. HL_{IMRR} values are significantly higher than before (and, with few exceptions, lie in the 3-5 year interval). The increase is even more important for the HL_m and HL_w measures, whose point estimates are, in most cases, larger than 15 years and have no finite upper bound.¹¹ It is also remarkable that cross-country heterogeneity increases considerably, raising doubts about the adequacy of panel estimates, that are computed under the hypothesis of cross-country sectoral heterogeneity.

(Table I about here)

If the errors are contemporaneously correlated, as is likely to be the case here, more efficient estimators than OLS can be employed. When N is relatively small with respect to T , the

¹⁰For the sake of brevity the figures are omitted but they are available upon request.

¹¹Similar results have been reported in a purely time series context after small-sample bias correction, by Murray and Papell (2002) and Lopez et al. (2003, 2004).

standard approach is to treat the group of equations as a system of seemingly unrelated equations (SURE) and then estimate the system by GLS, which would be efficient in this case. In addition to the SURE estimates, IMRR also present figures computed according to a common correlated effects procedure (CEE, Pesaran, 2006), based on regressions

$$q_{ict} = \sum_{k=1}^K \rho_{ick} q_{ict-k} + \sum_{h=0}^H \phi_{ich} \bar{q}_{t-h} + e_{ict},$$

where \bar{q} is the cross-sectional average of q_{ic} .

Table II presents analogous figures to Table I but, in this case, the SURE and the CCE estimators have been computed. For the sake of brevity, only small sample bias-corrected figures are reported since, as illustrated in Table I, this bias is substantial.¹²

Accounting for contemporaneous correlation in the errors produces a substantial decrease in persistence estimates but, otherwise, many of the conclusions drawn from Table I are still valid. The HL_{IMRR} estimates are always smaller than HL_m and HL_w and the size of the gap changes considerably with the estimation method. The SURE technique tends to homogenize the model estimates across sectors and, hence, the gap between the corresponding figures for HL_{IMRR} , HL_m and HL_w is smaller. In general, the three measures lie in the 3-5 year interval in this case. However, unlike the SURE, the CCE estimator reduces persistence in the mean but noticeably increases the variability across sectors. This brings about an important reduction in HL_{IMRR} estimates, which are close to those reported in the first column of Table 1, with a panel point estimate slightly higher than one year and a half and an upper bound of less than two years and a half. Nevertheless, the existence of a high variability in coefficient estimates across sectors (more specifically, the fact that a few sectors are very persistent) translates into very large values for both the HL_m and HL_w . It follows that the gap between IMRR's measure and ours is particularly large in this case: while the HL_{IMRR} panel estimate is around 20 months, the HL_m and HL_w panel estimates exceed 180 months. However, notice that according to the CCE estimates countries are very heterogeneous, so one should interpret panel estimates with caution since they are obtained under the assumption of country homogeneity.

¹²When computing panel estimates according to the SURE technique, N is in fact larger than T ($N=204$, $T=180$) and the SURE estimate is not feasible. So, as in IMRR, we use Engel's truncated version of the Eurostat dataset, which has fewer observations than theirs. However, in country-by-country calculations the same data as in Table I has been employed since that problem is not present. The number of lags for computing CCE estimates was chosen according to the AIC.

Summarizing, it turns out that when sectoral persistence is correctly measured, HL estimates are not below the “consensus view” since the standard result of half-lives (HLs) of real exchange rate adjustment falling into the 3 to 5 year range (or even higher values when small-sample bias corrections are introduced) is recovered.

(Table II about here)

4. RESULTS FOR AGGREGATE DATA

The aim of this section is to show that the existence of sectoral heterogeneity at the individual level does not necessarily introduce a bias in measures of persistence computed with aggregate data and that, in fact, highly compatible results can be obtained across aggregation levels using standard techniques. We assume that aggregate prices in country c are constructed as the geometric average of sector-specific prices, that is

$$P_{ct} = \prod_{i=1}^N p_{ict}^{\omega_{ic}}$$

where ω_{ci} are weights that verify $\sum_{i=1}^N \omega_{ci} = 1$ and are not time-varying and that $\omega_{ci} = \omega_{us,i}$ for all i , where $\omega_{us,i}$ are U.S price weights, the bilateral aggregate real exchange rate $Q_{c,t}$ can be written as

$$Q_{c,t} = \sum_{i=1}^N \omega_{ci} q_{ict}. \quad (14)$$

that is, it is a weighted sum of sectoral *RERs*. Since, in order to build a price index, a large number of individual prices are considered, the results in Section 2 suggest that the IRF associated with $Q_{c,t}$ can be reasonably approximated by the weighted average of goods-specific responses (see Mayoral, 2007 and Lewbel, 1994). In reality, however, weights are not equal across countries and $Q_{c,t}$ is equal to

$$Q_{c,t} = \sum_{i=1}^N \omega_i q_{ict} + \sum_{i=1}^N (\omega_{us,i} - \omega_{c,i}) q_{ict}, \quad (15)$$

that is, the aggregate RER is the sum of a weighted sum of individual RERs plus an additional term that captures cross-country differences in price weights. If weights are time-varying, additional terms should be included in (15).

The “aggregation bias” argument states that, even in the situation described in (14), i.e., when the aggregate RER is exactly a weighted sum of sectoral RERs, measures of persistence derived

from $Q_{c,t}$ would tend to overestimate average sectoral persistence if q_{ict} present heterogeneous dynamics. Hence, in order to isolate this potential source of bias from other sources of divergence derived from the non-constant and non-homogeneous character of price weights across time and countries, we have constructed an aggregate variable, computed as the weighted sum of sectoral prices, so that equation (14) holds exactly. In order to construct the artificial aggregate RERs, Eurostat harmonized price weights corresponding to 2006 have been employed. Notice that these are also the weights used in the elaboration of HL_w in Tables 1 and 2, so the corresponding figures are directly comparable.

Table 3 presents the HLs associated with the original aggregate RERs (denoted as $Q_{c,t}$) as well as the above-described artificially aggregated data ($Q_{c,t}^*$). Long autoregressive models have been fitted to the data and the order of the autoregression has been chosen according to a general-to-specific criterion.¹³ The first column of Table 1 reports the HLs computed with the original data while the second column displays similar values, this time computed with the artificially generated data.

HL values corresponding to $Q_{c,t}^*$ are very much in line with those obtained with sectoral data. The correlation coefficient between (non biased-corrected) HL_w in Table 1 and the figures reported in Table II is 0.93 and the mean divergence between the two measures is less than 4 months. These figures illustrate how close the results obtained with sectoral or aggregate data are. When the original aggregate RERs are analyzed, the problem described above applies since non-harmonized Eurostat weights are different across countries, are time-varying and, in addition, are unknown. This implies that the aggregate real exchange rates is a weighted average of sectoral exchange rates plus additional terms that can introduce a bias between sectoral and aggregate measures. Nevertheless, the figures reported in Table III, column 1, are still a good approximation of weighted sectoral HLs. The correlation coefficient is still very high (0.8) and, although the estimates in Table III are slightly higher than those reported in Table I, the mean difference is only 8.8 months. Furthermore, the qualitative conclusion does not change: when aggregate data is employed, HL estimates lie, in general, in the 3 to 5 year interval.

(Table III about here)

¹³The maximum number of lags was set equal to 30. Very similar results were obtained using the AIC.

5. CONCLUSIONS

This paper offers bad and good news. The bad news are that we do not yet have a convincing solution to the long-debated PPP puzzle. In particular, it is argued that aggregation does not solve the puzzle. We have shown that the divergence between IMRR's aggregate and sectoral persistence estimates is due to a downward bias affecting their sectoral estimates, rather than to an upward bias in aggregate estimates deriving from the existence of individual heterogeneity, as argued by IMRR. The source of the bias is the definition of “*average*” response function employed by these authors. This function is computed as in a model where coefficients are homogeneous and equal to the mean value of sector-specific coefficients. Clearly, by averaging the model's coefficients in the first stage, highly persistent sectors are eliminated so that, not surprisingly, lower estimates of persistence are obtained in the second stage. Nevertheless, when IRFs are computed for each sector and then averaged, standard estimates of persistence are recovered. Finally, it is also shown that very similar persistence values can be obtained when aggregate data is employed, as implied by the theoretical results in Mayoral (2007).

The good news, however, are that estimates derived from aggregate data are reliable even when the assumption of individual homogeneity is violated, which is likely to be the case in a wide variety of contexts. Thus, applied macroeconomists can still rely on aggregate data for their studies. Notwithstanding, when heterogeneity is suspected, it is important to remember that the dynamics of the aggregate process can be very complex. Hence, careful model specification is needed in this case since, otherwise, the aggregation problems highlighted in IMRR would apply.

REFERENCES

- Andrews, D. and H. Chen (1994), "Approximately median-unbiased estimation of autoregressive models," *Journal of Business and Economic Statistics* 12, 187-204.
- Berk, K.N. (1974), "Consistent Autoregressive Spectral Estimates," *The Annals of Statistics* 2, 489-502.
- Bils, M. and P. Klenow (2002), "Some Evidence on the Importance of Sticky Prices," *NBER* 9060.
- Cheung, Y.-W., M. Chin and E. Fuji (2001), "Market Structure and the Persistence of sectoral Real Exchange Rates," *International Journal of Finance and Economics* VI(II), 95-114.
- Engel, C. and S.S. Chen (2005), "Does 'Aggregation Bias' Explain the PPP Puzzle?" *Pacific Economic Review* 10, 49-72.
- Gadea, M.D., A. Montañes, and M. Reyes (2004), "The European currencies and the dollar: from the post-Bretton-Woods era to the euro," *Journal of International Money and Finance* 23, 1109-1136.
- Granger, C.W.J. (1980), "Long-memory relationships and the aggregation of dynamics models," *Journal of Econometrics* 14, 227-238.
- Hegwood, N.D. and D.H. Papell (1998), "Quasi Purchasing Power Parity. International," *Journal of Finance and Economics* 3, 279-289.
- Imbs, J, A. Mumtaz, M.O. Ravn and H. Rey (2005), "PPP Strikes Back: Aggregation and the Real Exchange Rate," *Quarterly Journal of Economics* CXX(1), 1-43.
- Imbs, J, A. Mumtaz, M.O. Ravn and H. Rey (2004), "Aggregation bias does not explain PPP puzzle" Manuscript.
- Kilian, L. (1998), "Small-Sample Confidence Intervals for Impulse Response Functions," *Review of Economics and Statistics* LXXX, 218-30.
- Koop, G., E. Ley, J. Osiewalski and M.F.J. Steel (1997), "Bayesian analysis of long memory and persistence using ARFIMA models," *Journal of Econometrics* 76, 149-169.
- Lewbel, A. (1994), "Aggregation and Simple Dynamics," *American Economic Review* 84, 905-918.
- Lopez, C., C. Murray and D.H. Papell (2003), "State of the Art Unit Root Test and the PPP Puzzle," *Journal of Money, Credit and Banking* 37, 361-369

- Lopez, C., C. Murray and D.H. Papell (2004), “More Powerful Unit Root Test and the PPP puzzle” Manuscript.
- Mayoral, L. (2007), “The law of conservation of persistence”, Manuscript.
- Michael, P., A.R. Nobay and D.D. Peel (1997), “Transactions costs and nonlinear adjustment in real exchange rates: and empirical investigation,” *Journal of Political Economy*, 105, 862-79.
- Murray, Ch. J. and D.H. Papell (2002), “The purchasing power parity persistence paradigm,” *Journal of International Economics* 56, 1-19.
- Murray, C. and D.H. Papell (2005), “The Purchasing Power Parity is Worse Than You Think,” *Empirical Economics* 30, 783-790.
- Ng, S., P. Perron (2001), “Lag length selection and the construction of unit root tests with good size and power,” *Econometrica* 69, 1519-54.
- Obstfeld, M. and K. Rogoff (2000), “The Six Major Puzzles in International Macroeconomics: Is There a Common Cause?,” *NBER Working Papers*, 7777.
- Pesaran, M. H. (2006), “Estimation and Inference in Large Heterogeneous Panels with a Multifactor Error Structure,” *Econometrica* 74, 967-1012.
- Pesaran, M. H. and R. Smith (1995), “Estimating Long-Run Relationships From Dynamic Heterogeneous Panels,” *Journal of Econometrics* LXVIII, 79-113.
- Robinson, P.M. (1978), “Statistical inference for a random coefficient autoregressive model,” *Scandinavian Journal of Statistics* 5, 163-168.
- Rogoff, K. (1996), “The Purchasing Power Parity Puzzle,” *Journal of Economic Literature* XXXIV, 647-68.
- Rossi, B. (2005), “Confidence Intervals for Half-Life Deviations From Purchasing Power Parity,” *Journal of Business & Economics Statistics* 33, 432-442.
- Stoker, T.M. (1984). “Completeness, Distribution Restrictions, and the Form of Aggregate Functions,” *Econometrica* 52, 887-907.
- Taylor, M.P., D.A. Peel and L. Sarno (2001), “Non-linear mean-reversion in real exchange rates: toward a solution to the purchasing power parity puzzles?,” *International Economic Review* 42, 1015- 1042.
- Yang, J. (1997), “Exchange Rate Pass-Through into U.S. Manufacturing Industries,” *Review of Economics and Statistics* LXXIX, 95-104.
- Zaffaroni, P. (2004), “Contemporaneous aggregation of linear dynamic models in large economies,”

Journal of Econometrics 120, 75-102.

APPENDIX 1

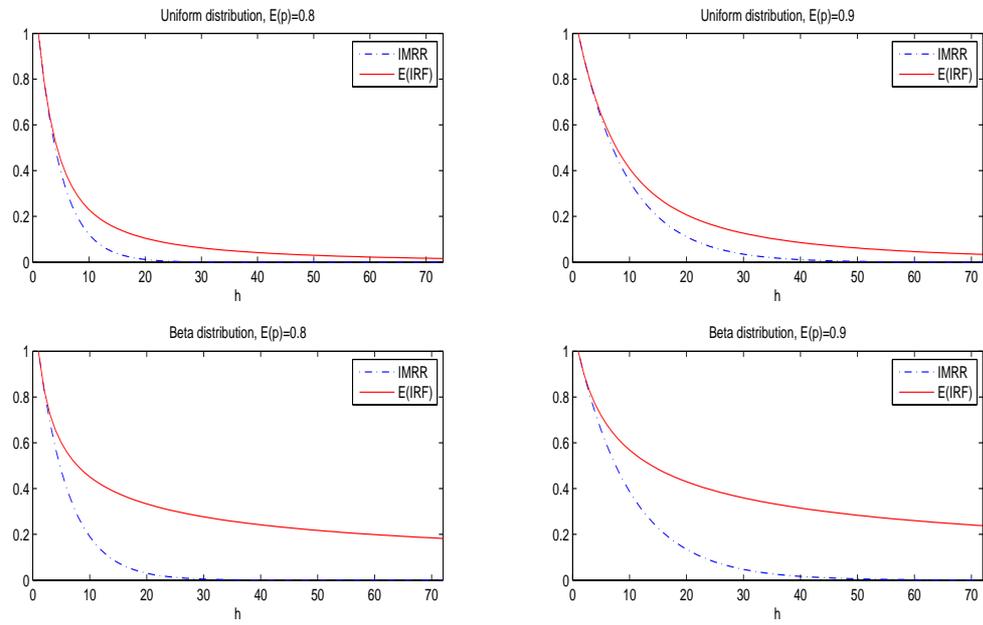


FIG I. IRFs of several processes with heterogeneous dynamics

TABLE I

HALF LIVES WITH DISAGGREGATE DATA

Panel	Mean Group (MG)			MG, Bias corrected		
	HL_{IMRR}	HL_m	HL_w	HL_{IMRR}	HL_m	HL_w
	26.40 (20.98,31.28)	36.39 (26.85,38.09)	37.42 (26.76,39.90)	43.04 (19.20,68.46)	>180 (51.84, ∞)	144.84 (53.60, ∞)
BE	27.96 (22.66,33.40)	37.68 (32.57,39.47)	36.81 (31.34,38.77)	59.68 (40.45,75.32)	>180 (78.36, ∞)	>180 (53.60, ∞)
DE	26.23 (20.56,31.55)	31.42 (20.09,33.76)	39.53 (19.54,52.68)	51.37 (19.15,63.81)	>180 (30.39, ∞)	>180 (28.87, ∞)
DK	33.10 (25.06,44.14)	45.43 (36.78,49.53)	43.90 (33.41,48.33)	87.65 (27.84,119.68)	>180 (74.72, ∞)	>180 (49.70, ∞)
ES	37.16 (28.78,45.98)	51.57 (44.41,55.35)	44.31 (35.51,49.21)	100.80 (32.92,138.34)	>180 (155.36, ∞)	>180 (88.58, ∞)
IT	29.78 (23.02,34.71)	37.36 (31.77,38.64)	35.63 (26.14,37.17)	96.71 (25.62,143.98)	>180 (98.41, ∞)	>180 (28.47, ∞)
FR	27.32 (22.48,31.46)	36.49 (27.95,39.55)	38.46 (27.88,41.97)	62.92 (24.79,81.89)	>180 (95.62, ∞)	>180 (61.93, ∞)
GR	26.38 (20.66,31.01)	32.73 (27.07,34.01)	28.01 (22.97,33.91)	47.17 (20.87,60.12)	178.53 (37.62, ∞)	99.12 (27.52, ∞)
NL	23.05 (18.90,26.94)	26.93 (19.94,27.37)	27.29 (19.79,28.47)	37.54 (20.54,48.04)	104.69 (31.24, ∞)	44.55 (26.34,52.57)
PT	32.79 (25.87,42.14)	57.60 (46.07,62.92)	59.07 (44.40,66.30)	110.19 (26.18,152.03)	>180 (120.29, ∞)	>180 (1,110.87, ∞)
FI	20.55 (17.28,24.23)	20.94 (17.93,21.89)	20.85 (17.80,22.28)	30.92 (19.19,39.32)	31.87 (25.52,54.85)	30.96 (25.11,49.58)
UK	17.74 (14.59,19.95)	21.53 (14.20,21.38)	21.07 (14.82,22.17)	33.12 (16.52,39.81)	42.10 (26.55, ∞)	38.03 (21.91, ∞)

TABLE II

HALF LIVES WITH DISAGGREGATE DATA

Panel	MG-SURE, Bias corrected			MG-CCE, Bias corrected		
	HL_{IMRR}	HL_m	HL_w	HL_{IMRR}	HL_m	HL_w
	42.87 (17.54,76.83)	45.57 (19.69,62.13)	49.70 (20.25,78.43)	20.09 (10.75,31.40)	>180 (50.93,>180)	>180 (48.53,>180)
BE	34.85 (27.23,42.49)	37.52 (29.75,47.32)	40.07 (29.93,51.76)	21.75 (16.57,26.27)	>180 (62.14,>180)	29.05 (21.38,91.05)
DE	40.76 (31.65,48.31)	47.58 (34.97,60.79)	46.35 (30.45,58.47)	27.83 (15.87,42.08)	81.59 (31.72,>180)	147.79 (25.20,>180)
DK	45.28 (33.14,55.21)	54.29 (39.78,69.16)	60.02 (39.79,79.30)	25.18 (14.76,33.37)	35.78 (15.42,49.98)	39.90 (15.90,71.80)
ES	35.46 (26.08,44.76)	38.61 (30.78,47.53)	40.48 (30.08,52.30)	25.05 (14.81,32.08)	157.02 (26.55,>180)	64.02 (22.59,150.06)
IT	40.32 (31.19,48.81)	43.57 (32.34,52.14)	48.11 (35.32,63.60)	23.75 (15.06,34.89)	44.76 (18.05,69.48)	38.66 (16.31,64.27)
FR	30.76 (24.18,38.27)	32.68 (24.22,39.85)	35.49 (28.47,42.98)	25.31 (14.54,34.50)	177.83 (37.39,>180)	94.72 (26.98,175.40)
GR	27.40 (23.31,32.83)	30.56 (24.28,37.66)	33.54 (25.87,42.58)	32.16 (44.44,60.65)	166.74 (28.22,>180)	179.62 (30.19,>180)
NL	33.48 (26.79,40.97)	35.55 (28.06,44.87)	38.77 (29.56,49.19)	20.89 (10.31,32.25)	32.71 (13.64,56.73)	39.95 (18.47,63.39)
PT	31.17 (24.58,37.60)	38.47 (28.89,49.03)	38.80 (29.75,45.91)	20.84 (14.23,27.30)	>180 (33.83,>180)	>180 (33.84,>180)
FI	30.93 (25.12,36.23)	32.04 (24.50,38.10)	33.38 (24.73,42.28)	13.85 (9.66,20.14)	24.41 (12.74,37.77)	24.05 (12.32,39.52)
UK	36.87 (28.74,46.85)	39.50 (29.35,49.07)	41.49 (30.35,50.07)	19.20 (12.36,25.88)	104.30 (26.59,170.95)	28.08 (14.01,42.48)

TABLE III

HALF LIVES WITH AGGREGATED DATABASE

	$Q_{c,t}$	$Q_{c,t}^*$
Panel	43.86 (37.61,47.43)	36.85 (29.72,38.41)
BE	45.20 (33.04,55.90)	40.28 (18.74,50.03)
DE	45.97 (16.13,72.42)	40.04 (16.22,87.98)
DK	49.12 (19.64,58.46)	42.20 (17.05,48.82)
ES	50.82 (41.36,56.61)	42.25 (39.22,64.11)
IT	37.44 (29.23,42.03)	38.19 (28.18,40.51)
FR	40.16 (18.26,49.86)	39.86 (17.38,50.60)
GR	57.37 (39.54,67.38)	39.71 (16.49,49.88)
NL	39.50 (16.07,50.76)	29.29 (17.04,37.57)
PT	73.60 (52.82,87.16)	70.54 (42.05,87.51)
FI	23.36 (16.46,24.41)	21.18 (16.00,23.94)
UK	31.58 (8.27,44.44)	17.82 (6.36,34.32)