

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

BY M. DOLORES GADEA<sup>a</sup> AND LAURA MAYORAL<sup>b</sup>

<sup>a</sup>University of Zaragoza; <sup>b</sup>Institute for Economic Analysis, (CSIC)

Accepted in Journal of Applied Econometrics

## Abstract

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 them, the source of the bias is the existence of heterogeneous price adjustment dynamics at the sectoral level that established time series or panel data methods fail to control for.

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), which prove that aggregate and micro models induce the same shock persistence behavior, we show that estimates based on aggregate and sectoral exchange rates are, in fact, highly consistent. Thus, aggregation is not the solution to the PPP puzzle.

JEL classification: F31, C22, C23.

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

---

\*Correspondence to: Laura Mayoral. Institute for Economic Analysis, XXXX, Barcelona. Tel.: +3493XXXXXX, Fax: +3493XXXXXX. E-mail: 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 noticed that the estimated half-lives (HLs) of real exchange rate adjustment obtained in studies based on panel and long-span data 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 high short-term exchange rate volatility.

The literature documenting 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 downwardly biased. When the bias is corrected, it is generally found that HL point estimates are well above the "consensus view", implying that the size of the puzzle is even larger than was originally believed (see Murray and Papell, 2002, 2005, Lopez et al., 2003, 2008).

In the opposite direction, there have been several attempts to solve the puzzle, most of them departing from linearity (such as nonlinear dynamics in real exchange rate adjustment or the existence of structural breaks)<sup>1</sup> but also, in a linear setting, highlighting 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

---

<sup>1</sup>By 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), Taylor et al. (2001) and Sarno et al. (2004). 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.

are upwardly biased. According to them, the existence of heterogeneous dynamics at the sectoral level (which is neither taken into account explicitly nor handled in an appropriate manner in most 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”*.<sup>2</sup>

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 (sectoral) HL, which, in turn, is computed from a ‘sectoral’ impulse response function. The source of the bias is precisely the definition of the sectoral impulse response function used by these authors. Instead of computing the individual impulse responses and averaging them in order to produce an estimate of the average sectoral 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’ impulse response function, as if the model was one of homogeneous coefficients given by the mean value of the heterogeneous AR coefficients. Since the impulse response function (IRF) is a highly nonlinear function, averaging the IRFs may yield very different results to averaging the AR coefficients and then computing the IRF. 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 the model coefficient estimates are

---

<sup>2</sup>This 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.

averaged, 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 those employed in their paper, we have quantified the size of the bias that affects IMRR's measures of persistence. It turns out that the bias is substantial and that, once it is corrected, sectoral persistence estimates increase considerably. Moreover, the classical result of 3-5 year half-life of PPP deviations is recovered and even larger estimates are obtained when small sample bias correcting techniques are employed. It is important to emphasize that 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), which 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 process 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 estimates of persistence, computed with either IMRR's aggregate or with sectoral data, which are highly consistent.

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

Hence, the bad news is that aggregation is not a convincing solution to the PPP puzzle. The good news, however, is that applied macroeconomists can rely on aggregate data for evaluating the persistence of aggregate shocks in the presence of individual heterogeneity, since, under the usual assumptions of correct specification, standard techniques should deliver micro and macro estimates that are very much alike.

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 that affects 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 consistent 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 that affects IMRR’s sectoral estimates. Secondly, we consider the question of whether the persistence of aggregate shocks, as measured by the IRF, changes when considered at different aggregation levels. It is shown that the standard IRF at the aggregate level is simply the expected value of the sectoral impulse responses. This implies a tight link between the aggregate and the sectoral shock response, since the former is just the average of the individual shock responses. It follows that the aggregate process is persistent if the sectors are themselves, on average, persistent, (and not because they are heterogeneous) and that, under the usual assumptions of correct specification, micro and macro estimates of shock response should be very similar. In other words, there is no *aggregation bias* that systematically increases persistence when more aggregated data is employed.

### 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.<sup>3</sup> In its simplest version, they assume that for each country  $c$ , each

---

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

sector in the panel can be represented as (see equation (1) in IMRR),

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

where  $i$ ,  $c$  and  $t$  denote sector, country and period, respectively,  $q_{c,i,t}$  is the real exchange rate for country  $c$ , sector  $i$  at time  $t$ ,  $\gamma_{c,i} = \bar{\gamma} + \eta_{c,i}^\gamma$ ,  $\rho_{c,i} = \bar{\rho} + \eta_{c,i}^\rho$ ,  $\bar{\gamma}$  and  $\bar{\rho}$  are constants, and  $\rho_{c,i}$  has support on the interval  $(-1, 1]$ . We further assume that  $E_s(\rho_c^h)$  exists for all  $h$ , where  $E_s(\cdot)$  denotes the expectation across the distribution of sectors of country  $c$ , and that the innovation  $\nu_{c,i,t} = u_{c,t} + \varepsilon_{c,i,t}$  is the sum of two orthogonal, zero-mean martingale difference sequences, one common to all sectors and one idiosyncratic, with variances  $\sigma_{u_c}^2 > 0$  and  $\sigma_{\varepsilon_{c,i}}^2$ , respectively. Finally, it is assumed that  $\eta_{c,i}^\gamma$  and  $\eta_{c,i}^\rho$  are *i.i.d* zero-mean random variables, mutually independent of  $\nu_{c,i,t}$ .

As argued by IMRR, impediments to arbitrage or nominal rigidities vary considerably across goods. Since these impediments are usually believed to be behind cross-country price differences, they could bring about important heterogeneity in the speeds of reversion to parity across sectors and countries.<sup>4</sup> Model (1) can account for different sources of heterogeneity: in addition to country and sector fixed effects (captured by the parameter  $\gamma_{c,i}$ ), it also allows for different speeds of shock adjustment by letting  $\rho_{c,i}$  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 (IRF), 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 sector  $i$  of country  $c$ , the response to a unitary aggregate shock occurring at time  $t$ ,  $h$  periods ahead, can be computed as the difference between two forecasts (see Koop et al., 1997),

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

where the operator  $E(\cdot | \cdot)$  denotes the best mean squared error predictor and  $z_{i,t-1} = \left( q_{c,i,t-1}, q_{c,i,t-2} \dots \right)'$ ; Applied to the simple model in (1), it yields that the

---

time  $t$  while  $P_{US,i,t}$  is the corresponding U.S. price.

<sup>4</sup>See Cheung et al. (2001) and Bills and Klenow (2002) for some evidence.

response of sector  $i$  in country  $c$  to a unitary aggregate shock in  $t$ ,  $h$  periods ahead is

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

If one is interested in the average response across sectors to this 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 in country  $c$  to a unitary shock  $h$  periods ahead, denoted as  $IRF_{micro}^c$ , is then given by

$$IRF_{micro}^c(t, h) = E_s(IRF^c(t, h)) = E_s(\rho_c^h), \text{ for } h \geq 0, \quad (4)$$

Then, the expected IRF associated with (1) is given by the  $h^{th}$  – *moment* of the distribution of  $\rho$ . 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 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 for country  $c$  as the value of  $h$  that verifies

$$E_s(IRF_{micro}^c(t, h = HL_{micro}^c)) = 0.5, \quad (5)$$

whereas the CIR is

$$CIR_{micro}^c = \sum_{h=0}^{\infty} IRF_{micro}^c(t, h). \quad (6)$$

Let us now revise how the calculations reported in IMRR relate to the measures defined in (4), (5) and (6). They assume that there is not country heterogeneity, and therefore  $\rho_{c_1,i} = \rho_{c_2,i} = \rho_i$  for all  $c_1, c_2$ . Their approach is to estimate the expected value of  $\rho_i$ ,  $\bar{\rho}$ , in a second step, to compute the IRF defined in (4) as if the true DGP was given by  $q_{i,t} = \bar{\gamma} + \bar{\rho}q_{i,t-1} + a_{i,t}$ , 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,

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

They estimate  $\rho$  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  $\rho$  is positive, which is a very realistic assumption in this case,  $\rho^h$  is strictly convex and application of Jensen’s inequality yields

$$IRF_{micro} = E_s(IRF(t, h)) > \overline{IRF}(t, h), \text{ for all } h > 1, \quad (8)$$

or, in other words, (7) systematically underestimates the average shock response.<sup>5</sup> 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 of the first autoregressive coefficient 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  $\alpha_{1i}$  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  $\alpha_1$  is positive, the IRF is a convex function and Jensen’s inequality ensures the result above.

To illustrate the inequality in (8), we have fitted AR(1) models to the sectoral exchange rate data employed by IMRR (see Section 3 below for a description of this data set) and

---

<sup>5</sup> Although, 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.

we have used the OLS estimated coefficients to compute (7) and (4), the IRF's according to IMRR's and this paper's approaches, respectively. The average of the estimated AR(1) coefficients is 0.98 and the average across the different countries of the standard deviations of these coefficients is 0.023. Figure I presents histograms of the AR(1) sectoral coefficients (left-hand side) as well as the IRFs for three countries in the sample, namely, Germany, Spain and France. The histograms show that sectoral data is highly persistent, with most sectoral coefficients concentrated in the neighborhood of 1. The mean value of the AR(1) coefficients,  $\bar{\rho}$ , is 0.975, 0.985 and 0.975 for Germany, Spain and France, respectively. However, there is enough heterogeneity in the coefficients for the inequality (8) to be important.<sup>6</sup> As Figure I illustrates, the gap between the alternative definitions of IRF is significant. The HL computed as in IMRR is less than two years and a half (29 months) for the three countries considered. When the HLs are computed using (5), estimates over three years (44, 41 and 40 months for Germany, Spain and France, respectively) have been obtained. Interestingly, these figures are similar to those presented in Section 3, where more complex models and more sophisticated techniques are applied to estimate the HLs.

**(Figure I about here)**

We have also carried out some simulations to see under what circumstances we should expect a large gap between the two alternative definitions of IRFs. We have generated 200 heterogeneous AR(1) processes of the form  $y_t = \beta_i y_{t-1} + \varepsilon_t$ , where  $\varepsilon_t \sim iN(0, 1)$  and the  $\beta_i$  have been drawn from a  $N(\mu, \sigma)$  distribution, for different values of  $\mu$ ,  $\mu \in \{0.9, 0.98\}$ , and  $\sigma$ ,  $\sigma \in \{0.02, 0.05, 0.1\}$ . To avoid explosive processes, values of  $\beta_i$  strictly greater than 1 have been replaced by 1 (an exact unit root).<sup>7</sup> Figure II presents the plots of the IRF

---

<sup>6</sup>The sample standard deviation of the AR(1) coefficients is 0.029, 0.015 and 0.026 for Germany, Spain and France, respectively.

<sup>7</sup>This implies that, after the truncation, the mean and the standard deviation of the distribution of  $\beta$  are slightly different from the values of  $\mu$  and  $\sigma$  above. More specifically, the sample mean and the sample standard deviation after the truncation are  $\mu' = \{0.901, 0.089, 0.089\}$  and  $\sigma' = \{0.02, 0.048, 0.088\}$  for the three graphs on the left-hand side of Figure 2, respectively, and  $\mu' = \{0.977, 0.969, 0.948\}$  and  $\sigma' = \{0.017, 0.035, 0.067\}$ , for the plots of the right-hand side of Figure II.

computed according to IMRR and to (4). It is seen that the more persistent and the more heterogeneous the individual processes are (that is, the larger the values of  $\mu$  and  $\sigma$ ), the more important the gap between the two IRFs is. For moderate values of  $\mu$  and  $\sigma$ , the difference between the two functions is small, as illustrated by the graph at the upper left corner of Figure II. However, the gap can be very important when either  $\mu$  or  $\sigma$  are large as the remaining plots show. The graph at the upper right corner of Figure II illustrates the difference between the alternative definitions of IRFs for values of  $\mu$  and  $\sigma$  that are close to those found in IMRR data set ( $\mu = 0.98$  and  $\sigma = 0.02$ ). For this particular case, the HL associated to IMRR's IRF is 33 periods, whereas that computed according to (4) equals 46 periods.

**(Figure II about here)**

Summarizing, the results above show that IMRR's sectoral persistence estimates 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.

## **2.2. Comparing aggregate and sectoral persistence**

The major argument of IMRR is that the existence of sectoral heterogeneity, which is not controlled for by standard time series or panel estimation techniques, introduces a bias in persistence measures computed with aggregate data such that “*the persistence of disaggregated relative prices is on average smaller than the persistence of the aggregate real exchange rate itself*”. To analyze the validity of this claim, we now turn to examine the relation between the average of the sectoral IRFs, 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.<sup>8</sup> Hence, before deriving the IRF of the aggregate model, we explicitly consider the 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 process as one given by the expected value across individuals of the disaggregate relations. The aggregate real exchange rate for country  $c$  could be obtained as

$$Q_{c,t} = E_s(\gamma_c) + E_s(\rho_c Q_{c,t-1}) + u_{c,t}, \quad (9)$$

where  $E_s(\cdot)$  denotes the expectation over the cross-sectional distribution of sectors of country  $c$ ,  $Q_{c,t} = E_s(q_{c,t})$  is the aggregate real exchange rate for country  $c$  at time  $t$  and  $u_{c,t} = E_s(\nu_{c,t})$  is the aggregate shock. Under the assumptions of Section 2.1 and assuming further that the number of micro-processes is (countably or uncountably) infinite, Lewbel (1994) showed that expression (9) is equivalent to,

$$Q_{c,t} = \sum_{s=1}^{\infty} A_s Q_{c,t-s} + u_{c,t}, \quad (10)$$

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

$$A_j = m_j - \sum_{r=1}^{j-1} m_{j-r} A_r. \quad (11)$$

where  $m_j = E(\rho_c^j)$  is the moment of order  $j$  of  $\rho_c$ . It follows that, under heterogeneity, the aggregate model might display very complicated dynamics even when the behavior of the micro units is very simple, as it is in this case.

As for the standard aggregate IRF associated with model (10), it can be computed as the difference between the forecasts

$$IRF_{macro}(t, h) = E(Q_{c,t+h} | u_{c,t} = 1; Z_{t-1}) - E(Q_{c,t+h} | u_{c,t} = 0; Z_{t-1}), \quad (12)$$

---

<sup>8</sup>See Granger (1980), Robinson (1978) and Zaffaroni, (2004).

where  $IRF_{AG}$  denotes the standard IRF computed with aggregate data and  $Z_{t-1} = (Y_{t-1}, Y_{t-2}, \dots)$ . Application of this definition to (10) yields,

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

and in general,

$$IRF_{macro}(t, h) = \sum_{j=1}^h A_j IRF_{macro}(h-j). \quad (13)$$

At first glance, there is no clear relation between the sectoral and the aggregate IRFs, defined in (4) and (13), respectively. However, the expression of the  $IRF_{macro}$  in (13) can be notably simplified. To do this, notice that (11) can be rewritten as  $m_j = \sum_{r=0}^{j-1} m_r A_{j-r}$ . Iterating this expression, one can easily check that  $IRF_{macro}(t, h) = A_1 = m_1$ ,  $IRF_{macro}(t, 2) = A_1^2 + A_2 = m_0A_2 + m_1^2 = m_2$ , and that, in general,

$$IRF_{macro}(t, h) = m_h = E_s \left( \rho_c^h \right).$$

That is, the aggregate IRF equals the non-centered moments of the distribution of the AR coefficients. Notice that this is precisely the value of the average of the sectoral IRFs, as shown in (4). This result also holds for more general micro  $AR$  dynamics and under less stringent assumptions than those considered here, as shown in Mayoral (2007).

Several considerations are worth emphasizing at this point. Firstly, the result above shows that the effect over time of aggregate shocks is the same, regardless of whether it is considered at the sectoral or at the aggregate level. Since the population values across aggregation levels are equal, under the usual assumptions of correct specification, consistent estimators applied to either type of data will provide similar estimates of shock persistence irrespective of the aggregation level, at least for sufficiently large sample sizes. This implies that there is no aggregation bias that systematically increases persistence estimates based on aggregate data.

Secondly, notice that the aggregate process  $Q_{c,t}$  might not admit a representation with a finite number of parameters, as (10) shows. However, even in these situations, it is still

possible to obtain consistent estimates of the autoregressive parameters of the aggregate model. As shown by Berk (1974), consistency can be achieved provided a sufficiently long AR( $k$ ) structure is specified, where  $k$  grows at an appropriate rate with respect to the sample size. Some simulations illustrating this point are provided in Section 4.

Thirdly, IMRR's analytical calculations to show that aggregate time series data overestimate persistence start by postulating the same model for the sectoral units and for the aggregate data, namely, an AR(1) model. Then, they consider whether the estimate of the autoregressive coefficient of the aggregate model is a consistent estimator of  $E_s(\rho_c)$ , the average of the sectoral AR(1) coefficients. But, as is clear from (10), this aggregate model is misspecified and so, not surprisingly, their estimates are biased. Thus, the source of the bias discussed by IMRR is due to the misspecification of the aggregate model rather than to the aggregation of heterogeneous processes.

### 3. RESULTS FOR SECTORAL DATA

This section quantifies the magnitude of the negative bias that affects IMRR's measures of sectoral shock persistence. We employ the same data set as in their paper, that is, nineteen monthly sectoral exchange rates of 11 European countries (Belgium, Denmark, Germany, Spain, Italy, France, Greece, Netherlands, Portugal, Finland and U.K.) 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

$$q_{c,i,t} = \gamma_{c,i} + \sum_{k=1}^K \rho_{c,i,k} q_{c,i,t-k} + \nu_{c,i,t}, \quad (14)$$

where they assume that sectors are homogeneous across countries, so that  $\gamma_{c,i} = \gamma_i$ ,  $\rho_{c,i,k} = \rho_{i,k}$  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 recompute the MG, MG-SURE and MG-CCE, correcting for the OLS small-sample bias using Kilian’s (1998) bootstrap-after-bootstrap method. Then they 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. 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 comparison, IMRR’s notation for the different estimation approaches has been preserved and HLRs computed according to their procedure are also reported.

Table I contains the HLRs 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, it is the HL associated with the sectoral IRF computed by averaging the AR coefficients.  $HL_m$  and  $HL_w$  are the HLRs associated with a simple arithmetic average and with a weighted average of the individual IRFs, respectively. Ideally, in order to facilitate comparison with the HLRs 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. Our preferred measure of sectoral persistence is  $HL_w$ , since it weights sector-specific responses in a similar manner to which the aggregate function weights 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, where  $p$  was chosen according to a general-to-specific criterion with a maximum number of lags of 20.<sup>9</sup> Confidence intervals have been calculated using bootstrap techniques.

In order to compute bias-corrected estimates, Kilian’s (1998) bootstrap-after-bootstrap method has been employed. IMRR consider two alternative implementations of this technique: the “*indirect*” approach, which consists of first correcting the bias of the autoregressive coefficients and then computing the HL, and the “*direct*” approach, which directly corrects the downward bias of the HL. They study, by simulation, which technique behaves best and conclude that the direct approach provides a better fit in their case. Analogously, we have also conducted a similar Monte Carlo exercise to determine which method performs best for our definition of IRF. The Appendix presents details of the computation of Kilian’s bootstrap-after-bootstrap algorithm as well as the results of our simulation study. It turns out that the direct approach tends to underestimate the true HL substantially whereas the indirect one performs reasonably well. Thus, the indirect bias-correcting approach has been employed to perform the bias-corrected estimates in Tables I and II.

**(Table I about here)**

Several conclusions can be drawn from Table I. 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 that affects 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),

---

<sup>9</sup>The AIC was also employed and very similar results were obtained.

whereas  $HL_m$  and  $HL_w$  are, in general, above this 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 the  $HL_{IMRR}$  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.<sup>10</sup> 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.

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 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 (CCE, Pesaran, 2006), based on the regressions

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

where  $\bar{q}$  is the cross-sectional average of  $q_{c,i}$ .

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.<sup>11</sup>

<sup>10</sup>Similar 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).

<sup>11</sup>When 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

**(Table II about here)**

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 I, 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) leads to 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.

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.

---

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.

#### 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 into persistence estimates computed with aggregate data and that, in fact, sectoral and aggregate estimates are very much alike.

As explained in Section 2.2., the fact that the aggregate process contains an infinite number of parameters is not an obstacle for obtaining consistent estimates since, as shown by Berk (1974), fitting a long autoregression can be sufficient to achieve consistency. To illustrate this argument, a Monte Carlo simulation, showing that standard estimation techniques yield similar impulse response trajectories when computed with macro or micro data, has been carried out. We have generated 200 AR(1) processes of the form  $y_{it} = \beta_i y_{it-1} + (u_t + \varepsilon_{it})$ , where  $u_t \sim N(0, 1)$ ,  $\varepsilon_{it} \sim N(0, 1)$  and  $\beta_i$  has been generated as a  $N(\mu, \sigma)$  distribution, for different values of  $\mu$  ( $=\{0.5, 0.7, 0.9, 0.95\}$ ) and  $\sigma$  ( $=\{0.02, 0.05\}$ ) and values of  $\beta_i$  greater than 1 have been replaced by 1 (an exact unit root), to avoid explosive processes. Aggregate data has been generated as the simple average of the individual processes, i.e.,  $Y_t = N^{-1} \sum_{i=1}^N y_{it}$  and AR(p) models have been fitted to the micro and to the aggregate data, where  $p$  has been chosen according to the AIC. The  $IRF_{micro}$  and the  $IRF_{macro}$  have been computed as in (4) and (13), respectively. Figure III presents the average trajectories over the number of replications (300) of the micro and macro impulse responses. The sample size used in this experiment is  $T=200$ , to match this paper's data set.

For all values of  $\mu$  and  $\sigma$ , the estimated micro and macro IRFs are fairly close. It is noteworthy that when persistence increases (bottom right-hand corner), both functions underestimate the true IRF, due to the well-known downward small-sample bias of OLS estimates. However, the size of the bias is very similar with either type of data and, even in these cases, micro and macro estimates are very much alike.

**(Figure III about here)**

We turn next to analyze Eurostat's aggregate real exchange rates dataset. Assuming that

aggregate prices in country  $c$  are constructed as the geometric average of sector-specific prices, that is

$$P_{c,t} = \prod_{i=1}^N P_{c,i,t}^{\omega_{i,c}}$$

where  $\omega_{c,i}$  are weights that verify  $\sum_{i=1}^N \omega_{c,i} = 1$ ,  $\omega_{c,i} = \omega_{US,i}$  for all  $i$  and where  $\omega_{US,i}$  are U.S. price weights and are not time-varying, then the bilateral aggregate real exchange rate,  $Q_{c,t} = \log(S_{c,t}P_{c,t}/P_{US,t})$ , can be written as a weighted sum of sectoral RERs, i.e.,

$$Q_{c,t} = \sum_{i=1}^N \omega_{c,i} q_{c,i,t}. \quad (15)$$

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  $Q_{c,t}$  and the weighted average of goods-specific impulse responses should be close.

In reality, however, weights are not equal across countries. This implies that  $Q_{c,t}$  is equal to

$$Q_{c,t} = \sum_{i=1}^N \omega_{c,i} q_{c,i,t} + \sum_{i=1}^N (\omega_{US,i} - \omega_{c,i}) q_{c,i,t}, \quad (16)$$

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 (16).

The ‘‘aggregation bias’’ argument states that, even in the situation described in (15), 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  $q_{c,i,t}$  presents 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 over time and across countries, we have constructed an artificial aggregate variable, computed as the weighted sum of sectoral prices, so that equation (15) holds exactly. In order to construct this variable, 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 I and II.

Table III 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.<sup>12</sup> The second column of Table III reports the HLs computed with the original data while the third column displays similar values, this time computed with  $Q_{c,t}^*$ .

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 I and the figures reported in Table III 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. The HL values reported in Table III, column 1, corresponding to the analysis of the original aggregate data set, are also a good approximation to the 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. Moreover, the qualitative conclusion does not change: when aggregate data is employed, HL estimates lie, in general, in the 3 to 5 year interval.

Several reasons can account for this small divergence between sectoral and aggregate estimates. As mentioned above, since weights vary across countries and over time, aggregate real exchange rates are a weighted average of sectoral exchange rates plus additional terms, as shown in (16). These terms can introduce some discrepancies between aggregate and sectoral estimates. More importantly, sectoral estimates are usually believed to be more affected by measurement error than aggregate ones. Therefore, they may suffer from more severe biases than aggregate estimates.

**(Table III about here)**

---

<sup>12</sup>The maximum number of lags was set equal to 30. Very similar results were obtained using the AIC.

## 5. CONCLUSIONS

This paper offers good and bad news. The bad news is that the aggregation bias argument does not seem to be a convincing solution to the long-debated PPP 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 has also been 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, is 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. Significant differences between sectoral and aggregate persistence estimates should not be interpreted as different micro-macro predictions but as the sign of misspecification of the sectoral or the aggregate model. 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 appear.

## REFERENCES

- Andrews, D., Chen, H. 1994. Approximately median-unbiased estimation of autoregressive models. *Journal of Business and Economic Statistics* **12**: 187-204.
- Berk, KN. 1974. Consistent Autoregressive Spectral Estimates. *The Annals of Statistics* **2**: 489-502.
- Bils, M., Klenow, P. 2002. Some Evidence on the Importance of Sticky Prices. NBER 9060.
- Cheung, YW., Chin, M., Fuji, E. 2001. Market Structure and the Persistence of sectoral Real Exchange Rates. *International Journal of Finance and Economics* **VI(II)**: 95-114.
- Engel, C., Chen, SS. 2005. Does 'Aggregation Bias' Explain the PPP Puzzle? *Pacific Economic Review* **10**: 49-72.
- Gadea, MD., Montañes, A., Reyes, M. 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, CWJ. 1980. Long-memory relationships and the aggregation of dynamics models, *Journal of Econometrics* **14**: 227-238.
- Hegwood, ND., Papell, DH. 1998. Quasi Purchasing Power Parity. *International Journal of Finance and Economics* **3**: 279-289.
- Imbs, JA., Mumtaz, MO., Ravn, Rey, H. 2005. PPP Strikes Back: Aggregation and the Real Exchange Rate, *Quarterly Journal of Economics* **CXX(1)**: 1-43.
- Imbs, JA., Mumtaz, MO., Ravn, Rey, H. 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., Ley, E. Osiewalski, J., Steel, MFJ. 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., Murray, C., Papell, DH. 2003. State of the Art Unit Root Test and the PPP Puzzle, *Journal of Money, Credit and Banking* **37**: 361-369.

Lopez, C., Murray, C., Papell, DH. 2008. Median Unbiased Estimation in DF-GLS regressions and the PPP puzzle. Manuscript.

Mayoral, L. 2007. The law of conservation of persistence. Manuscript.

Michael, P., Nobay, AR. Peel, DD. 1997. Transactions costs and nonlinear adjustment in real exchange rates: and empirical investigation, *Journal of Political Economy* **105**: 862-79.

Murray, C., Papell, DH. 2002. The purchasing power parity persistence paradigm, *Journal of International Economics* **56**: 1-19.

Murray, C., Papell, DH. 2005. The Purchasing Power Parity is Worse Than You Think, *Empirical Economics* **30**: 783-790.

Obstfeld, M., Rogoff, K. 2000. The Six Major Puzzles in International Macroeconomics: Is There a Common Cause? NBER Working Papers, 7777.

Pesaran, MH. 2006. Estimation and Inference in Large Heterogeneous Panels with a Multifactor Error Structure, *Econometrica* **74**: 967-1012.

Pesaran, MH., Smith, R. 1995. Estimating Long-Run Relationships From Dynamic Heterogeneous Panels, *Journal of Econometrics* **LXVIII**: 79-113.

Robinson, PM. 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.

Sarno, L., Taylor, MP. Chowdhury, I. 2004. Nonlinear dynamics in deviations from the law of one price: abroad empirical studies, *Journal of International Money and Finance* **23**: 1-25.

Stoker, TM. 1984. Completeness, Distribution Restrictions, and the Form of Aggregate Functions, *Econometrica* **52**: 887-907.

Taylor, MP., Peel, DA., Sarno, L. 2001. Non-linear mean-reversion in real exchange rates:

toward a solution to the purchasing power parity puzzles?, *International Economic Review* **42**: 1015- 1042.

Zaffaroni, P. 2004. Contemporaneous aggregation of linear dynamic models in large economies, *Journal of Econometrics* **120**: 75-102.

## APPENDIX

This appendix summarizes our implementation of Kilian’s (1998) bias correction procedure and compares, via Monte Carlo simulation, how the *direct* and the *indirect* corrections perform when applied to compute our definition of sectoral HLs.

### Kilian’s Bias correction method

As mentioned in the text, the *direct* approach performs the bootstrap correction directly on the HL, while the *indirect* method corrects the bias of the autoregressive coefficients that are employed to estimate the IRF and the HL. We summarize below the steps that have been followed to compute the direct and indirect bias corrections.

#### **Indirect method.**—

Step 1: For each sector, estimate the AR(p) model coefficients.

Step 2: Obtain bias-corrected estimates of the autoregressive parameters by using the bootstrap-after-bootstrap method (see Kilian, 1998).

Step 3: For each sector, compute the impulse response function using the corrected estimates. Average these functions to obtain an estimate of the sectoral IRF.

Step 4: Calculate the bias-corrected HL by applying formula (5) to the sectoral IRF obtained in step 3.

#### **Direct method.**—

Step 1: For each sector, estimate the AR(p) model coefficients.

Step 2: Calculate the associated IRFs and average these functions to obtain an estimate of  $IRF_{micro}$ . Compute an estimate of the HL associated with  $IRF_{micro}$ , denoted by  $\widehat{HL}$ .

Step 3: Generate  $r=1000$  bootstrap samples of the innovations using non-parametric bootstrap techniques and use the parameters obtained in Step 1 to generate  $r$  artificial samples of sectoral real exchange rates.

Step 4: Repeat steps 1 and 2 using the artificially generated series. Compute the esti-

mated bootstrap-HL as the sample mean of the  $r$  HLs obtained from the bootstrap replications, that is,  $\overline{HL}_b = r^{-1} \sum_{i=1}^r HL_i^*$ , where  $\overline{HL}_b$  denotes the average bootstrap HL and  $HL_i^*$  denotes the HL of the  $i^{th}$ -bootstrap replication.

Step 5: The bias-corrected HL is given by  $\widetilde{HL} = 2\widehat{HL} - \overline{HL}_b$ .

### Monte Carlo evidence on bias correction

This section reports the results of a Monte Carlo experiment to examine how the direct and indirect approaches perform for our definition of sectoral IRF and sectoral HL. IMRR (2004) have also explored this issue (see Table 3) and, in order to obtain comparable results, we have closely followed the design of their experiment.

We have generated processes of the form

$$\begin{aligned} q_{it} &= \alpha_i + \beta_i q_{it-1} + x_t + \nu_{it}, \quad i = 1, \dots, N, \quad t = 1, \dots, T \\ x_t &= \lambda x_{t-1} + \xi_t, \\ \nu_{it} &= u_t + \varepsilon_{it} \end{aligned}$$

where  $\alpha_i \sim N(0, 1)$ ,  $\beta_i \sim U[0.93, 0.99]$ ,  $\varepsilon_{it} \sim iN(0, 1)$ ,  $\xi_t \sim N(0, 1)$ ,  $u_t \sim N(0, 1)$ ,  $T=200$  and  $N=20$ . The number of replications is 1000. The MG and the CCE methods have been employed to compute estimates of the parameters and the small-sample bias has been corrected using either the direct or the indirect method.

The results of the experiment are reported in Table A1. The second column of this table reports the true sectoral HL for the generated samples. The third and the fourth columns present the direct and indirect bias-corrected estimates obtained using the MG estimator and the fifth and sixth columns display analogous values computed with the CCE method.

The MG method only delivers consistent estimates for the case where  $\lambda = 0$ . In this case, the direct method considerably underestimates the true HL, while the indirect method performs very well. When  $\lambda > 0$ , the MG method performs poorly since it cannot control for the common correlated component,  $x_t$ . Both the direct and the indirect methods tend to overestimate the true HL in this case. The CCE method, however, is able to control for

this effect. It is noteworthy that, in all cases, the indirect method delivers values that are very close to the true HL while the direct approach is severely downwardly biased.

**TABLE A1**

DIRECT AND INDIRECT CORRECTION APPROACHES

	HL TRUE	HL MG		HL CCE	
$\lambda$		<i>direct</i>	<i>indirect</i>	<i>direct</i>	<i>indirect</i>
0	18.35	13.08	17.90	14.29	18.24
0.2	18.46	20.74	26.61	14.21	18.52
0.4	18.47	25.88	32.39	14.59	18.05
0.6	18.35	25.16	45.32	14.86	17.10

Note: The DGP is  $(1 - \beta_i L)q_{it} = \alpha_i + x_t + \nu_{it}$ ,  $(1 - \lambda L)x_t = \xi_t$ ,  $\nu_{it} = u_t + \varepsilon_{it}$ ,  $\alpha_i \sim N(0, 1)$ ,  $\beta_i \sim U[0.93, 0.99]$ ,  $\varepsilon_{it} \sim iN(0, 1)$ ,  $u_t \sim N(0, 1)$ ,  $\xi_t \sim N(0, 1)$ , T=200 and N=20.

## TABLES

### TABLE I

HALF LIVES WITH DISAGGREGATE DATA

	Mean Group (MG)			MG, Bias corrected		
	$HL_{IMRR}$	$HL_m$	$HL_w$	$HL_{IMRR}$	$HL_m$	$HL_w$
Panel	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, $\infty$ )	144.84 (53.60, $\infty$ )
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, $\infty$ )	>180 (53.60, $\infty$ )
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, $\infty$ )	>180 (28.87, $\infty$ )
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, $\infty$ )	>180 (49.70, $\infty$ )
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, $\infty$ )	>180 (88.58, $\infty$ )
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, $\infty$ )	>180 (28.47, $\infty$ )
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, $\infty$ )	>180 (61.93, $\infty$ )
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, $\infty$ )	99.12 (27.52, $\infty$ )
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, $\infty$ )	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, $\infty$ )	>180 (1,110.87, $\infty$ )
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, $\infty$ )	38.03 (21.91, $\infty$ )

**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)

## FIGURES

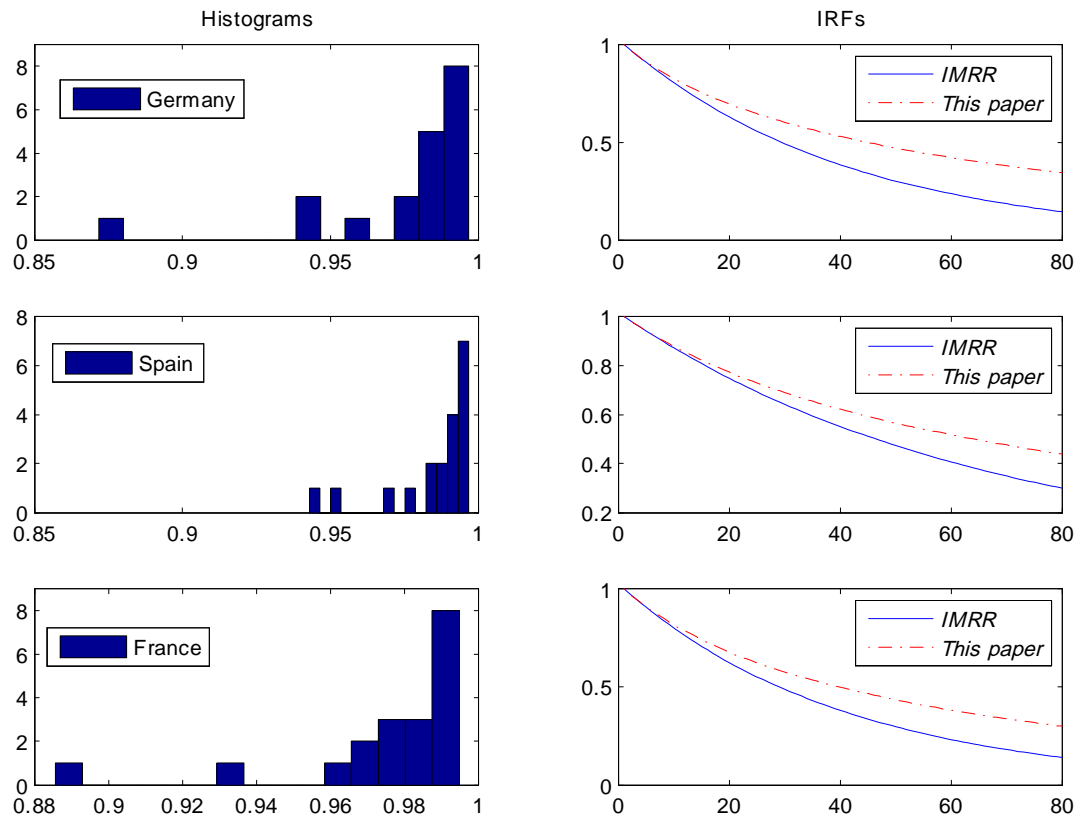


FIG. I. Alternative definitions of IRFs based on AR(1) model.

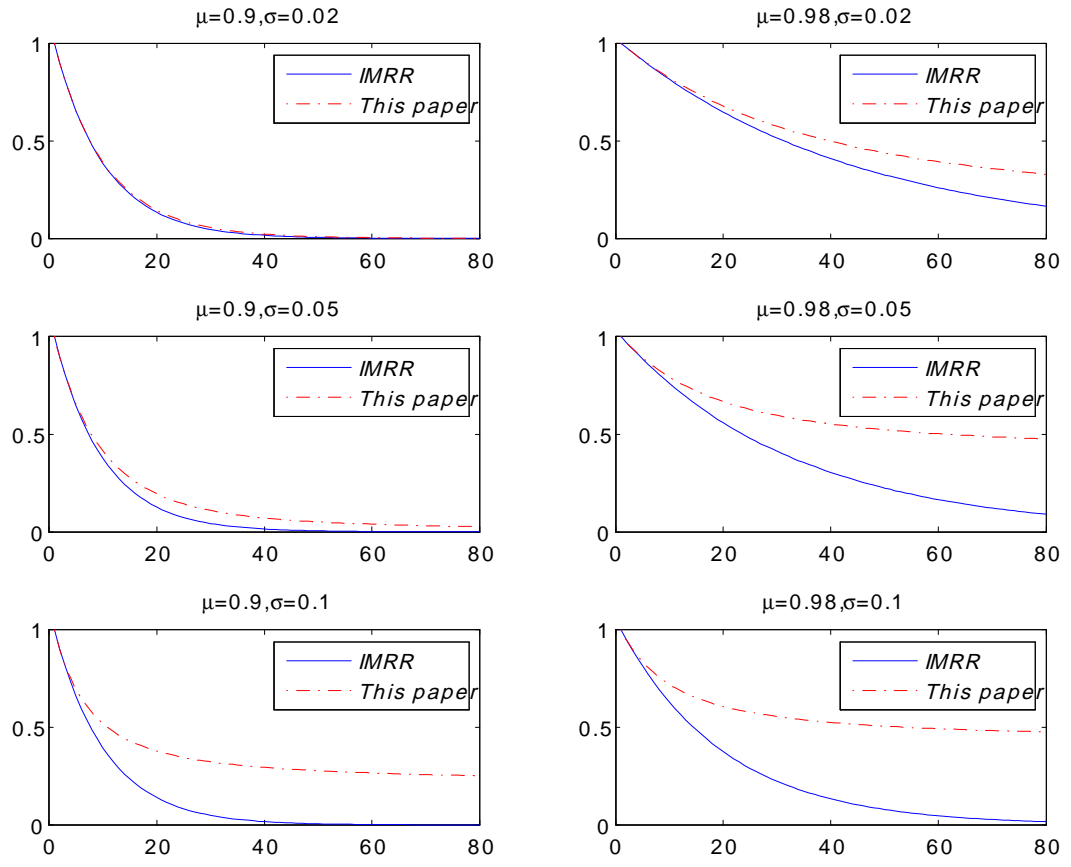


FIG. II. Simulated IRFs for different values of  $\mu$  and  $\sigma$ .

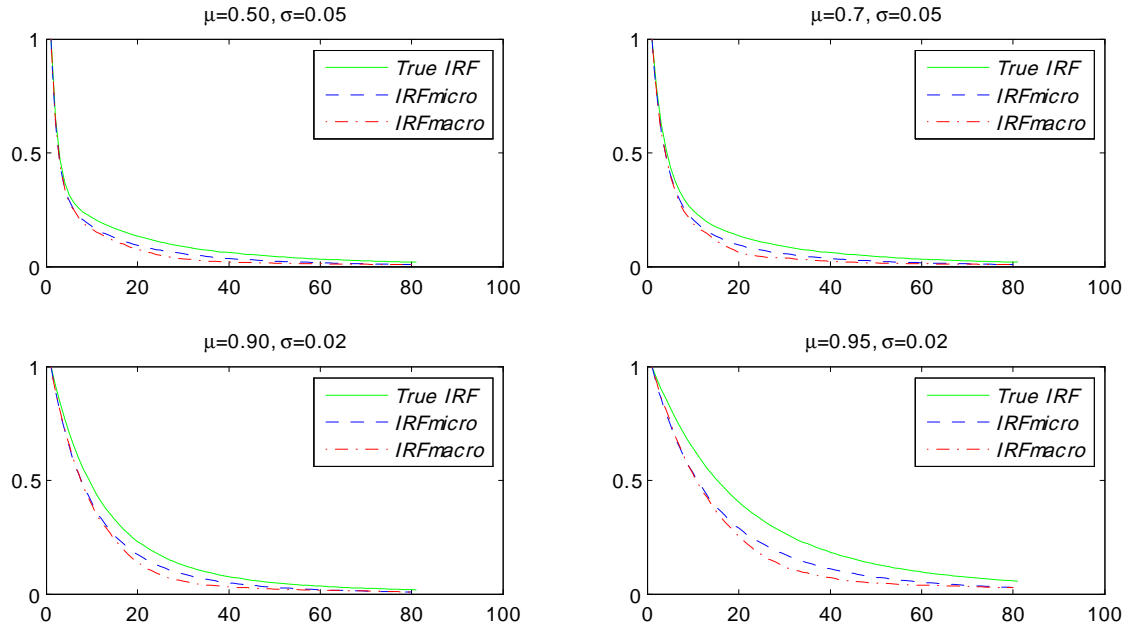


FIG. III. Simulated micro and macro IRFs.