Measuring the PPP puzzle with aggregate data in the presence of sectoral heterogeneity

By Laura Mayoral[a] and Maria Dolores Gadea*
[a]Institute for Economic Analysis, (CSIC) and
[b]Dept. of Applied Economics, University of Zaragoza

January 2008

Abstract

A decade after its formulation, the so-called Rogoff's PPP puzzle is still an open question and, in spite of the abundant literature, no agreement has been reached regarding either its theoretical explanation or the most appropriate way to measure it. Recently, the use of aggregate real exchange rate (RER) data has been criticized because it has been argued that the high degree of heterogeneity observed at the sectoral level is not controlled by established time series or panel data methods. This induces a positive bias in aggregate persistence estimates that gives rise to the puzzle.

However, aggregate data has important advantages over sectoral data, both in terms of availability and quality. The purpose of this paper to use aggregate RERs to measure the PPP puzzle but allowing explicitly for sectoral heterogeneity. To do so, the aggregate process corresponding to a standard heterogenous sectoral model has been derived. It turns out that the response to aggregate shocks is the same, regardless the level of aggregation at which it is measured. Thus, no aggregation bias arises if aggregate data and appropriate estimation techniques are employed. Next, this theory has been applied to the a data set of aggregate RERs. Finally, we have compared our results to previous findings in the literature, pointing out similarities and differences, and providing plausible explanations for the later case.

JEL classification: C22, F31, O11

Keywords: PPP puzzle; Real exchange rates; Heterogeneous dynamics; Fractional integration; aggregation bias.

*We are grateful to the participants in EEA-ESEM 2006 and the 10th *International Conference on Macroeconomic Analysis and International Finance* for their useful comments. We acknowledge financial support from the Spanish Ministry of Education through grants CYCIT SEJ2005-00215, SEC2003-04429 and SEC2003-04476.

1. INTRODUCTION

In contrast to the view prevailing in the 1980's, in recent years a general consensus about the long-run holding of the purchasing power parity (PPP) has emerged. This consensus has been made possible by enhancing the power of unit root tests, either by considering longer spans of data for a single country (Taylor (2002), Lothian and Taylor (1996), Abuaf and Jorion (1990), etc.) or panels of several contries (see Wei and Parsley, (1995), Frankel and Rose, (1996), etc.)

However, the rate at which deviations from PPP decay has been found to be extremely slow. As Rogoff (1996) has pointed out, the coincident 3-5 year half-lives (HL) of deviations found in the above-mentioned empirical studies are too lasting to be reconciled with the short-run variability exhibited by real exchange rates (RER), necessarily provoked by monetary factors. Furthermore, it has recently been argued that the reported 3-5 year half-lives are underestimated, implying that the puzzle is even bigger than was first thought. Murray and Papell (2002) have shown that traditional measures underestimate persistence because they rely on estimates of the AR parameter, which is known to be downwardly biased in finite samples if the data is highly persistent. Extending their analysis, López et al. (2003, 2004) have reported half-life estimates between 8 and 11 years, with lower bounds over 3 years and upper bounds that do not allow the rejection of the failure of the PPP.[1]

The conclusions from the papers above have been drawn from the analysis of aggregate exchange rate data of different countries. However, the use of this type of data to compute persistence estimates has been recently criticized. It has been argued that when dealing with aggregate data, the papers above implicitly assume that sectoral exchange rates have homogeneous dynamics. Nevertheless, there is substantial evidence on the existence of a high degree of sectoral heterogeneity (see for instance Engel (2002), Cheung et al. (2001), Campa and Goldberg (2002), Crucini et al. (2005), etc.). Building on the results in Pesaran and Smith (1995), Imbs et al. (2005) have argued that established time series and panel methods applied to aggregate data fail to control for this heterogeneity. This induces a positive bias in persistence estimates computed with aggregate data (what they call the "aggregation bias"), which, in turn, gives rise to the above-described puzzle. Using Eurostat sectoral data, Imbs et al. (2005) report estimates of persistence that reflect much shorter half-lives (around 1 year) and, since these values

[1] Similar conclusions have been reached by Cashin and Mcdermott (2003), Rossi (2003), Caporale et al. (2004) and Murray and Papell (2005).

are highly compatible with models based on nominal rigidities, they claimed to have found the solution to the PPP puzzle. Similar results are also reported by Crucini and Shintani (2006) using a similar methodology on a different disaggregate data set.

On the other hand, it has recently been shown that there exists a tight link between aggregate and sectoral persistence even when sectors present heterogeneous dynamics. Mayoral (2007) has shown that if sectors can be represented as heterogeneous AR processes, the standard impulse response function (IRF) computed at the aggregate level equals the average of sectoral impulse responses, provided the number of sectors is sufficiently large. This implies that the (population) aggregate response of a shock could be estimated either using disaggregate or aggregate data and, provided estimates are consistent, both set of results should be highly compatible.

Dealing with aggregate as opposed to sectoral time series data presents important advantages in both quantitative and qualitative terms: aggregate data is usually widely available, is much more employed and is less affected by measurement error. An additional difficulty of sectoral studies is that it is usually assumed that the idiosyncratic (sectoral) and the aggregate shocks share the same dynamics, as in Imbs et al. (2005) and Crucini and Shintani (2006). If this hypothesis is violated, which is likely to be the case in many instances, sectoral estimates would be inconsistent. This problem does not arise when modeling aggregate data since (under certain assuptions) the idiosyncratic component vanishes in the aggregate so that one should only care about the dynamics of common shocks. Finally, measures of persistence computed with sectoral data are less developed and worse understood than those based on aggregate data. The use of inappropriate measures of sectoral persistence has lead to inaccurate estimates in many instances, see Section 5 below, Mayoral (2007) and Gadea and Mayoral (2007) for more details on this issue.

Thus, the purpose of this paper is twofold: firstly, using a data set of aggregate RERs, we investigate how they can be modelled and estimated in the presence of sectoral heterogeneity and secondly, the estimated models are employed to provide new evidence about the holding of the PPP and the speed of reversion to parity using a data set of aggregate RERs. To do so, we first examine the stochastic properties of the aggregate model when sectoral heterogeneity is introduced. It turns out that when sectoral adjustment heterogeneity is allowed for, the aggregate model admits an AR(∞) representation, that can be stationary or not, depending on the distribution of the sectoral adjustment parameter. More specifically, the behavior of the distribution of that parameter would determine whether the aggregate RER is $I(0)$, $I(1)$ or a fractionally integrated process of order $d \in (0, 1)$. Clearly, the appropriate estimation strategy to be employed depends on the former characteristics of the process and we will review how consistent estimates can be obtained in either case.

Next, we apply this theory to the data set of RER elaborated by Taylor (2002). A wide range of econometric tools, under both the classical and the Bayesian approach have been employed to characterize the properties of the processes and to obtain parameter estimates. Fractional orders of integration, d, strictly positive but smaller than 1 have, in general, been obtained. The former result ($d > 0$) can be interpreted as a sign of the importance of allowing for sectoral heterogeneity, while the latter ($d < 1$) supports the long-run holding of the PPP hypothesis. Impulse response functions (IRFs) and half-lives (HLs) are provided and it is shown that, in general, the latter are considerably higher than the traditional 3-5 year *consensus,* in line with some recent literature that argues that the puzzle is larger than it was first believed.

Finally, we examine how the results presented in this paper are related to previous findings. We show how the existence of FI behavior in RER is consistent and can help to understand previous findings, such as the low power of unit root tests or the results obtained by Murray and Papell (2002) and Lopez et al. (2003, 2004). Our results are also very different to those obtained by studies that, like ours, allow for heterogenous sectoral behavior but use disaggregate data, such as Imbs et al. (2005), and we will also examine the source of the divergence.

The remainder of the paper is organized as follows. Section 2 examines the stochastic properties of the aggregate process when sectoral heterogeneity is allowed for and reviews how consistent estimates of the aggregate model can be obtained in this case. Sections 3 and 4 contain the empirical results. Section 3 reports the results of tests that help to determine the stochastic properties of the processes and identifies and estimates the most appropriate models. Section 4 presents half-life estimates and other measures of persistence. Section 5 discusses our results and compares them to what has been found in previous analyses. Finally, Section 6 concludes. Appendix 1 and 2 include some complementary material.

2. HETEROGENEOUS SECTORAL DYNAMICS IN RER AND AGGREGATION

As noticed by Imbs et al. (2005) "*one recurrent conclusion in most of the existing work [on sectoral exchange rates] is heterogeneity, both across sectors and across countries.*" Indeed, evidence supporting this statement is ample see, for instance, Engel (2002), Cheung et al. (2001), Campa and Goldberg (2002), Crucini and Shintani (2006), Gopinath and Rigobon (2006) and Campa and González (2006).

Sectoral heterogeneity might stem from a wide variety of sources. In an early study, Dornbusch (1987) suggests that the response of relative prices to exchange rate movements critically depends on the degree of market integration or separation, the degree of sustitutability between domestic and foreign variants of a product and on market structure. Differences in these aspects across industries and countries would bring about sectoral exchange rate heterogeneity.

In order to evaluate exchange rate persistence, studies that allow for heterogeneity in price adjustment dynamics typically consider the standard adjustment model for each of the sectors. The (log) of the sectoral real exchange rate for sector i and country j at time t with respect to a reference country, for instance the United States, is defined as

$$q_{it}^{j} = p_{it}^{j} - p_{it}^{usa},$$

where p_{it}^{j} is the log of the price of good i in country j and $p_{it}^{usa} = \log(P_{it}^{usa} \times E_t^j)$ is the log of the price of good i in the United States multiplied by the nominal exchange rate, E_t^j, defined as domestic currency units *per* U.S. dollar. In the following, the super index j will be dropped to simplify the notation.

In its simplest formulation, the adjustment model amounts to consider an AR(1) specification for the dynamics of q_{it} that allows for heterogeneity in the autoregressive parameter and in the level,

$$q_{it} = \alpha_i + \lambda_i q_{it-1} + \nu_{it}, \; i = 1, ..., \; N, \; t = 1, ...T, \tag{1}$$

$$\nu_{it} = \rho_i u_t + \varepsilon_{it}, \tag{2}$$

where α, λ, ρ, are random variables and u_t and ε_{it} represent a common and an idiosyncratic shock, respectively. Sectoral data is employed to compute estimates of persistence for each of the sectors and then, averages of these quantities are used as measures of real exchange rate persistence, see for instance Imbs et al. (2005) and Crucini and Shintani (2006). It is usually found that persistence estimates based on aggregate data are significatively higher than those derived from sectoral data.[2] For instance, Imbs et al. (2005) report estimates of averaged sectoral half-lives of around 1 year, in sharp contrast with the usual 3 to 5 year range typically found in

[2] A similar result has also been reported in the context of inflation persistence, see also Altissimo et al., (2006, 2007), Clark (2006), Lünnemann and Mathä, (2004), etc.

studies based on aggregate data. The reason that has been used to justify this divergence has been that established time series methods fail to control for heterogeneity and this induces a positive bias in persistence estimates. As Imbs et al. (2005) put it *"the aggregate real exchange rate is persistent because its components have heterogeneous dynamics".*

However, as argued in the Introduction, aggregate time series data presents important advantages with respect to sectoral data. Thus, one would ideally like to estimate exchange rate persistence with aggregate data controlling simultaneously for sectoral heterogeneity. In the following we examine how this can be accomplished.

If aggregate data can be employed for measuring persistence under sectoral heterogeneity, it has to be the case that aggregate and sectoral estimates of persistence are highly compatible. For this to happen, two necessary conditions should be verified; Firstly, shock persistence should be constant across aggregation levels, in such a way that the (population) effect over time of a common shock is the same, regardless of whether the measuring is considered at the sectoral or at the aggregate level. Secondly, it should be the case that consistent estimates of the relevant persistence measures could be obtained when aggregate data is employed.

Regarding the first condition, Mayoral (2007) has recently demostrated that there exists a tight link between aggregate and sectoral persistence even when sectors present heterogeneous dynamics. She has shown that if sectors behave as in (1), the standard impulse response function (IRF) computed at the aggregate level equals the average of sectoral impulses responses to a common shock, provided the number of sectors is sufficiently large. Since this result is crucial for developing the results of this paper, Appendix 1 presents a more detailed description of it, see also Mayoral (2007). It follows that the (average) persistence of a common shock remains contant across aggregation levels and thus, the aggregation of heterogeneous processes does not generate additional shock response, as has been argued in the literature. In other words, if shocks to the aggregate process are persistent, there are so because the micro units are, on average, persistent and not only because they are heterogeneous.[3]

With respect to the second condition, in order to obtain consistent estimates of shock persistence we should consider first estimation of the aggregate model, since most persistence measures (such as the IRF, the cumulative impulse response (CIR), the sum of the autoregressive coefficients (SAC), etc.) are usually functions of the corresponding model coefficients.

[3]See Section 5 for a description of the reasons why macro and micro estimates of persistence have found to be different.

If sectoral exchange rates behave as AR(1) processes with homogenous dynamics (that is, if λ in (1) is a constant), then the aggregate exchange rate would also be an AR(1) process. However, as is well-known the dynamics of the aggregate process becomes much more complicated under heterogeneity. Following Stoker (1984), the aggregate real exchange rate, Q_t, is obtained as the expected value of (1)

$$Q_t = E_s(\lambda q_{t-1}) + u_t, \tag{3}$$

where $Q_t = E_s(q_t)$, $E_s(\nu_t) = u_t$ and $E_s(.)$ denotes expectation across the distribution of sectors. Lewbel (1994) has shown that, under certain conditions (see Appendix 1), expression (3) can be written as

$$Q_t = \sum_{k=1}^{\infty} A_k Q_{t-k} + u_t, \tag{4}$$

for constants A_1, $A_2, ...,$ defined as $A_k = E(\alpha_k)$, where $\alpha_1 = \lambda$ and $\alpha_k = (\alpha_{k-1} - A_{k-1})\lambda$ for $k > 1$.

It follows that even when sectors have a simple AR(1) dynamics, under heterogeneity the aggregate real exchange becomes an AR(∞) process.[4]

If consistent estimates of the coefficients in (4) are obtained, aggregate data can be employed to compute persistence measures and the results should be highly compatible with those obtained when sectoral data is employed. However, (4) contains an infinite number of coefficients and thus, estimation is not a straight-forward task. In fact, the appropriate estimation strategy depends on the stochastic properties of Q_t and it is well-known that these properties crucially depend on the behavior of the distribucion of λ around 1. This would determine whether Q_t is integrated of order 0, I(0) in the following, integrated of order one, $I(1)$, or fractionally integrated of order d, *with* $0 < d < 1$, (henceforth, $FI(d)$), among other possibilities. For the sake of clarity, the following section describes the relation between the distribution of λ and the statistical properties of Q_t and next, we will focus on estimation.

[4] The assumption of AR(1) dynamics is just made for simplicity. It can be relaxed without any change in the main conclusion, see Lewbel (1994).

Stochastic properties of the aggregate real exchange rate under heterogeneity

Aggregate real exchange rate data is usually constructed as a (weighted) average of sectoral exchange rates. Assume, for simplicity, that aggregate data is given by

$$\bar{Q}_t = N^{-1} \sum_{i=1}^{N} q_{it}.$$

Building on previous results by Robinson (1978) and Granger (1980), Zaffaroni (2004) has shown that the stochastic properties of $\bar{Q}_t$ when $N \to \infty$ depend on the behavior of the distribution of λ, the heterogeneous AR coeficient in (12), around 1. Suppose that the support of λ is [0,γ_2], where negative values of λ are excluded for simplicity's sake. If $\gamma_2 < 1$, Zaffaroni (2004) shows that the aggregate process behaves as a stationary $I(0)$ process. However, if $\gamma_2 = 1$ and the P($\lambda = 1$) > 0, then the limit of $\bar{Q}_t$ as $N \to \infty$ is an I(1) process. An interesting intermediate case arises whenever the distribution of λ is absolutely continuous in the interval [0, 1). To characterize the properties of $\bar{Q}_t$ in this case, Zaffaroni (2004) uses a semiparametric characterization of the distribution of λ around unity,

$$f(\lambda) \sim c_d (1-\lambda)^{-d}, \text{ as } \lambda \to 1, \ 0 < c_d < \infty, \tag{5}$$

where '$\sim$' stands for asymptotic equivalence and $d \in (0,1)$ is a real parameter. This is a mild semiparametric specification of the cross-sectional distribution of λ that leaves the behavior of the density function of this parameter completely unspecified for any given interval [0, δ] with $\delta < 1$.[5] If the distribution of λ verifies this condition, then $\bar{Q}_t$ converges to a fractionally integrated process of order $d \in (0,1)$.[6]

Fractionally integrated processes were first introduced by Granger and Joyeux (1981) and Hosking (1980). The process y_t is said to be a fractionally integrated process of order d, FI(d), in reference to the degree of integration if

$$(y_t - \mu) = (1-L)^{-d} X_t, \tag{6}$$

[5] A particular case of this general family of distributions is the *Beta* distribution.

[6] As for the relation between the aggregate sectoral exchange rate $Q_t = E_s(q_t)$ defined in (3) and $\bar{Q}_t = N^{-1}\sum_{i=1}^{N} q_{it}$, it holds that $\bar{Q}_t \overset{L_2}{\to} E_s(q_t) = Q_t$ as N tends to infinity, provided the limit of $\bar{Q}_t$ is a stationary process. Otherwise, one should take first differences and in this case $(1-L)\bar{Q}_t \overset{L_2}{\to} (1-L)Q_t$. Thus, whenever nonstationarity is detected, the usual procedure of first differentiating the data would be sufficient in order to guarantee the convergence of $(1-L)\bar{Q}_t$ to $(1-L)Q_t$. In these cases, the IRF of Q_t can be estimated by first, estimating the IRF associated with $(1-L)\bar{Q}_t$, and then cumulating the corresponding values, see Mayoral (2007) for more details.

where X_t is an $I(0)$ process given by

$$X_t = \sum_{i=0}^{\infty} \alpha_i \varepsilon_{t-i},$$

with α_0=1, $\sum_{i=0}^{\infty} |\alpha_i| < \infty$, $\{\varepsilon_i\}_{i=0}^{\infty}$ is a martingale difference sequence and d is a real number. The fundamental long-term properties of y_t are governed by d and can be described in terms of interval regions for this parameter. For values of $d \in (-0.5, 0.5)$, y_t is stationary and invertible. If $d \in (0, 0.5)$, y_t is said to be a *long-memory* process, characterized by slowly decaying (nonsummable) autocorrelations, $\rho(k) \sim ck^{2d-1}$ for large k, where $\rho(.)$ represents the autocorrelation function and c is a constant. In other words, long memory is implied by a hyperbolic decay of correlations, as opposed to the case $d = 0$ where correlations decay exponentially fast. Finally, if $d \in [0.5, 1)$, y_t is non-stationary (it has unbounded variance) and yet mean reverts in the sense that shocks eventually die out.[7] Values of d greater than or equal to 1 imply a permanent behavior of shocks.

Notice that fractionally integrated models nest as a particular case the traditional $I(0) - I(1)$ framework and, therefore, this formulation is particularly convenient to model aggregate data under individual heterogeneity since it can accommodate its most relevant stochastic features, described at the beginning of this section.

A popular parametric characterization of y_t can be obtained by assuming that X_t admits an ARMA(p,q) representation. It is said in this case that y_t belongs to the class of ARFIMA(p,d,q) (autoregressive fractionally integrated moving average) processes, which are a natural generalization of the traditional ARIMA framework and is given by

$$\Phi_p(L)(1-L)^d (y_t - \mu) = \Theta_q(L)\varepsilon_t, \tag{7}$$

where $\Phi_p(z)$ and $\Theta_q(z)$ are the autoregressive and moving average polynomials, respectively, with zeros outside the unit circle and d is a real number. The polynomial $(1-L)^d = \sum_{i=0}^{\infty} \pi_i(d) L^i$, where $\pi_i(d) = \Gamma(i-d) / (\Gamma(-d)\Gamma(i+1))$ and $\Gamma(.)$ denotes the gamma function. Then, if the aggregate RER, Q_t defined in (3), admits a representation as in (7), then it can be rewritten as

$$\sum_{k=0}^{\infty} A_k Q_{t-k} = u_t = \Phi_p(L)\Theta_q(L)^{-1}(1-L)^d Q_t, \tag{8}$$

where the last term in (8) is also an AR(∞) process but only depens on a finite number of parameters.

[7] If y_t is nonstationary it will be assumed that $y_t = 0$ for all $t < 0$.

Estimation of the aggregate real exchange rate under heterogeneity

There is a widespread belief that if the dynamics of the micro units is heterogenous, estimates based on aggregate data can be highly misleading. For instance, Pesaran and Smith (1995) analyzed this issue by considering a micro model very similar to that described in (1). They concluded that "*in the dynamic case if the coefficients differ across groups, [.] aggregating give inconsistent and potentially highly misleading estimates of the coefficients.*"

Pesaran and Smith (1995) draw their conclusions by analyzing the properties of the OLS estimator of A_1 in a model like

$$Q_t = A_1 Q_{t-1} + a_t. \tag{9}$$

and they showed that $\hat{A}_1$ does not converge to the expected value of the heterogeneous coefficient, $E(\lambda)$. Notice, however, that model (9) is misspecified, since the DGP of Q_t is an AR(∞) process, as shown in (**??**). Furthermore, if Q_t is an I(0) process (whose spectral density is positive and finite around the zero frequency), Berk (1974) has shown that the AR(∞) polynomial can be approximated by an AR(k) process and consistent estimates can be achieved provided k grows at an appropriate rate with respect to the sample size T.[8] In practice, the order of the autoregression can be chosen according to the usual information criteria and good results are obtained even for moderate values of k. Hence, the lack of consistency in Pesaran and Smith aggregate estimates is a result of the incorrect specification of the aggregate model and not a consequence of the heterogeneity.

To illustrate this result, we have carried out a simple simulation study. For each replication, 1000 draws from the distribution of λ, the heterogeneous AR parameter, have been obtained, where $E(\lambda) = 0.77$ and $\sigma_\lambda = 0.23$.[9] Figure 1 presents the histogram of λ for one of the replications. Then, 1000 heterogeneous AR(1) sectors have been generated using these values of λ. The corresponding aggregate process was computed as the simple average of the micro units. Two models were fitted to the aggregate process: an AR(1) and an AR(k) process, where k is chosen according to the AIC. Notice that, according to model (4), the coefficient associated to Q_{t-1} is precisely $E(\lambda)$. The average value over all replications of $\hat{A}_1$ was 0.855 and 0.768, with standard

[8] More specifically, consistency is achieved provided $k, T \to \infty$ and $k^3/T \to 0$.

[9] More specifically, the distribution of λ was generated as a Beta(r,s) distribution on the interval (0,1) with parameter values equal to r=1.3 *and* s=0.3 and the resulting values were multiplied by 0.95, to bound the support of λ away from unity; see Figure 1.

deviations of 0.05 and 0.08, when the AR(1) or the AR(k) model was fitted, respectively. Not surprinsingly, the former value is far from the true one (0.77), as the results in Pesaran and Smith (1995) predict. However, when larger AR models are fitted, very reasonable estimates are obtained.

Finally, three IRFs were also computed. IRF$_{micro}$ is the average of the IRFs associated to each of the individual processes. IRF$_{macro}$-AR(k), refers to the IRF based on an AR(k) model fitted to the aggregate data, where k was chosen according to the AIC. Finally, IRF$_{macro}$-AR(1), denotes the IRF derived from an AR(1) specification for the aggregate data. Figure 2 plots the average over all the replications of the three IRFs described above. The number of replications was equal to 500 and the sample size was 200 observations.

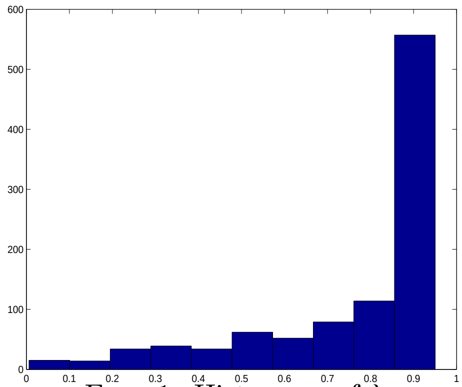

FIG. 1. Histogram of λ

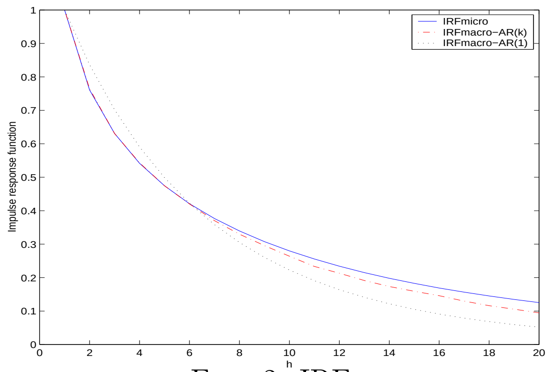

FIG. 2. IRFs

Figure 2 illustrates that highly compatible micro and macro estimates can be obtained using established time series techniques even when sectors are highly heterogeneous: the average of the micro IRFs and the IRF derived from the aggregate process when an AR(k) process was fitted are very close. The function IRFmacro-AR(1) corresponds to the misspecification case described in Pesaran and Smith. Not surprinsingly, the shape of this function is very different since in this case estimates are inconsistent.

If the integration order of Q_t is greater than zero, a natural estimation strategy would be to determine the order of integration and take the appropriate transformation so that the properties of the I(0) estimation framework are recovered. Thus, if the data is I(1), the first differences would be $I(0)$, and the transformed process can be consistently estimated by fitting a sufficiently large AR(k) process. Then, the IRF of the original data can be computed by cumulating the IRF of the differenced series. On the other hand, if the data is fractionally integrated, the vast literature on the estimation of FI(d) processes can be employed (see, for instance, Robinson,

2003) in order to determine the value of d and to obtain estimates of the remaining parameters. With respect to the estimation of the IRF for FI processes, it has been considered by Koops et al. (1997).

Fortunately, the above-mentioned characteristics of Q_t can be tested using only aggregate data and techniques that are already standard in the time series literature. The next section reviews some of them.

Summaring, under individual heterogeneity the use of aggregate data is perfectly justified and good estimates, in line with those that would be obtained if sectoral data was available, can be achieved.

3. MODELING AGGREGATE REAL EXCHANGE RATE DATA

In this section we investigate the statistical properties of a set of aggregate real exchange rates. We use the database elaborated by Taylor (2002) consisting of a sample of 20 countries over a period running from 1850 to 1996 (although some series start later), to which we have appended the IMF's *International Financial Statistics* extending the data time span to 2004. The countries included in the study are Argentina (ARG), Australia (AUS), Belgium (BEL), Brazil (BRA), Canada (CAN), Chile (CHL), Denmark (DNK), Finland (FI), France (FR), Germany (GE), Italy (IT), Japan (JPN), Mexico (MEX), Netherlands (NLD), Norway (NOR), Portugal (PRT), Spain (SP), Sweden (SWE), Switzerland (SWI) and Great Britain (GBR).[10] The data is annual and has been constructed as,

$$Q_t^j = P_t^j - P_t^{usa} - e_t^j$$

where P_t^j and P_t^{usa} are the logs of the price index for country j and the United States, respectively, and e_t^j is the log of the nominal exchange rate, defined as domestic currency units per US dollar. Hence, an increase in Q_t^j represents an appreciation of domestic currency in real terms. PPP holds if deviations from equilibrium have a transitory character, that is, if Q_t^j is a mean-reverting process. However, the holding of the PPP (mean reversion of Q_t^j) does not require the stationarity of Q_t^j, as is frequently imposed in this literature. If Q_t^j possesses an integration order strictly smaller than 1, the PPP will hold since, as illustrated in the previous

[10] Since the analysis is basically univariate, the different sample sizes are not difficult to deal with. In addition, some data is missing for specific periods such as the World Wars and some hyperinflation episodes. Following Taylor (2002), we have interpolated the series using the Tramo-Seats program (Gómez and Maravall, 1996).

section, shocks to Q_t^j are transitory in this case. In other words, the long-run holding of the PPP only requires the degree of integration of Q_t^j to be strictly less than one.

In the following subsections we apply a battery of tests to investigate the integration order of the series and then, we identify and estimate the corresponding models.

3.1. Unit root tests

We have begun our analysis by computing some standard unit root tests, namely, the MZt-GLS (Ng and Perron, 2001) and the KPSS tests (Kwiatkowski et al., 1992). The former considers that the model contains a unit root under the null hypothesis, while the latter sets the I(0) model as H_0. Table I presents the corresponding figures. The results vary greatly across countries: the existence of a unit root can be rejected for many of them (14 out of 20) but, nonetheless, the hypothesis of I(0) is also rejected for 11 countries and, surprisingly, it is possible to reject both hypotheses for 6 of them.

These ambiguous results have often been interpreted as an indicator of a behavior midway between the I(0) and the I(1) models. Hence, in the following subsection we explicitly allow for an order of integration in betweeen 0 and 1.

(TABLE I ABOUT HERE)

3.2. Integer versus fractional integration

We start by considering tests of integer ($d = 0$ or $d = 1$) versus fractional orders of integration. The hypothesis of $d = 1$ has clear economic implications: if $d = 1$ cannot be rejected in favor of a smaller order of integration, it would imply the failure of the PPP. It also implies that at least some sectors contain a unit root. On the other hand, rejection of $d = 0$ could be interpreted as a sign of the existence of a sizeable mass of highly persistent sectors, as discussed in Section 2.

To test the null hypothesis of $d = 1$ versus $d < 1$, the Fractional Dickey-Fuller (FDF) test has been applied to this data (see Dolado et al., 2002, 2006). This test generalizes the traditional Dickey-Fuller approach to test for $I(1)$ against $I(0)$ to the more general framework of $I(1)$ versus FI(d), with $d < 1$.[11] The results are reported in Table II. The conclusion of this table

[11] The test is based on the t-ratio associated with the coefficient of $(1-L)^d y_{t-1}$ in a regression of $(1-L) y_t$ on $(1-L)^d y_{t-1}$ and, possibly, some lags of $(1-L) y_t$ to account for the short-run autocorrelation of the process

is clear: the unit root model is clearly rejected against fractionally integrated alternatives in all countries, with the exception of Japan, in a regression where the only deterministic component is an intercept. If a trend is also introduced in the regression, the I(1) hypothesis can also be rejected for Japan. Analogously, tests of the hypotheses FI(d) vs. I(0) have also been computed following a similar approach (see Dolado et al., 2006) and, with the exception of Finland, the null hypothesis of FI(d) could not be rejected in any case.[12]

(TABLE II ABOUT HERE)

3.3. Estimation of aggregate RER data

The next step is to fit fractionally integrated models to our data set. To test the holding of the PPP, the parameter of interest is d, since it completely characterizes whether the exchange rate is mean-reverting or not. However, if one wants to study how convergence takes places, estimates of the remaining parameters of the model are also needed.

In order to check how robust our results are to the method employed, we have adopted two estimation approaches, classical and Bayesian, and we have computed several estimators from each class. The Bayesian approach has the advantage of providing the posterior distribution of d, and this allows one to compute interesting probabilities, such as the holding of the PPP, i.e., $\mathrm{P}(d < 1)$.

Within the classical approach we have considered the semiparametric methods (that only provide estimates of d) introduced by Geweke and Porter-Hudak (1983) (hereafter GPH) and the (feasible) local Whittle estimator (FELW), see Shimotsu, (2006).[13] Then, the corresponding $\hat{d} - th$ difference of the original process was computed and large $AR(k)$ models were fitted to the transformed process. Within the parametric class, the exact maximum likelihood (EML) and the nonlinear least squares method (NLS) proposed by Sowell (1992) and Beran (1994),

and/or some deterministic components if the series displays a trending behavior or initial conditions different from zero. To compute the tests, an estimated value of d under H_1 is required and the values of the EML estimator, reported in the fourth column of Table III, have been used.

[12]The results of these tests have been omitted for the sake of brevity, since similar conclusions can also be drawn from tests based on confidence intervals of the memory parameter, d, reported in Table III below.

[13]The feasible exact local Whittle estimator (FELW) is a recent version of the local Whittle estimator, initially proposed by Robinson (1995), that is consistent when the DGP is nonstationary and contains deterministic components.

respectively, have been computed. These estimation methods provide estimates of all the model parameters.

The results of the classical estimation are displayed in Table III.[14]

(TABLE III ABOUT HERE)

Several conclusions can be drawn from this table. Firstly, although the results vary slightly across the different estimation methods and countries, in general, fractional values of d, far from both 0 and 1, are found. Secondly, most countries exhibit values of d in the non-stationary but mean-reverting range ($0.5 \leq d < 1$). The fact that d is strictly less than 1 implies that the process is mean-reverting or, in other words, that the PPP holds in the long run. Nevertheless, if $d \geq 0.5$, the corresponding variable is non-stationary and shows a very slow rate of convergence to equilibrium. This confirms previous findings about the nature of shocks to RER, which are usually characterized as transitory but very persistent. Moreover, it also shows why it is difficult to reject the unit root hypothesis if a short sample is used since unit root tests have very low power against this type of FI process (Diebold and Rudebusch, 1991).

One problem often associated with classical estimators is the fact that they are very sensitive to the selection of the specific model in the parametric case, or to the choice of bandwidths in the semiparametric case. To overcome this problem, we have re-estimated the series adopting a Bayesian approach, which in this framework presents at least three advantages with respect to classical estimation. Firstly, the method takes model uncertainty into account when making inferences on d, which is something that cannot be done using the classical parametric approach. Secondly, (posterior) exact finite sample distributions of d are obtained. This allows us to compute interesting probabilities, as the probability that the PPP holds for country j, $P(d_j < 1)$, or the probability that Q_{jt} is a stationary process, $\mathrm{P}(d_j < 0.5)$. Finally, it has been argued that, although FI models contain the integer $I(1) - I(0)$ class as a particular case, in practice it is very unlikely that an integer value of d is obtained when the true value is in fact and integer. As pointed out by Hauser et al. (1999), this could be a drawback of FI models, especially when they are used for measuring persistence. Nevertheless, as discussed by Koop et al. (1997) and Gadea and Mayoral (2006), the Bayesian approach allows us to attach prior probability mass to integer values of d so that the ARIMA and ARFIMA specifications are on an equal footing.

[14] Only values of d are reported. The remaining AR and ARMA parameters obtained in the parametric methods are available upon request.

This will be particularly useful in Section 4 when persistence measures are computed.

We have followed the method proposed by Koop et al. (1997). Nine ARFIMA specifications for each country have been estimated, considering all possible combinations of ARFIMA models with p, $q \leq 2$.[15] Table IV reports the mean and standard deviations of d for three types of models: best, mean and weighted. The *best* model is the most likely according to the posterior probabilities while the *weighted* and *mean* are constructed taking the average of all possible specifications according to their posterior probability and as a simple average, respectively. Figures 3a and 3b display the graphs of the posterior distributions of d.

The results are in accordance with the main conclusions obtained using classical estimation. For all countries, values of d far from 0 and 1 are obtained, confirming the highly persistent, though transitory, character of shocks to Q_t^j. Figures 3a and 3b also highlight the important variability associated with the estimation of this parameter. Using the posterior density of d, it is easy to compute the probabilities that d belongs to each of the regions of interest. Figure 5 reports the probabilities that the PPP holds, P(d^j<1), and that the RER is stationarity, P(d^j < 0.5), for each country j. The figures have been computed using the *mean* and the *weighted* distributions (denoted as $P_m(.)$ and $P_w(.)$, respectively). The average probability that $d < 1$ is very high, around 85%. Nevertheless, the average probability of stationarity is much lower, around 37%. Thus, almost 50% of the posterior probability mass corresponds to a case where the PPP holds but deviations from PPP are non-stationary so that convergence is very slow.

(TABLE IV ABOUT HERE)

4. HALF LIVES AND OTHER MEASURES OF PERSISTENCE

Section 3 shows that the PPP holds in the long run for our long-span RER data set, in line with recent literature. This section investigates how the convergence takes place. To do so, we

[15] A preliminary analysis showed that lags higher than 2 were not significant, a plausible result when working with annual data. A flat prior probability was attached to each model and a uniform density for d in the interval $[0, 1.5]$ was assumed. Hence, the method puts an equal 1/3 of the prior probability mass on values of d that imply stationary shocks, non-stationary but mean-reverting shocks and, finally, permanent shocks, corresponding to the intervals [0, 0.5), [0.5,1) and [1, 1.5], respectively.

compute IRFs and Half-lives that are derived from the models estimated in Section 3.[16]

The IRF measures "*the effect of a change in the innovation ε_t by a unit quantity on the current and subsequent values of y_t*" (Andrews and Chen, 1994). The Half-life (HL) is defined as the number of periods that a shock needs to vanish by 50 percent and can be easily calculated from the IRF function as

$$IRF(HL) = 0.5. \tag{10}$$

In the ARFIMA context, the IRF(h) is defined as the h^{th} coefficient of the polynomial, $A(L) = (1-L)^{-d}\,\Phi(L)^{-1}\,\Theta(L)$. These coefficients can be computed according to the formula (see Koop et al., 1997 for details), $IRF(h) = \sum_{i=0}^{h} \pi_i(-d)\,J(h-i)$, where the $\pi_i(-d)'s$ come from the binomial expansion of $(1-L)^{-d}$ in powers of L and $J(.)$ is the standard ARMA(p,q) impulse response, given by $J(i) = \sum_{j=0}^{q} \theta_j f_{i+1-j}$, with $\theta_0 = 1$, $f_h = 0$ for $h \leq 0$, $f_1 = 1$ and $f_h = -\left(\phi_1 f_{h-1} + ... + \phi_p f_{h-p}\right)$, for $h \geq 2$.

Table V reports the estimated HLs while Figures 5a and 5b present the evolution of the IRFs for a horizon of twenty years. The second column of Table V presents HL estimates obtained using Sowell's (1992) EML method while columns 3 to 5 display the results obtained from the three Bayesian models reported in Table IV.[17]

(TABLE VI ABOUT HERE)

Using the Bayesian posterior densities, it is possible to compute the probabilities that the HL lies in a pre-specified range. Figure 6 reports the probability that the HL is less than k years for $k = 3, 5$, and 10. From here, it is easy to compute the probability of the so-called Rogoff's interval (HLs in the 3-5 year range). Two types of probabilities are reported for each value of k, computed from the *mean* and the *weighted* models, denoted as P_m and P_w, respectively.

Several interesting conclusions can be drawn from the inspection of Figure 6. The probability of half-lives being inferior to 5 years is bigger than 0.5 for only seven of the 20 countries analyzed (ARG, AUS, BEL, CHL, FIN, FRA and MEX). Thus, in general, the probabilities of HLs being in the so-called Rogoff's interval (3-5 years) are fairly small. Also, only the half of countries have

[16] A detailed revision of persistence measures in a fractional integration context can be found in Gadea and Mayoral (2006).

[17] HLs and IRFs were also computed using the semiparametric estimates of d and the large AR(k) models fitted to the residuals. The results were very similar as those obtained with the parametric methods and then, they are not reported for the sake of brevity.

$P(HL < 10$ years) bigger than 50%, which confirms the very persistent character of deviations from equilibrium.

Hence, these results differ considerably from those reported in the original work of Taylor (2002), who obtains a mean half-life of around 4 years for the whole period. Our findings are closer to those of Murray and Papell (2002, 2005a) and Lopez et al. (2004) who also find HLs above the 3-5 Rogoff consensus. Using Taylor's (2002) database and median-unbiased estimators, Lopez et al. (2004) find a median HL of 11.34 and 7.55 in DF-GLS and ADF regressions, respectively, and Murray and Papell (2005b) apply different lag selection methods and find that long-run PPP held for the real exchange rates of only 9 out of the 16 industrialized countries in Taylor's sample Murray and Papell (2005a) use the data of Lothian and Taylor (1996) -two centuries of franc-sterling real exchange rate- and obtain a median HL between 5.95 and 16.98 -depending the number of lags selected- with respective bounds of [3.49,16.98] and $[16.98,\infty)-$ if they use the unbiased-median estimator. In the following section we discuss how the latter set of results are related to ours.

5. RELATIONS WITH PREVIOUS RESULTS

The literature on the PPP puzzle is very large and the conclusions about the duration of deviations to PPP differ greatly depending on the data, the techniques and the assumptions adopted. In this section we examine previous results in the area in the light of the new findings reported in this paper.

Two main conclusions arise from the previous sections. Firstly, if sectors present heterogeneous dynamics, aggregate data can still be employed to obtain consistent estimates of persistence measures. However, since aggregate dynamics might become considerably more involved in that case, a careful examination is needed to identify the appropriate model. Secondly, our results suggest that FI processes provide a good fit for this data set. Conditional on this finding, we obtain substantial support of the holding of the PPP and half lives that are considerably higher than the usual 3-5 year consensus or, in other words, our results are in accordance with a recent strand of literature that claims that the size of the puzzle is bigger than was first thought.

We now compare these results to some previously obtained in the literature.

a. Results based on sectoral data.

In Section 2 it was argued that persistence, as measured by the IRF and related measures, remains constant across aggregation levels. However, the results reported in this paper are in sharp contrast to those presented by papers that use sectoral data to draw their conclusions. For instance, Imbs et al. (2005) analyze a data set of monthly sectoral real exchange rates and report estimates of price adjustment that are completely in line with models of slow nominal price adjustment, with an 'average' HL of price adjustment of about 1 year.

Imbs et al. (2005) use the IRF (and related measures such as the HL and the cumulative impulse response, CIR) to establish persistence comparisons across aggregation levels. However, their definition of sectoral IRF differs from the one employed in Appendix 1 to establish the equivalence of persistence measures across aggregation levels. Instead of averaging the sectoral impulse responses that, for the simple model considered in (1) yield $IRF\,(h)_{\text{sect.}} = E_s\left(\lambda^h\right)$, see (16), their estimates of sectoral response are computed by firstly, averaging the AR coefficients of the different sectors and secondly, computing the standard IRF of this "averaged" model. Thus, for the simple sectoral model in (1) Imbs et al. (2005) present estimates of the function

$$\overline{IRF}\,(h) = E\,(\lambda)^h\,,\ \text{for } h \geq 0, \tag{11}$$

as measures of sectoral persistence. Clearly, since the IRF is a highly nonlinear function, averaging the IRFs or averaging the model coefficients and then computing the IRF may yield very different results. In fact, for most empirically relevant cases, Jensen's inequality implies that the former measure is larger than the latter.

Furthermore, Gadea and Mayoral (2007) have shown that it is precisely this definition of sectoral IRF what drives Imbs et al. (2005) results. Using the same data set as the latter authors, it is shown that when the average of the sectoral IRFs is employed to evaluate sectoral persistence, standard values of persistence are recovered. Moreover, sectoral results are highly compatible with those obtained with aggregate data, as the theory predicts.[18]

b. *Results based on aggregate data.*

Our conclusions about the existence of FI behavior in RER data can help to explain some of the findings reported in the literature. For instance if the data is FI(d) with $d < 1$, unit root tests are in general consistent (reject the I(1) hypothesis with a probability that tends to 1) but lack power to reject the (false) unit root hypothesis unless long data spans are employed, see

[18] Mayoral (2007) has also demostrated that other persistence measures, such as the sum of the autoregressive coefficients or the largest autoregressive root are not suitable for establishing persistence comparisons across aggregation levels, see the latter paper for details.

Diebold and Rudebush, (1991) and Diebold and Senhadji (1996).

On the other hand, FI(d) models can be approximated by a long autoregression. However, one should expect an important downward bias of the autoregressive parameters in small samples when processes are very persistent, as in the case where RERs are considered. However, the use of median-unbiased estimators applied on a sufficiently long autoregression could yield a good approximation of the true IRF when the model is FI. Then, one should expect that the results based on bias-correcting techniques and those computed with FI(d) are highly compatible, as it actually happens. Notice however that the FI parametrization has several advantages over a long autoreggresion since it is usually more parsimonious and estimates of d do not present a systematic small-sample bias as opposed to AR parameters in very persistent processes. Thus, no correction has to be performed even when the process is very persistent.

Finally, nonlinear models have become a popular alternative for modelling the dynamics of real exchange rates. According to this approach, transaction costs in international arbitrage drive real exchange rates to follow a nonlinear mean-reverting process where the speed of adjustment is directly related to the size of the deviations from equilibrium. By introducing a threshold in the real exchange rate adjustment, several authors have found further evidence of reversion, as in Michael et al. (1997), as well as noticeably reduced half-life deviations, as in Taylor et al. (2001).

However, it is well-known that there is an identification problem between FI(d) models and some type of non linearities (see Mikosh and Starika, (2004), Bhattacharya et al. (1983), Diebold and Inhoue (2001), Perron and Qu (2004) among many others). An analysis of whether the nonlinearities found in RERs are spurious and due to the existence of FI is beyond the scope of this paper but it is currently under investigation by the authors.

6. CONCLUSION

In recent years, there has been a rebirth of interest in the sources and measurement of PPP deviations. However, in spite of the large quantity of literature on this subject, there is still an open debate. In this paper, we allow for sectoral heterogeneity in the speed of adjustment to parity and we show that, under mild assumptions, the aggregate RER can be still be used to compute estimates of persistence that should be very similar as those that computed if sectoral data was available. Using a wide range of econometric tools we find robust empirical evidence of fractional integration in the real exchange rates. The memory parameter, d, takes values within the interval (0,1), implying that real exchange rates are mean-reverting but highly persistent processes. This is enough to re-establish the PPP hypothesis as a valid rule for the very long run, although the size of Rogoff's puzzle (the 3-5 year consensus only has a probability of around 15%) has increased considerably.

We have also examined how the results presented in this paper are related to previous findings. It has been shown how the existence of FI behavior in RER is consistent and can help to understand previous findings, such as the low power of unit root tests or the findings obtained by Murray and Papell (2002) and Lopez et al. (2003, 2004). Our results are also very different to those obtained by studies that, like ours, allow for heterogenous sectoral behavior but use disaggregate data, such as Imbs et al. (2005). Explanations about the source of the discrepances have also been provided.

REFERENCES

Abuaf, N. and Ph. Jorion, 1990. "Purchasing Power Parity in the Long Run" *Journal of Finance* 45, 157-74.

Andrews, D. and H. Chen, 1994. "Approximately median-unbiased estimation of autoregressive models" *Journal of Business and Economic Statistics* 12, 187-204.

Baum, Ch. F., Barkoulas, J.T., and M. Caglayan, 1998. "Long memory or structural breaks: Can either explain non stationary real exchange rates under the current float?" Manuscript.

Beran, J., 1994. *Statistics for Long-Memory Processes.* Chapman & Hall.

Caporale, G.M, M.Cerrato and N. Spagnolo, 2004. "Measuring half-lives using a non-parametric bootstrap approach" *Public Policy Discussion Papers* Brunel University 04-13.

Cashin, P. and C.J. McDermott, 2003. "An Unbiased Appraisal of Purchasing Power Parity" *IMF Staff Papers* 50, 321-351.

Cheung, Y.-W. and K.S. Lai, 1993. "A Fractional Cointegration Analysis of Purchasing Power Parity" *Journal of Business and Economics Statistics* 11, 103-12.

Crucini, M.J. and M. Shintani, 2006. "Persistence in Law-of-One-Price Deviations: Evidence from Micro-data" UCLA Department of Economics.

Davidson, J. and P. Sibbertsen, 2005. "Generating schemes for long memory processes: regimes, aggregation and linearity" *Journal of Econometrics* 128, 253-282.

Diebold, F.X., and A. Inoue, 2001. "Long memory and regime switching", *Journal of Econometrics"* 105, 131-159.

Diebold, F.X., and G. Rudebusch, 1991. "On the power of the Dickey-Fuller tests against fractional alternatives" *Economic Letters* 35, 55-160.

Diebold, F.X., Husted, S., Rush, M, 1991. "Real Exchange Rates under the Gold Standard" *Journal of Political Economy* 99, 1252-71.

van Dijk, D., P.H. Franses and R. Paap, 2002. "A nonlinear long memory model for US unemployment" *Journal of Econometrics* 102, 135-165.

Dolado J.J., J. Gonzalo and L. Mayoral, 2002. "A Fractional Dickey-Fuller test for unit roots" *Econometrica* 70, 1963-2006.

Dolado J.J., J. Gonzalo and L. Mayoral, 2005. "Testing I(1) against I(d) alternatives in the presence of deterministic components" Working Paper, Universitat Pompeu Fabra.

Dolado J.J., J. Gonzalo and L. Mayoral, 2006. "What is what: A simple time-domain test for long-memory vs. structural breaks" Working Paper, Universitat Pompeu Fabra.

Doornik, J. and M. Ooms, 2006. "A package for estimating, forecasting and simulating ARFIMA models: ARFIMA package 1.4 for OX" Mimeo.

Dueker M. and A. Serletis, 2000. "Do Real Exchange Rates have Autoregressive Unit Roots? A Test under the Alternative of Long Memory and Breaks" *Federal Reserve Bank of St. Louis* 2000-016A.

Elliott, G., Rothenberg, T.J.and J.H. Stock, 1996. ”Efficient test for an autoregressive unit root” *Econometrica* 64, 813-36.

Frankel, J.A., 1986. "International Capital Mobility and Crowding out in the U.S. Economy: Imperfect Integration of Financial Markets or of Goods Markets?" In *How Open Is the U.S. Economy?* edited by RW. Hafer. Lexington, Mass: Lexington.

Gadea, M.D. and L. Mayoral, 2006. "The Persistence of Inflation in OECD Countries: a Fractionally Integrated Aproach" *International Journal of Central Banking* 2, 51-104.

Geweke, J. and S. Porter-Hudak, 1983. "The estimation and application of long memory time series models" *Journal of Time Series Analysis* 4, 221-238.

Gil-Alana, L.A., 2000. "Mean Reversion in the real exchange rates" *Economics Letters* 69, 285-288.

Gómez, and A. Maravall, 1996. "Programs TRAMO and SEATS" Banco de España, Documento de Trabajo.

Gopinath, G. and R. Rigobon, 2006. "Sticky Borders" NBER, Working Paper 12095.

Granger, C.W.J., 1980. "Long-memory relationships and the aggregation of dynamics models" *Journal of Econometrics* 14, 227-238.

Granger, C.W.J and K. Joyeux, 1980. "An introduction to long memory series" *Journal of Time series Analysis* 1, 15-30.

Granger, C.W.J. and M. Hyung, 2004 "Occasional structural breaks and long memory with an application to the S&P 500 absolute stock returns" *Journal of Empirical Finance* 11, 399-421.

Hauser, M.A., B.M. Pötscher and E. Reschenhofer, 1999. "Measuring persistence in aggregate output: ARMA models, fractionally integrated ARMA models and nonparametric procedures" *Empirical Economics* 24, 243-269.

Henry, M. and P. Zaffaroni, 2002. "The long range dependence paradigm for Macroeconomics and Finance." in P. Doukhan, G. Oppenheim and M. Taqqu (ed). *Long range dependence: Theory and applications.* Birkhäuser: Boston.

Hosking, J.R.M., 1981. "Fractional differencing" *Biometrika* 68, 165-176.

Imbs, J. A., Muntaz, M.O. Ravn and H. Rey, 2005. "PPP Strikes Back: Aggregation and the Real Exchange Rate" *Quarterly Journal of Economics* CXX(1), 1-43.

Knetter, M. M., 1993. "International Comparisons of Price-to-Market Behavior". *American Economic Review* 83, 473-486.

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.

Kwiatkowski, D., P.C.B. Phillips, P. Schmidt and Y. Shin, 1992. "Testing the null hypothesis of stationarity against the alternative of a unit root" *Journal of Econometrics* 54, 159-178.

Lewbel, A., 1994. "Aggregation and simple dynamics" *American Economic Review* 84, 905-918.

Lopez, C., Ch. Murray and D.H. Papell, 2003. "State of the Art Unit Root Test and the PPP Puzzle" Manuscript.

Lopez, C., Ch. Murray and D.H. Papell, 2004. "More Powerful Unit Root Test and the PPP puzzle" Manuscript.

Lothian, J.R. and M.P. Taylor, 1996. "Real exchange rate behavior: the recent float from the perspective of the past two centuries" *Journal of Political Economy* 104, 488-509.

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, 2005a. "The Purchasing Power Parity is Worse Than You Think" *Empirical Economics* 30, 783-790.

Murray, C. and D.H. Papell, 2005b. "State of the Art Unit Root Tests and Purchasing Power Parity" *Journal of Money, Credit and Banking* 37, 361-369.

Newey, W. and K. West, 1994. "Automatic Lag Selection in Covariance Matrix Estimation." *Review of Economics Studies* 61, 631-653.

Ng, S. and P. Perron, 2001. "Lag length selection and the construction of unit root tests with good size and power" *Econometrica* 69, 1519-54.

Nunes, Luis C, Ch.-M. Kuan and P. Newbold, 1995. "Spurious break" *Econometric Theory* 11, 736-749.

Perron, P. and Z. Qu, 2004. "Estimating Restricted Structural Changes Models" *Journal of*

Econometrics (forthcoming).

Robinson, P.M., 1978. "Statistical Inference for a Random Coefficient Autoregressive Models" *Scandinavian Journal of Statistics* 5, 163-168.

Robinson, P.M., 1995. "Gaussian semiparametric estimation of long range dependence" *Annals of Statistics* 23, 1630-1661.

Rogoff, K., 1996. "The Purchasing Power Parity Puzzle" *Journal of Economic Literature* XXXIV, 647-68.

Rossi, B., 2003. "Confidence intervals for half-life deviations from Purchasing Power Parity" Manuscript.

Saikkonen, P. and R. Luukkonen, 1988. "Lagrange multiplier tests for testing non-linearities in time series models" *Scandinavian Journal of Statistics* 15, 55-68.

Shimotsu, K., 2006. "Exact Local Whittle Estimation of Fractional Integration with Unknown Mean and Time Trend" Manuscript revised in March 2006.

Smallwood A. D., 2005. "Joint Test for Long Memory and Non-linearity: The Case of Purchasing Power Parity" *Studies in Nonlinear Dynamics & Econometrics* 9, article 7.

Sowell, F., 1990. "The Fractional Unit Root Distribution". *Econometrica* 58, 495-505.

Sowell, F., 1992. "Maximum likelihood estimation of stationary univariate fractionally integrated time series" *Journal of Econometrics* 53, 165-188.

Taylor, A,M., 2002. "A century of Purchasing-power Parity" *The Review of Economics and Statistics* 84, 139-150.

Taylor, M.P., D.A. Peel and L. Sarno, 2001. "Non-lineal mean-reversion in real exchange rates: toward a solution to the purchasing power parity puzzles?" *International Economic Review* 42, 1015- 1042.

Teverosky, V. and M. Taqqu, 1997. "Testing for long-range dependence in the presence of shifting means or a slowly declining trend, using a variance-type estimator" *Journal of Time Series Analysis* 18, 279-304.

Yang, J., 1997. "Exchange rate pass-through in 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.

TABLES

TABLE I

Unit root and stationarity tests

	MZ_t-GLS	KPSS
ARG	-4.09**	0.11
AUS	-2.43**	0.81**
BEL	-3.37**	0.74**
BRA	-2.53**	0.16
CAN	-1.77	0.78**
CHL	-1.02	1.06**
DNK	-2.00	0.68**
FIN	-5.70**	0.18
FRA	-2.33*	0.89**
GE	-2.67**	0.39
ITA	-3.99**	0.07
JPN	0.17	1.14**
MEX	-2.50*	0.81*
NLD	-2.58**	0.43
NOR	-2.37*	0.46*
PRT	-1.85	0.40
SP	-2.87**	0.31
SWE	-3.21**	0.59*
SWI	-0.68	0.94**
GBR	-2.58**	0.38

TABLE II

FDF Test (I(1) versus FI(d)).

$H_0 : d = 1$; $H_1 : d < 1$		
	with intercept	with intercept and trend
ARG	-4.47**	-4.53**
AUS	-2.99**	-3.41**
BEL	-4.10**	-4.46**
BRA	-2.57**	-2.04*
CAN	-2.99**	-3.31**
CHL	-2.26*	-2.96**
DNK	-2.74**	-3.45**
FIN	-6.20**	-6.24**
FRA	-4.66**	-4.75**
GE	-2.43*	-2.47*
ITA	-3.87**	-4.08**
JPN	-0.97	-3.01**
MEX	-3.54**	-3.56**
NLD	-2.55*	-2.83**
NOR	-4.22**	-4.56**
PRT	-1.68+	-3.61**
SP	-3.50**	-3.61**
SWE	-3.99**	-4.12**
SWI	-2.95*	-3.60*
GRB	-2.29*	-3.31*

Notes: (**) , (*) Significant at the 1% and 5% level, respectively. An intercept was included to compute the statistics. The lag length was chosen according to the SBIC criterion in the MZt-GLS and Bartlett's window was used as a kernel estimator in the KPSS. The bandwidth was chosen according to Newey and West (1994).

TABLE III

Classical Estimation of FI(d) models

	GPH	FELW	EML	NLS
ARG	0.40 (0.293)	0.47 (0.158)	0.64 (0.096)	0.61 (0.097)
AUS	0.74 (0.257)	0.49 (0.144)	0.50 (0.187)	0.46 (0.181)
BEL	0.46 (0.274)	0.26 (0.151)	0.31 (0.090)	0.34 (0.103)
BRA	-0.10 (0.293)	0.89 (0.158)	0.92 (0.090)	0.91 (0.090)
CAN	0.92 (0.257)	0.56 (0.144)	0.48 (0.137)	0.46 (0.213)
CHL	0.44 (0.347)	0.70 (0.177)	0.46 (0.261)	0.94 (0.151)
DNK	0.52 (0.273)	0.62 (0.151)	0.66 (0.108)	0.68 (0.102)
FIN	-0.21 (0.273)	0.14 (0.151)	−0.30 (0.325)	−0.26 (0.312)
FRA	0.18 (0.274)	0.35 (0.151)	0.77 (0.283)	0.68 (0.171)
GE	0.61 (0.274)	0.81 (0.151)	0.17 (0.219)	0.55 (0.275)
ITA	0.33 (0.273)	0.42 (0.157)	0.43 (0.117)	0.39 (0.117)
JPN	0.74 (0.293)	0.57 (0.158)	0.47 (0.335)	0.90 (0.188)
MEX	0.76 (0.293)	0.51 (0.158)	0.50 (0.107)	0.53 (0.096)
NLD	0.72 (0.257)	0.63 (0.144)	0.34 (0.379)	0.42 (0.301)
NOR	0.59 (0.257)	0.54 (0.144)	0.53 (0.281)	0.32 (0.215)
PRT	0.83 (0.293)	0.69 (0.158)	0.73 (0.129)	0.56 (0.157)
SP	0.78 (0.273)	0.63 (0.151)	0.56 (0.106)	0.53 (0.106)
SWE	0.38 (0.273)	0.47 (0.151)	0.41 (0.135)	0.46 (0.122)
SWI	0.54 (0.293)	0.60 (0.158)	0.52 (0.125)	0.71 (0.158)
GRB	0.76 (0.257)	0.44 (0.144)	0.67 (0.100)	0.59 (0.106)

Notes: Standard errors in brackets. Computations for the GPH, EML and NLS estimators were carried out using the ARFIMA Package 1.04 for OX, (Doornick and Ooms, 2006) while the FELW was implemented using a

MATLAB code provided by the author. Parametric models were selected using the AIC.

TABLE IV

BAYESIAN ESTIMATION OF FI(d)

	best	mean	weighted
ARG	0.27 (0.18)	0.41 (0.22)	0.39 (0.20)
AUS	0.51 (0.20)	0.64 (0.23)	0.57 (0.22)
BEL	0.34 (0.14)	0.44 (0.20)	0.34 (0.20)
BRA	0.51 (0.28)	0.72 (0.24)	0.72 (0.23)
CAN	0.51 (0.25)	0.70 (0.26)	0.67 (0.23)
CHL	0.74 (0.10)	0.65 (0.24)	0.69 (0.19)
DNK	0.51 (0.21)	0.66 (0.22)	0.60 (0.22)
FIN	0.28 (0.17)	0.40 (0.23)	0.32 (0.22)
FRA	0.40 (0.21)	0.61 (0.23)	0.52 (0.27)
GE	0.52 (0.28)	0.80 (0.24)	0.72 (0.28)
ITA	0.42 (0.20)	0.53 (0.81)	0.44 (0.29)
JPN	0.68 (0.20)	0.76 (0.23)	0.71 (0.22)
MEX	0.56 (0.13)	0.59 (0.25)	0.54 (0.22)
NLD	0.63 (0.20)	0.71 (0.22)	0.64 (0.22)
NOR	0.65 (0.17)	0.67 (0.27)	0.54 (0.23)
PRT	0.58 (0.21)	0.68 (0.21)	0.63 (0.20)
SP	0.58 (0.21)	0.74 (0.20)	0.67 (0.29)
SWE	0.32 (0.22)	0.55 (0.23)	0.45 (0.23)
SWI	0.48 (0.24)	0.69 (0.23)	0.61 (0.24)
GRB	0.49 (0.20)	0.67 (0.24)	0.61 (0.22)

Notes: Standard errors in brackets. All variables are considered in differences. Estimates have been computed from 25,000 replications of the 9 possible ARFIMA models with p,q<=2.

Notes: $(^{**})$, $(^{*})$ Rejection at the 1% and the 5% level, respectively. Critical values: N(0,1). The FDF invariant regression $\Delta y_t = \alpha_1 \tau_{t-1}(d) + \phi \Delta^d y_{t-1} + \sum_{j=1}^{k} \Delta y_{t-j} + a_t$ has been run and the number of lags of Δy_t was chosen according to the AIC. The term $\tau_t(\delta)$ is defined as $\tau_t(\delta) = \sum_{i=0}^{t-1} \pi_i(\delta)$, where the coefficients $\pi_i(\delta)$ come from the power expansion of $(1-L)^{\delta}$.

Notes: $(^{**})$, $(^{*})$ Rejection at the 1% and the 5% level, respectively. These figures have been computed with a MATLAB code provided by the author.

TABLE VI

Estimation of half lives

	Classical	Bayesian		
		best	mean	weighed
ARG	2.8	2.0	2.2	2.1
AUS	5.8	9.3	8.6	8.7
BEL	2.5	2.8	4.7	3.6
BRA	>10	>10	9.7	9.9
CAN	8.5	>10	>10	>10
CHL	2.7	7.4	5.7	6.5
DNK	8.9	9.9	9.4	9.4
FIN	1.7	2.2	7.6	2.2
FRA	2.3	4.8	5.9	5.1
GE	8.7	>10	>10	>10
ITA	3.8	7.4	6.7	6.1
JPN	7.9	>10	6.0	5.8
MEX	2.8	3.4	4.3	4.0
NLD	7.3	>10	>10	>10
NOR	5.7	>10	>10	>10
PRT	>10	>10	>10	>10
SP	8.8	2.5	5.8	4.2
SWE	3.6	4.8	6.7	6.2
SWI	6.9	>10	>10	>10
GRB	7.9	6.9	7.7	7.8

Notes: HLs are calculated from IRFs. Under the classical approach, HLs have been computed using the EML estimates.

APPENDIX 1

Equivalence of persistence measures across aggregation levels

This appendix shows that the average response to an aggregate shock is the same across aggregation levels. Consider the sectoral adjustment model described in (1), given by

$$q_{it} = \alpha_i + \lambda_i q_{it-1} + \nu_{it},\ i = 1, ...,\ I,\ t = 1, ...T, \quad (12)$$

$$\nu_{it} = \rho_i u_t + \varepsilon_{it}, \quad (13)$$

where α, λ, ρ, are random variables and u_t and ε_{it} are $i.i.d$ processes representing a common and an idiosyncratic shock, respectively; ν_t is assumed to be independent of λ. It is also assumed that the distribution of λ has bounded support (typically, $\lambda \in [-1, 1]$) and that $E_s\left(\lambda^k\right)$ exists for all $k > 0$, where $E_s(.)$ denotes expectation over the distribution of sectors.

Suppose now that one wants to evaluate the average response over time of sectoral exchange rates to a unitary shock. Then for each sector i, the response to a unitary common shock can be evaluated through the impulse response function (IRF), defined as the difference between two forecasts (see Koops et al., 1996 and Jorda, 2005)

$$IRF_i(t, h) = E\left(q_{it+h}|u_t = 1; z_{ijt-1}\right) - E\left(q_{it+h}|u_t = 0; z_{it-1}\right) \quad (14)$$

where the operator E(.|.) denotes the best mean squared error predictor and $z_{it-1} = \begin{pmatrix} q_{it}, & q_{it-1}, & ... \end{pmatrix}'$; Application of this definition to (12) yields

$$IRF_i(t, h) = \rho_i \lambda_i^h,\ \text{for } h \geq 0. \quad (15)$$

Then, a natural measure of the average sectoral response to a unitary common shock can be obtained by averaging (15) over the distribution of sectors of country j, that is,

$$IRF\left(t, h\right)_{\text{sect.}} = E_s\left(IRF(t, h)\right) = E_s\left(\lambda^h\right),\ \text{for } h \geq 0, \quad (16)$$

since $E_s(\rho)$ has been normalized to 1.

To show that exactly the same values are obtained when one computes the IRF from the aggregate model, aggregation of sectors should be considered first. To derive the aggregate model, we follow Lewbel (1994) and we further assume that the distribution of sectors is countably or