Aggregation Bias and the PPP Puzzle

By
Rudolf Rajczi

Submitted to
Central European University
Department of Economics

In partial fulfilment of the requirements for the degree of
Master of Arts in Economics

Supervisor: Professor Attila Ratfai

Budapest, Hungary
2013

Abstract

Using an extensive data set of European disaggregate price indices I reexamine Imbs et al.'s (2005*b*) findings on the importance of dynamic aggregation bias in explaining the Purchasing Power Parity puzzle. My results suggest that there is clear evidence for such bias in sectoral data, but allowing for heterogeneous adjustment processes across panel members can in large extent explain the aggregation bias even on the country-level. Once heterogeneity is taken into account and the cross-sectional correlation of the error terms is controlled for, the estimated half-life of shocks hitting the real exchange rates can drop below one year, providing further empirical support against Rogoff's famous puzzle.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Introduced by Cassel (1918) in the field of modern macroeconomics, the theory of Purchasing Power Parity is one of the oldest concepts and one of the most intensively researched areas in international economics. It builds on the notion of the Law of One Price (LOP), which states that if markets function properly, there is no room for arbitrage on the level of individual goods irrespective from the point of purchase, so that once expressed in a common currency, prices of identical goods should not differ across countries. The PPP is a direct extension of the LOP to aggregate price indices and assumes a co-movement between the relative price levels and the nominal exchange rates for any given pair of countries. It is most commonly analyzed in terms of movements of the real exchange rates (which is defined as the nominal exchange rate adjusted for relative national price levels), since, as it follows from the previous definition, observed variations in the real exchange rates indicate deviations from the PPP.

Of course, the LOP and the PPP seldom hold in practice: transportation costs, customs or other types of trade restrictions, menu costs, higher costs associated with foreign transactions (transaction costs, opportunity cost of leisure) or differences in preferences even for the very same good can prevent international relative prices to closely follow the changes in the nominal exchange rates. Therefore, as Rogoff notes, "while few empirically literate economists take PPP seriously as a short-term proposition, most instinctively believe in some variant of purchasing power parity as an anchor for long-run real exchange rates" (Rogoff, 1996, p. 647). As a result, the notion of the PPP is generally used in international economics when analyzing long-term phenomena and is widely used for calculating equilibrium exchange rates, comparing national levels of income or calculating international poverty measures.

Starting from the 1970s, the empirical analysis of the PPP has gathered substantial attention and led to a vast number of studies focusing on the empirical justification of the PPP-theory.[1] These studies (especially those based on samples from the post-Bretton Woods floating exchange rate periods) faced difficulties in finding support for the stationarity of the real exchange rates and hence, justification of the PPP as a valid long-run phenomenon.[2] As researchers associated the failure to reject the null hypothesis of random walks in the context of real exchange rate with the low power of the employed statistical tests[3], the focus shifted on improving the power of the tests by extending the dimensions of the analyzed data sets.

[1]For some of the early contributors see e.g. Isard (1977); Frenkel (1980).

[2]For an extensive review of the early literature see e.g. Froot and Rogoff (1995); Rogoff (1996); Taylor (2002).

[3]As Taylor (2001) notes, with annual single time series data and a true parameter value of an underlying AR(1) process of $\rho = 0.87$ (which translates to a half-life of 5 years for the shocks), one would require 128 observations to be able to reject the null hypothesis of a random walk at the 5% significance level in case of a standard Dickey-Fuller unit root test.

One way to do that was to increase the length of the sample periods; however, the availability of long-term time series is in general fairly limited[4], moreover, the results are potentially biased by the sturctural breaks (such as exchange regime changes) in the sample.[5] As an alternative approach to increas the number of observations in the sample, the vast majority of empirical studies have focused on the joint analysis of several real exchange rates in the panel context.[6] As Rogoff (1996) notes, both approaches led to remarkably similar results regarding the persistence of the deviations from the PPP and conludes that shocks hitting real exchange rates damp out at a pace of roughly 15 % on a yearly basis, putting the estimated half-lives of such shocks to three to five years.[7] Based on these findings, Rogoff argues that the empirical results indicating such high persistence are hard to reconcile with the high volatility of real exchange rates in the short term. Since such slow rates of mean reversion cannot be explained by theoretical models of nominal rigidities, he introduces the notion of the "PPP-puzzle".

With the extending availability of international price data at the disaggregated level, as well as the discouraging results in explaining Rogoff's puzzle based on aggregate price indices, the focus of recent research has shifted to the analysis of the PPP on the sectoral or even goods level.[8] Following this recent trend, my thesis tries to establish empirical evidence for the validity of the PPP. By combining the micro and macro aspects present

[4] As an interesting example for such studies, Froot, Kim and Rogoff (1995) analyze annual commodity data observed in England and Holland over 7 centuries.

[5] Amongst others, Taylor (2002), Abuaf and Jorion (1990) and Lothian and Taylor (1996) investigte the role of exchange rate regimes in the context of the PPP.

[6] See e.g. Frankel and Rose (1996).

[7] Other papers such as Froot and Rogoff (1995) or Taylor (2001) report similar "consensus estimates" of four to five years.

[8] See e.g. Crucini and Shintani (2008); Goldberg and Verboven (2005); Haskel and Wolf (2001); Parsley and Wei (2007).

in the literature I perform a comparative analysis on real exchange rates based on both aggregate and sectoral price indices.

My methodology closely builds on the recent study of Imbs et al. (2005*b*), which shows that aggregation bias has a substantial positive impact on the estimated price adjustment dynamics and once that bias is controlled for, persistence estimates obtained on disaggregated price level data indicate a faster pace of conversion than the "consensus view" established by Rogoff (1996).[9] During my analysis I also take into account the key points of the "PPP-debate" triggered by Imbs et al. (2005*b*)[10].

In addition, I extend Imbs et al.'s (2005*b*) analysis in the sense that I investigate the presence of such bias on three different levels of aggregation. Along these lines I show that although aggregation bias can result in positively biased estimates of the persistence of real exchange rates, this bias does not necessarily appear during the aggregation of sector-specific real exchange rates but can also evolve when neglecting country-specific heterogeneity on the aggregate level. Finally, I demonstrate that "intra-European" real exchange rates show a faster of conversion than those based on US-prices, therefore, choosing the US as the numeraire country might lead to overestimated half-lives for European countries.

The remainder of my thesis is organized as follows: in Chapter 2 I formally introduce the concept of the PPP and summarize the theoretical derivations of Imbs et al. (2005*b*) on the different factors of aggregation bias. Chapter 3 describes the most important

[9]The PPP-literature has introduced several forms of bias which can account for the high persistence of the PPP-deviations. See, for instance, Taylor, Peel and Sarno (2001) on the importance of non-linear model specifications or Taylor (2001) on temporal aggregation bias.

[10]Chen and Engel (2005) criticize the key findings of Imbs et al. (2002) along several dimensions, while Imbs et al. (2005*a*) provides additional arguments for the validity of their earlier results.

econometric methods used in recent empirical studies on the validity of the PPP. These tools include panel unit root tests, heterogeneous panel estimators, as well as methods correcting for cross-sectional correlation or small sample bias. In Chapter 4 I give a brief summary of the most important features of my data set. Chapter 5 describes the results of my empirical analysis, while in Chapter 6 I provide additional support for my findings by performing several robustness checks. Finally, Chapter 7 concludes.

Chapter 2

Theoretical Background

2.1 The Theory of PPP

According to the fundamental identity of the PPP, domestic and foreign prices should be equal, once they are accounted in the same currency. The basic intuition behind the theory is that in an environment of perfect competition and properly functioning markets, the possibility of arbitrage prevents sellers from demanding different prices for the very same product. Therefore, for any good i

$$P_i = SP_i^* \tag{2.1}$$

where P_i is the domestic price of good i, P_i^* represents its foreign counterpart while S is the nominal exchange rate (home price of a unit of foreign currency). On the single good level, the literature usually refers to this concept as the Law of One Price (LOP). Assuming that aggregate price measures are constructed based on the same weights across countries, Eq.

2.1 can be generalized to a basket of goods as well, arriving at the fundamental equation of the absolute PPP:

$$\sum_{i=1}^{N} \omega_i P_i = S \sum_{i=1}^{N} \omega_i^* P_i^* \tag{2.2}$$

where ω_i and ω_i^* are the weights applied during aggregation in the home and foreign country, respectively (again, with the implicit assumption that $\omega_i = \omega_i^* \,\forall\, i$).

In practice this identity holds only in the most extreme cases, since market imperfections, transportation- and other types of transaction costs or differences in preferences allow producers to set different prices even for the outmost identical goods. As an other important issue, even if the LOP holds on the goods level, the assumption of the application of identical weights for different countries is seldom true in practice. In most of the cases, price indices such as the CPI or WPI are constructed by governmental agencies without any international cooperation, since they are meant to measure the representative goods baskets of the respective country.

As a direct consequence, applied economic research seldom focuses on the validity of the absolute PPP but analyzes the question whether the observed differences are bounded in a way or prices move independently from each other. In order to allow for constant price differential between aggregate price indices, an alternative definition of the PPP was introduced: focusing on the changes in price levels and those in the nominal exchange rate, the relative form of the PPP states that

$$\Delta P_t = \Delta S_t \Delta P_t^* \tag{2.3}$$

where S_t denotes the period-t value of the exchange rate, while P_t and P_t^* represent the price index of the same basket of goods measured in the home and foreign country in period t, respectively. The relative PPP states that changes in the relative price levels should be offset by the movements of the nominal exchange rate, without requiring that the absolute form of the PPP holds.

2.2 Heterogeneity and Aggregation Bias

This section introduces the different forms of bias arising during the estimation of autoregressive time series models due to the heterogenity of the underlying disaggregated processes. Analyzing the characteristics of such models has substantial implications for empirical PPP-research as real exchange rates are usually assumed to follow AR(p) processes. The section summarizes the related findings of Imbs et al. (2005*b*), who associated the empirical failure of the PPP with the presence of such bias; however, in a pure econometric context this problem is thoroughly explored by Pesaran and Smith (1995).

2.2.1 Inconsistency of the Lambda-Type Estimators in Heterogeneous Dynamic Panels

Consider the disaggregated time series following first order autoregressive processes[1]:

$$y_{i,t} = \alpha_i + \delta_i y_{i,t-1} + \xi_{i,t}, \quad i = 1, ..., N, \quad t = 1, ..., T \tag{2.4}$$

[1] Pesaran and Smith (1995) use a more general framework allowing for contemporaneous exogenous variables as well; however, in the context of real exchange rates, these can be omitted without loss of generality.

where $\alpha_i = \alpha + \eta_i^\alpha$ and $\delta_i = \delta + \eta_i^\delta$, with η_i^α and η_i^δ having zero mean and constant covariance, $\delta_i \in (-1, 1)$ and $\xi_{i,t}$ being distributed $iid(0, \sigma_i^2)$. Performing a pooled estimation with restricting the persistence parameters to be identical across i yields a model as

$$y_{i,t} = \alpha_i + \delta y_{i,t-1} + \varepsilon_{i,t}, \quad \varepsilon_{i,t} = \eta_i^\delta y_{i,t-1} + \xi_{i,t} \tag{2.5}$$

As we can see, the lagged dependent variable now enters the error term, leading to inconsistent results of the usual OLS estimator. Moreover, the within or fixed effect estimator δ_W is also inconsistent, as in the full asymptotic case

$$\text{plim}_{N\to\infty, T\to\infty}\delta_W = \delta + \sum_{i=1}^{N}(\delta_i - \delta)\gamma_i \tag{2.6}$$

with

$$\gamma_i = \frac{\sigma_i^2}{1-\delta_i^2} / \left[\sum_{i=1}^{N}\left(\frac{\sigma_i^2}{1-\delta_i^2}\right)\right] \tag{2.7}$$

For the proof of the above statement I refer to Pesaran and Smith (1995). Imbs et al. (2005*b*) show that for large N, the asymptotic bias of δ_W is "positive *if and only if* $cov(\tilde{\delta}, \tilde{\alpha}) > 0$, i.e. the covariance between the vector of persistence parameters $\tilde{\delta} = \{\delta_i\}_{i=1}^N$ and the vector of coefficients $\tilde{\gamma} = \{\gamma_i\}_{i=1}^N$ is positive" (Imbs et al., 2005*b*, p. 4). Under these conditions, more persistent disaggregated time series have larger weights when summing up $\sum_{i=1}^N(\delta_i - \delta)\gamma_i$, resulting in a positive bias of δ_W.

2.2.2 Aggregation Bias Due to Cross-Correlated Error Terms

The second type of bias Imbs et al. (2005*b*) introduce arises due to the neglected cross-sectional depence structure of the error terms, hence I refer to this form as "cross-dependence bias". They derive the characteristics of such bias starting from the same model specification as described by Eq. 2.2.1; however, the assumptions regarding the distribution of the error terms are relaxed in the sense that they are allowed to be correlated across the cross section: $E(\xi_{i,t}, \xi_{j,t}) = \sigma_{i,j}$ for $i \neq j$. Aggregating the time series across i then leads to

$$\bar{y}_t = \bar{\alpha} + \delta \bar{y}_{t-1} + \bar{\xi}_t \tag{2.8}$$

where $\bar{y}_t = \sum_{i=1}^{N} \omega_i y_{i,t}$, $\bar{\alpha} = \sum_{i=1}^{N} \omega_i \alpha_i$, $\bar{\xi}_t = \sum_{i=1}^{N} \omega_i \xi_t + \sum_{i=1}^{N} \omega_i \eta_i^{\delta} y_{i,t-1}$, and ω_i are just the weights associated with the i-th cross sectional unit used for aggregation. Following the same intuition as in Subsection 2.2.1, one can show that the OLS estimator of δ is inconsistent with an asymptotic bias defined as

$$\text{plim}_{N\to\infty, T\to\infty} \delta_{OLS} = \delta + \sum_{i=1}^{N} (\delta_i - \delta)\kappa_i \tag{2.9}$$

with

$$\kappa_i = \left[\frac{\sigma_i^2 \omega_i^2}{1-\delta_i^2} + \sum_{i\neq j}^{N} \frac{\sigma_{i,j}\omega_i\omega_j}{1-\delta_i\delta_j} \right] \Big/ \left[\sum_{i=1}^{N} \left(\frac{\sigma_i^2 \omega_i^2}{1-\delta_i^2} + \sum_{i\neq j}^{N} \frac{\sigma_{i,j}\omega_i\omega_j}{1-\delta_i\delta_j} \right) \right] \tag{2.10}$$

Again, Imbs et al. show that the sign of the bias is the same as the sign of the covariance between the vector of persistence parameters and the vector of coefficients $\tilde{\delta} = \{\delta_i\}_{i=1}^{N}$ and the vector of coefficients $\tilde{\kappa} = \{\kappa_i\}_{i=1}^{N}$.

In the empirical part of my thesis I calculate the previously derived correlation conditions for my data to find support for the presence of aggregation bias arising from either of the two described sources. Section 5.1 describes the results of this analysis.

Chapter 3

Econometric Issues

3.1 Panel Unit Root Tests

Applied researchers generally test the validity of the PPP by analyzing the time series properties of the (logged) real exchange rates defined as

$$q_{it} = s_t - p_{it} + p^*_{it} \tag{3.1}$$

An increase in the real exchange rate can be interpreted as the depreciation of the domestic country's currency in real terms, thereby leading to a gain in competitiveness compared to the foreign country.

Investigating the presence of unit root processes in the real exchange rates had substantial economic implications for the PPP hypothesis: if PPP holds, the real exchange rate will revert to its long-run equilibrium value given by PPP after hit by shocks; if, on the other hand, real exchange rates seem to be driven by integrated processes, then the PPP serves no purpose in practice.

From a practical point of view, verifying the stationarity of the real exchange rates has posed a significant challenge to applied researchers: even though coefficients of the fitted autoregressive processes are estimated to be smaller than one, still, real exchange rates show high persistence, so that unit root tests could rarely reject the null of an integrated underlying process.

In order to deal with the low power[1] of these tests, some researchers[2] suggested using longer time series; however, those are difficult to obtain and are likely to be distorted by structural changes (e.g. exchange rate regime changes). As an alternative way to improve power, the cross-sectional dimension of the investigated samples was extended.[3]

Using panel data to investigate the PPP has several advantages: as a commonly cited improvement over time series data, a data set with multiple dimensions is less likely to suffer from multicollinearity; moreover, panel data sets can more easily overcome the issue of structural changes in the data given that the time dimension is shorter. Therefore, the following subsections give a brief introduction of the most important panel unit root tests while the results of these tests applied to our data are discussed in Section 4.2.

[1] The power of a test is defined as the probability of rejecting the null hypothesis when it is indeed false.

[2] See e.g. Taylor (2002).

[3] For an early study of this kind see Frankel and Rose (1996) while some of the more recent studies include e.g. Crucini and Shintani (2008); Goldberg and Verboven (2005).

3.1.1 The LLC-test

One of the first panel unit root tests was developed by Levin, Lin and James Chu (2002) (henceforth LLC-test). In their most general framework, the stochastic process $\{y_{it}\}$ observed over T time periods for a panel of N individuals is assumed to be generated by the following model:

$$\Delta y_{i,t} = \alpha_{0,i} + \alpha_{1,i}t + \rho y_{i,t-1} + \sum_{L=1}^{P_i} \theta_{i,L}\Delta y_{i,t-L} + \xi_{i,t} \tag{3.2}$$

where $-2 < \rho \leq 0$ for $i = 1, ..., N$ and the error terms $\xi_{i,t}$ are assumed to be distributed independently in the cross-sectional dimension and follow an invertible ARMA process: $\xi_{i,t} = \sum_{j=1}^{\infty} \theta_{i,j}\xi_{i,t-j} + \varepsilon_{i,t}$. As we can see, the LLC-test is a direct extension of a univariate ADF-test[4] into panel data framework. The test potentially allows for an intercept and time trend in the regression, a different order of serial correlation in the error terms; however, it restricts the slope coefficient to be identical across cross-sections so that under the alternative hypothesis, the speed of convergence is the same for all time series in the sample. As a result, testing the null hypothesis of unit roots in all series against the alternative of the stationarity of all time series is equivalent to testing $H_0 : \rho = 0, H_1 : \rho < 0$.

Levin, Lin and James Chu (2002) introduce a three-step procedure to carry out the test. In the first step, separate ADF-regressions as 3.2 are estimated for each cross-section, in order to obtain the individual-specific lag orders P_i. After observing P_i for each N, orthogalized residuals are calculated by two auxiliary regressions:

[4] Dickey and Fuller (1979)

$$\hat{e}_{i,t} = \Delta y_{i,t} - \hat{\alpha}_{0,i} - \hat{\alpha}_{1,i} t - \sum_{L=1}^{P_i} \hat{\pi}_{i,L} \Delta y_{i,t-L} \tag{3.3a}$$

$$\hat{v}_{i,t-1} = y_{i,t-1} - \tilde{\alpha}_{0,i} - \tilde{\alpha}_{1,i} t - \sum_{L=1}^{P_i} \tilde{\pi}_{i,L} \Delta y_{i,t-L} \tag{3.3b}$$

For each cross sectional unit, the residuals are then normalized by the standard errors of the residuals from Equation 3.2, $\hat{\sigma}_{\varepsilon,i}$:

$$e^*_{i,t} = \hat{e}_{i,t}/\hat{\sigma}_{\varepsilon,i}, \quad v^*_{i,t-1} = \hat{v}_{i,t-1}/\hat{\sigma}_{\varepsilon,i} \tag{3.4}$$

In the second step of the test, the ratio of the short and long run variance of the model is estimated:

$$\hat{S}_N = N^{-1} \sum_{i=1}^{N} \sum_{L=1}^{P_i} |1 - \hat{\theta}_{i,L}| \tag{3.5}$$

In the third step of the procedure the slope parameter $\delta = \rho + 1$ is estimated by regressing $e^*_{i,t}$ on $v^*_{i,t}$:

$$e^*_{i,t} = \delta v^*_{i,t-1} + u_{i,t} \tag{3.6}$$

After performing the three steps, the final test statistic is an adjusted t-statistic of δ taking the form

$$t^*_\delta = \frac{t_\delta - NT\hat{S}_N \hat{\sigma}_\delta}{\hat{\sigma}^2_{\varepsilon,i}} \times \frac{\mu_T}{\sigma_T} \tag{3.7}$$

where the adjustment term μ_T/σ_T (which may vary for different model specifications depending on the inclusion of a constant and/or a time trend in Equation 3.2) is provided by of Levin, Lin and James Chu (2002). In terms of asymptotic behavior, depending on the model setup, asymptotic normality of the test statistics holds as long as the cross-sectional dimension increases at a lower pace than the time dimension (more specifically, either $\sqrt{N}/T \rightarrow 0$ or $N/T \rightarrow 0$ as $N, T \rightarrow \infty$). As Levin, Lin and James Chu (2002) show, the derived test statistic has reasonable size properties for small samples ($N = 10$, $T = 25$), while for sufficient power larger samples ($N = 25$, $T = 100$) are required. Nevertheless, the implementation of the test depends upon the independence assumption across the cross-sectional units so that the test might provide biased results if cross-sectional correlation is present.

3.1.2 The IPS-test

Although the previously introduced LLC-test is fairly general in the sense that it allows for a different order of serial correlation across series, imposing the same first order autoregressive coefficient on the individual time series might be too restrictive. The approach developed by Im, Pesaran and Shin (2003) (henceforth the IPS-test) relaxes the latter assumption, as it allows both the constant and the autoregressive term to vary across individuals. Formally, their model takes the following form similarly to the initial setup of the LLC-test:

$$\Delta y_{i,t} = \alpha_{0,i} + \alpha_{1,i}t + \rho_i y_{i,t-1} + \sum_{L=1}^{P_i} \theta_{i,L} \Delta y_{i,t-L} + \xi_{i,t}, \quad i = 1, ..., N, t = 1, ..., T \tag{3.8}$$

The approach pools N separate ADF unit root tests (one for each panel member) to evaluate the presence of a unit root under the $H_0 : \rho_i = \rho = 0 \,\forall\, i$ against the alternative $H_1 : \rho_i < 0$ for $i = 1, ..., N_1$ and $\rho_i = 0$ for $i = N_1 + 1, ..., N$. As we can see, under the alternative, the model allows the rate of convergence to differ across individuals and does not require all time series to be stationary. The test statistic itself is the average of the t-ratios corresponding to the $\hat{\rho}_i$-s of the individual ADF-regressions:

$$\bar{t}_{NT} = N^{-1} \sum_{i=1}^{N} t_{i,T} = N^{-1} \sum_{i=1}^{N} \frac{\hat{\rho}_{i,T}}{\sqrt{Var(\hat{\rho}_{i,T})}} \tag{3.9}$$

The authors also address the issue of cross-sectional correlation and suggest to subtract the cross-sectional means from the observed data to mitigate its effect.

The proper asymptotic behavior of the test requires $T \to \infty$ followed by $N \to \infty$. Using Monte-Carlo simulations Im, Pesaran and Shin (2003) show that the proposed test performs well with both N and T sufficiently large (the largest sample they consider is $N = 100$ and $T = 50$) if the lag orders in the individual ADF-regressions (P_i) are "large enough". According to the simulation results, the finite sample performance of the IPS-test dominates that of the LLC-test.

3.1.3 Pesaran's (2007) CADF Test

The first generation panel unit root tests described in the previous subsections relied on the rather restrictive assumption of cross-sectional independence among the panel members. As Breitung and Das (2005) notes, ignoring the correlation across panel members can lead to over-rejection of the null hypothesis in the previously described tests, thus falsely concluding stationarity of integrated processes. In order to relax this constraint, these early tests were usually augmented by de-meaning the time series prior to carrying out the actual unit root tests. This correction, however, can only account for the cross-correlatedness of the error terms when all non-diagonal elements of the variance-covariance matrix turn out to be equal.

Pesaran (2007) introduced an alternative approach making use of a residual one-factor model. Consider a heterogeneous dynamic panel model expressed in a Dickey-Fuller regression form:

$$\Delta y_{i,t} = \alpha_{0,i} + \rho_i y_{i,t-1} + \xi_{i,t}, \quad i = 1, ..., N, t = 1, ..., T \tag{3.10}$$

where $\rho_i = \delta_i - 1$ and $\xi_{i,t}$ denotes a composite error structure compiled of an unobserved common common component f_t and an idiosyncratic term $\varepsilon_{i,t}$: $\xi_{i,t} = \gamma_i f_t + \varepsilon_{i,t}$ with both having zero mean, constant variance, finite fourth order moments and being independently distributed along all possible dimensions.[5] As we can see, through γ_i the setup allows for heterogeneous effects of the shared common factor on the individual time series. Similarly to the IPS-test, the null hypothesis of the proposed test is $H_0 : \rho_i = 0 \, \forall \, i$ against the alternative $H_1 : \rho_i < 0$ for $i = 1, ..., N_1$ and $\rho_i = 0$ for $i = N_1 + 1, ..., N$.

In practice, the stationarity of such a model can be evaluated by estimating the regression

$$y_{i,t} = a_i + b_i y_{i,t-1} + c_i \bar{y}_{t-1} + \Delta \bar{y}_t + e_{i,t} \tag{3.11}$$

As we can see, the unobserved common factors are proxied by the cross-sectional averages of both the dependent and the independent variables of the model. Due to the presence of these terms, Pesaran (2007) denoted this form as the Cross-sectionally Augmented Dickey-Fuller (CADF) regression. He proposes the use of either the $\bar{t}_{N,T}$ statistic of Im, Pesaran and Shin (2003) (as defined in 3.9) or $Z(N, T)$, its equivalent assuming normal distribution:

$$CIPS(N, T) = \bar{t}_{N,T} = N^{-1} \sum_{i=1}^{N} t_i(N, T), \quad Z(N, T) = N^{-0.5} \sum_{i=1}^{N} \Phi^{-1}(p_{i,T}) \tag{3.12}$$

[5] In the current setup the model does not allow for any sort of serial correlation within the compound error term; however, Pesaran (2007) also derives a modified version of the proposed test statistic, allowing for serial correlation.

where $p_{i,T}$ is the p-value of the Dickey-Fuller test carried out on the cross-sectional unit i in the form defined by Equation 3.11. It is clear from Equations 3.11 and 3.12 that a great advantage of the CADF-test relies in its simple methodology as it is a direct extension of the usual Dickey-Fuller test. Moreover, the asymptotic properties of the test are quite suitable to our empirical analysis, as they are derived for $N/T \to k$ while $N \to \infty, T \to \infty$. As Pesaran's (2007) Monte-Carlo simulations show, the test has satisfactory power and size if both dimensions of the panel are larger than 30.

3.2 Panel Estimators

3.2.1 Standard Panel Estimators

Consider a standard AR(1) model in a three dimensional panel context of the following form:

$$y_{i,c,t} = \alpha_{i,c} + \rho y_{i,c,t-1} + \varepsilon_{i,c,t} \tag{3.13}$$

where i and c denote the two cross-sectional dimensions of the panel while t is the time index and $\varepsilon_{i,c,t} \sim (0, \sigma^2_{i,c})$ is assumed as usual. In a dynamic panel context, due to the presence of constant sector-specific effects, the explanatory variables are necessary correlated with the unobserved panel-level effects of the error term, resulting in biased estimates of the lambda-type estimators. To tackle this problem, Equation 3.13 can be rewritten in first differences, thereby eliminating the time-fixed error components. Nevertheless, the lagged dependent variable remains endogenous in the differenced form. There have been several suggestions in the literature to overcome this problem: as $E(y_{i,c,t-p}, \varepsilon_{i,c,t}) = 0$ for

$p > 1$, Arellano and Bond (1991) derived a consistent generalized method of moments (GMM) estimator for this context by using all available lags of $y_{i,c,t-1}$ as instruments for the differenced lagged dependent variable.

As a drawback of this estimator, it has been shown that Arellano-Bond estimator could perform poorly if the true persistence parameter is close to unity. As an alternative GMM-type estimator, Blundell and Bond (1998) proposed to estimate the level and the first differenced equations jointly, using the lagged levels as instruments in the differenced equation and lagged differences as instruments in the level equation, respectively. By enabling for an extended set of moment conditions, the Blundell-Bond sys-GMM estimator is expected to have better properties in case of highly persistent processes.

3.2.2 Panel Estimators Allowing for Heterogeneous Slopes

By extending 3.13, consider a panel model with heterogeneous coefficients:

$$y_{i,c,t} = \alpha_{i,c} + \rho_{i,c} y_{i,c,t-1} + \varepsilon_{i,c,t}, \quad \alpha_{i,c} = \alpha + \eta^{\alpha}_{i,c}, \quad \rho_{i,c} = \rho + \eta^{\rho}_{i,c} \tag{3.14}$$

Standard panel estimators rely on the assumption that the coefficients of the model are identical across the cross-sectional groups ($\rho_{i,c} = \rho \; \forall \; i, c$). One can think of several economic problems when this assumption is misleading. Suppose instead that the coefficients may be different across sectors and one is interested in the average of these heterogeneous effects. As Pesaran and Smith (1995) note, in the static panel context unbiased coefficient means can be obtained by either aggregating, pooling, averaging group estimates or cross-sectional regressions. In a dynamic panel setup, however, the first two approaches lead to inconsistent estimates if the regressors are serially correlated. As a solution to this problem, they propose the application the Mean Group (MG) model, where the estimator can be obtained by performing standard OLS-regressions on each group of the panel and then taking the average of these estimates:

$$\hat{\rho}^{MG} = I^{-1}C^{-1}\sum_{i=1}^{I}\sum_{c=1}^{C}\hat{\rho}_{i,c}^{OLS} \tag{3.15}$$

where I and C denote the number of groups across cross-sections in our previous model setup. Pesaran and Smith (1995) show that MG-estimator provides consistent estimations for the coefficient means for large N and T.[6]

The Mean Group model assumes that heterogeneity across groups is deterministic in the sense that $\eta_{i,c}^{\alpha}$ and $\eta_{i,c}^{\rho}$ are fixed effects in the model. Alternatively, the group-specific deviations of the average coefficients could be considered as random variables with zero mean amd constant covariances $(E(\eta_{i,c}^{j}) = 0, E(\eta_{i,c}^{j}, \eta_{i,c}^{j'}) = \Gamma)$ and finite moments of higher order). In such a setup, the implementation of the Random Coefficients (RC) model first proposed by Swamy (1970) is recommended. While the MG-estimator simply takes the unweighted arithmetic average of the group-specific coefficients, the RC model performs a GLS regression and so aggregates the group-specific estimates by optimally weighting them, using the variance-covariance matrix of the estimated residuals.

To be more explicit, consider Equation 3.14 in matrix notation:

$$y_{i,c} = X_{i,c}\beta_{i,c} + \varepsilon_{i,c} \tag{3.16}$$

where

[6]The paper does not provide Monte-Carlo evidence for which sample sizes the estimator can be considered as consistent.

$$y_{i,c} = \begin{pmatrix} y_{i,c,1} \\ \vdots \\ y_{i,c,T} \end{pmatrix}, \quad \beta_{i,c} = (\alpha_{i,c}, \rho_{i,c}), \quad X_{i,c} = \begin{pmatrix} 1 & y_{i,c,0} \\ \vdots & \vdots \\ 1 & y_{i,c,T-1} \end{pmatrix}, \quad \varepsilon_{i,c} = \begin{pmatrix} \varepsilon_{i,c,1} \\ \vdots \\ \varepsilon_{i,c,T} \end{pmatrix} \tag{3.17}$$

Then, the MG-estimator $\hat{\rho}^{MG}$ is obtained by vertically averaging the second column of matrix $\hat{\beta}_{i,c}$ while the RC-estimator $\hat{\rho}^{RC}$ can be calculated as

$$\hat{\rho}^{RC} = \sum_{i,c} W_{i,c} \hat{\rho}_{i,c}^{OLS} \tag{3.18}$$

with

$$W_{i,c} = \Big[\sum_{i,c} (\Gamma + V_{i,c})^{-1} \Big] (\Gamma + V_{i,c})^{-1}, \quad V_{i,c} = \sigma_{i,c}^2 (X'_{i,c} X_{i,c})^{-1} \tag{3.19}$$

3.3 Other Panel-Related Econometric Issues

3.3.1 Cross-Sectional Dependence

The assumption of cross-sectional independence across panel members can turn out to be quite unrealistic, especially when the cross-sectional dimension of the panel is large. Such cross-dependencies can arise due to a variety of reasons such as spatial dependence, supranational institutional connections or other sources of common shocks which is not properly accounted for in the model. Addressing the problem of cross-sectional dependence has two advantages in our case: first and foremost, aggregation bias caused by cross-correlated error terms (as presented in Section 2.2.2) crucially relies on the assumption of cross-sectional correlation in the residuals, hence the issue should be included in the empirical part of this analysis too. Secondly, incorporating the dependence structure into the estimation process can potentially improve the efficiency of the estimators, thereby leading to more precise results.

Before turning to the methods of addressing this issue, first I introduce two approaches for testing cross-sectional dependence in the error terms. The test proposed by Breusch and Pagan (1980), which evaluates the null hypothesis of absence of cross-correlatedness, computes an LM-statistic of the form

$$BP = T \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \hat{\sigma}_{i,j}^2 \tag{3.20}$$

where N and T denote the cross-sectional and time dimensions of the panel, respectively while $\hat{\sigma}_{i,j}^2$ is the estimated correlation between the residuals of series i and j. Under the H_0, the test statistic follows a χ^2 distribution with $N(N-1)/2$ degrees of freedom. The test is generally suitable for small-N large-T panels and can exhibit substantial size distortions in panels with large cross-sectional dimensions.

By the properties of our sample, the Breusch-Pagan test is only applicable in the case of the headline price indices as the sample of the sectoral real exchange rates includes hundreds of observations in both dimensions (see Section 4.1 for details on our data set). Therefore we introduce an alternative test which is not constrained to panels with small cross-sectional dimensions. The test was proposed by Pesaran (2004) and relies on the following test statistic:

$$CD = \sqrt{\frac{2T}{N(N-1)}} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \hat{\sigma}_{i,j} \tag{3.21}$$

where N, T and $\hat{\sigma}_{i,j}^2$ are the same as in the case of the Breusch-Pagan test. Under H_0 the test statistic asymptotically follows a standard normal distribution. As Pesaran (2004) shows, the test shows favorable properties in terms of power even in the case of large-N panels, moreover, it is valid in the case of highly persistent heterogeneous dynamic panel models too.

We now turn to the approaches which allow the standard estimators to account for cross-dependence. Among the previously discussed estimation procedures only the Random Coefficients model exploits the variance-covariance matrix of the residuals whereas the Mean Group model as well as the standard Least Squares estimators generally assume the diagonality of this matrix and thus, independence across cross-sections. Therefore we now discuss two procedures which can extend these estimators to the cross-correlated context.

The most widely used method to deal with cross-dependence is to treat the cross-sectional units as a system of equations and perform Seemingly Unrelated Regression Estimations (SURE) based on GLS-techniques. The approach largely builds on the work of Zellner[7]. As in the case of all GLS-type estimators, the variance-covariance matrix of the error terms is estimated with help of the residuals of separate OLS-regressions and then the final estimators are obtained by correcting the OLS-estimates with the matrix, in a similar manner as shown in Equations 3.18 and 3.19.

The SURE approach is an acknowledged and widely used method to deal with cross-sectional correlation in panel context; has, however, one significant drawback: it can only be applied to panels whose cross-sectional dimension is smaller than the time dimension. Since in the empirical part of the thesis this is not always the case, the "Common Correlated Effects" method proposed by Pesaran (2006) will also be implemented. Intuitively,

[7] Zellner (1962); Zellner (1963); Zellner and Huang (1962).

the approach filters the individual-specific regressors from their cross-sectional averages, thereby accounting for the effects of common unobserved factors.[8] It allows for heterogeneous effects of these common error components on the cross-sectional units as well as for potential correlation among themselves or with the individual-specific errors. Compared to SURE, apart from its applicability to large panels, the CCE correction has a further advantage of simplicity as in practice it only requires to calculate the cross-sectional averages for both the dependent and explanatory variables and then include them as auxiliary regressors to the model.

3.3.2 Small Sample Bias

It is a well documented fact in the literature that autoregressive processes with high persistence can lead to underestimation of the coefficients in the case of the least squares estimators (see for instance Andrews (1993)), thereby reporting shorter half-lives for deviations from the PPP. Several procedures have been proposed to address this problem: Andrews (1993) suggests the use of the median unbiased estimator in the case when the distribution of the innovations of the AR(1) model as well as the median function of the slope coefficient are known. Alternatively, Kilian's (1998) "Bootstrap-after-Bootstrap" method does not require such prior knowledge of the theoretical distributions; however, it requires substantial computation capacities. A second technique relying on bootstrapping, called the "grid bootstrapping" method, has been proposed by Hansen (1999).

[8]The intuition behind the test is the same as in the case of the CADF-test.

In this subsection we discuss the method introduced by So and Shin (1999). They claim that in the case of positively correlated autoregressive processes the small sample bias of the least-squares estimator family is "due in large part to lack of efficiency in mean adjustment or trend adjustment". Therefore they propose that instead of the usual mean adjustment $\tilde{y}_t = y_t - T^{-1}\sum_{t=1}^{T} y_t$ performed during the LS-type estimations a recursive mean adjustment should be applied following the formula $\ddot{y}_t = y_t - t^{-1}\sum_{i=1}^{t} y_i$. In the case of a simple AR(1) model[9] the modified OLS-estimator takes the form

$$\hat{\rho}_{RMA}^{OLS} = \frac{\sum_{t=2}^{T}(y_{t-1} - \ddot{y}_{t-1})(y_t - \ddot{y}_{t-1})}{\sum_{t=2}^{T}(y_{t-1} - \ddot{y}_{t-1})^2} \tag{3.22}$$

As we can see, So and Shin's (1999) method does not require prior knowledge of the underlying process or intense computations and is simple to apply, since it requires to replace the overall sample mean with the recursive sample mean.

[9] $y_t = \alpha + \rho y_{t-1} + \varepsilon_t$. with $\varepsilon_t \sim iid(0, \sigma_\varepsilon^2)$

Chapter 4

Data

4.1 Description of the Data Set

Our data set consists of monthly observations of the harmonized index of consumer prices (HICP) for each current member state of the European Union compiled by the Eurostat (Eurostat, 1996-2013*c*). As the most important feature of this data set, price indices are not only reported for the general consumption basket of goods and services but at 4 different levels of aggregation – enabling us to examine the different forms of aggregation bias introduced in Section 2.2 at three levels. As the second ingredient of the real exchange rates, nominal exchange rates were also obtained from the statistical agency of the European Union (Eurostat (1996-2013*a*), Eurostat (1996-2013*b*)). Since observations for the goods and services are gathered by the national statistical agencies throughout each month continuously, the nominal exchange rates correspond to monthly averages of the daily official fixings instead of their month-end balances. Finally, for the abandoned currencies of the current EMU-members a data smoothing procedure was performed by Eurostat in order to account for the one-time jumps of the national exchamge rates at their entrance to the Euro-zone.

Turning to the structure of my data set, the 1^{st} level of the COICOP-classification contains only one price index, the headline inflation measure. On the 2^{nd} level of the COICOP-structure one can find the 12 most important subindices (e.g. "Food and non-alcoholic beverages"). Going one step further to the 3^{rd} level, the number of indices jumps to 40, while the most detailed 4^{th} level consists of 78 series, adding up to 131 price indices in total (examples for the last two levels are "Food" and "Bread and cereals", respectively). However, at the last two levels of the classification some price indices are not necessarily compiled for each country (therefore time series which do not appear among the components of the headline index of a respective country were dropped of the sample. Out of the potential 3,406 time series (131 price indices for 26 countries), 117 did not contain any observations, resulting in a cross-sectional dimension of $N = 3,289$.

The Eurostat allows access to HICP-data starting from January 1996 to the present times; however, we restricted my sample to the 1996:01–2012:12 period ($T_{max} = 204$). It has to be noted that the national statistical agencies of the new member states (NMS) started to comply with Eurostat's data gathering standards only during the accession process to the EU; therefore, most time series for the NMS for the 3^{rd} and 4^{th} COICOP-levels only start in January 2001, resulting in a highly unbalanced panel for the first years of my sample. Nevertheless, the last 12 years of the sample can be regarded as a strongly balanced panel.

After removing the observations which do not enter the aggregation process of the Eurostat, we arrive at 624,382 observations for our whole sample. Table A.1 in Appendix A summarizes the data coverage of our panel. As we can see, regarding the subsamples for the first two COICOP-levels, our data coverage is almost complete, with some observations missing for Bulgaria and Belgium for the first time periods. For levels 3 and 4 the coverage is somewhat reduced (especially for Romania, Hungary and Slovenia); nevertheless, out of the maximal 204 the sample includes 186 and 178 data points on average for the included price indices, respectively. Overall, the sample consists of 183 observations on average, with most missing values coming from the NMS for the pre-2001 period.

Using such a detailed data set has numerous advantages: most importantly, the PPP theory requires the underlying goods to be identical - although this is hard to implement in practice, extending our analysis to such detailed categories increases the credibility of our assumption regarding the homogeneity of the underlying goods. Moreover, using the Eurostat database as opposed to the usual consumer price indices compiled by the statistical authorities of the member states overcomes the issue of aggregation bias caused by applying different weights during the calculations. Lastly, extending the cross-sectional dimension of my sample should also improve the robustness of our results.

4.2 Stationarity of the Real Exchange Rates

The calculation of real exchange rates requires a reference country whose price indices serve as a basis of comparison – in our analysis the benchmark country is Germany. This choice was motivated by several factors: firstly, Germany is the biggest economy of the

European Union – since the PPP literature generally relies on a big advanced economy, the choice is in line with the usual empirical approach.[1] Secondly, since deviations from the PPP are probably less important in the case when there is limited economic activity between the countries, choosing the most important partner in terms of intra-EU trade increases the robustness of our results. Finally, this choice was supported from a practical point too: Germany has basically full data coverage in the sample, thus enabling us to maximize the number of investigated real exchange rates. Nevertheless, it does have the drawback of being part of the Euro-zone, thereby resulting in fixed nominal exchange rates for several countries and time periods. Therefore, as a robustness check I carried out the analysis with choosing the United Kingdom as a benchmark country as well.

Following the methodology of Imbs et al. (2005*b*), I performed the LLC- and IPS-tests on both the aggregate and the sectoral real exchange rates of our panel, with and without a deterministic trend. Since these tests belong to the family of first generation panel unit root tests and thus have a tendency of over-rejecting the true null hypothesis of a unit root in the case of cross-correlated panels, as a robustness check we also applied Pesaran's (2007) CADF-test (which allows for cross-dependence) to our data.[2]

Table 4.1 summarizes the results of the panel unit root tests. As we can see, the tests overwhelmingly reject the null hypothesis of the presence of a unit root for each COICOP-level, irrespective of the a priori assumptions of the test. As an interesting side experiment we can also conclude that the favorable qualitative results of the first generation tests

[1] Empirical PPP-studies generally select the US as the reference country; however, data in the COICOP-classification is not available for levels 3 and 4 for the US, hence choosing the latter as the reference economy would have led to a significant reduction of my sample.

[2] See Section 3.1 for the properties of the tests.

TABLE 4.1: Results of the panel unit root tests.

Test	Trend	COICOP-level 1	2	3	4
LLC	No	-9.8591	-21.1925	-17.8114	-19.088
		(0.0000)	(0.0000)	(0.0000)	(0.0000)
LLC	Yes	-10.0989	-8.7683	-12.205	-15.9453
		(0.0000)	(0.0000)	(0.0000)	(0.0000)
IPS	No	-4.6021	-10.0041	-9.7364	-14.2668
		(0.0000)	(0.0000)	(0.0000)	(0.0000)
IPS	Yes	-13.7814	-13.8077	-21.6201	-34.5233
		(0.0000)	(0.0000)	(0.0000)	(0.0000)
CADF	No	-4.6021	-10.0041	-9.7364	-14.2668
		(0.0000)	(0.0000)	(0.0000)	(0.0000)
CADF	Yes	-13.7814	-13.8077	-21.6201	-34.5233
		(0.0000)	(0.0000)	(0.0000)	(0.0000)

The table includes the test statistics and p-values (in parenthesis) for the respective panel unit root tests. All regressions include a constant. The LLC-test restricts the slope parameters to be indentical under the H_1, while the IPS- and the CADF-tests allow for heterogeneous dynamics across the panel members. The first two tests allow for a serial correlation up to 12 lags in the residuals, where the optimal lag length was selected based on SIC for each panel member. For the CADF-test, the table report statistics of tests factoring for first order serial correlation for each cross-sectional unit; however, tests including up to six lags did not produce qualitatively different results. The CADF-tests were carried out using the Stata package of Lewandowski (2006).

were not biased by cross-sectional correlation, a finding similar to the empirical results of Breitung and Das (2005).

Chapter 5

Results

After the description of the theoretical background as well as the methodological issues related to my analysis I summarize the empirical findings of my research. First, I investigate the presence of aggregation bias based on the formulae derived by Imbs et al. (2005*b*), then I present the regression estimates. Later, in Chapter 6 I run additional regressions in different setups to support the robustness of my results.

5.1 Presence of Heterogeneity Bias

In Section 2.2 I provided a short summary on how heterogeneity at a sectoral level can lead to biased estimates of the persistence parameters on the aggregate level, as introduced by Imbs et al. (2005*b*). I distinguished to forms of bias: the first one, denoted as 'heterogeneity bias' arises if one tries to fit autoregressive models with identical slope parameters to a panel whose cross-sectional members are in reality driven by different rates of adjustment. The second type of bias, called 'cross-dependence bias' arises when cross-sectional correlation across the panel members is not taken account during the estimation process. As demonstrated earlier, under certain conditions the sign of both types of bias should be positive resulting in the overestimation of the true autoregressive parameters.

In this section I calculate the bias coefficient Imbs et al. (2005*b*) derived for the case of heterogeneity bias. In order to do so, I fit separate AR(1) regressions using the OLS-estimator to each country and sector of our sample and record the estimated slope coefficients as well as the variance of the residuals. I then calculate the bias coefficient using the formulas defined in Equation 2.7.

Figure A.1 in Appendix A summarizes the results by plotting the obtained bias coefficients against the slope estimates whereas Table A.2 gives a detailed overview on the calculations performed on the aggregate level. The figures indicate a small but positive correlation between the two coefficients of interest. This is confirmed analytically as well: for the whole sample, the estimated correlation coefficients were 0.1360, 0.0304, 0.0142, 0.0253 for the COICOP levels 1-4, respectively. As a robustness check of our results I truncated the sample to the time series with estimated autoregressive coefficients between 0.8 and 1 (the right columns of Figure A.1 demonstrates our results for the truncated samples). After eliminating the outliers I was able to observe a substantial increase in the sample correlations, with the coefficients now being equal to 0.2691, 0.1272, 0.0984 and 0.0798, respectively.

Comparing our results to those of Imbs et al. (2005*b*), one can observe both similarities and differences: on the one hand, the presence of a positive heterogeneity bias found support in both data sets, emphasizing that allowing for cross-sectional heterogeneity could help us explain the large persistence parameters estimated by OLS or FE estimators. On the other hand; however, the support for the positive bias was less clear cut in our sample than in Imbs et al. (2005*b*) as the obtained correlation coefficients are only slightly positive. Unfortunately, these results cannot be compared directly to Imbs et al. (2005*b*) as they report covariances instead of correlations, thereby making any kind of comparison infeasible.

5.2 Regression Estimates

In this section I summarize the results of my empirical analysis. First, I report the results obtained by fitting standard OLS and Fixed Effects (FE) models to the four subsamples defined by each COCIOP-level of my data set. These techniques are restrictive in the sense that they assume identical slope coefficients among the different members of the panel[1]. In order to relax this assumption, I apply the Mean Group (MG) and Random Coefficients (RC) estimators described in Section 3.2.2, which allow for different adjustment dynamics across both countries and sectors. As the next step, these estimators are augmented to control for cross-sectional dependence within the panel by either the SURE approach (when applicable) or as in most of the cases, by Pesaran's (2006) CCE correction method.

The regression results presented in this Section are not interpreted in their original form but are subject of a simple conversion: as a widely used measure in the context of

[1] Besides the homogeneity of the persistence parameters, the OLS estimator also constrains the constants of the model to be the same across the cross sections, whereas the FE estimator allows for heterogeneous constants.

stationarity of the real exchange rates, I refer to the estimated coefficients in terms of half-lives which correspond to the number of periods necessary for a shock hitting the real exchange rate to lose half its initial effect. With regards to AR(1) processes, half-lives can be easily computed using the formula $HL = ln(0.5)/ln(\hat{\delta})$, where $\hat{\delta}$ is the estimated slope coefficient of the process. In the case of higher order autoregressive processes, half-lives are obtained based on the impulse response function (IRF) of the model and correspond to the first period after which the impulse response function remains below the 0.5 level.

Both Imbs et al. (2005*b*) and Chen and Engel (2005) point out that the standard specification in the empirical studies on the PPP is to fit AR(p) processes to the data, with most of the studies settling for $p = 1$. Out of the two studies, the former describes the process of obtaining the optimal p lag length so as to analyze the impulse response functions of the fitted AR(p) models and then selects the one for which the IRF is continuous around 0.5. However, this specification is fairly vague and thus hard to reproduce, therefore in my thesis I follow the most common specification of the literature and fit AR(1) models to my data. As a result, half-lives can be easily obtained by the formula described above without the rather subjective analysis of the impulse response functions.

5.2.1 Aggregate Real Exchange Rates

As the starting point of the analysis I estimated Equation 3.13 on the sample of aggregate real exchange rates by OLS, thereby restricting both the slope coefficients and constants to be identical across countries. As shown in the first row of Table 5.1, the OLS regression indicates a remarkably high level of persistence: the point estimate of the regression implies a half-life of 35 years for the shocks while the estimated confidence interval translates into half-lifes of 18-306 years. The results clearly indicate higher persistence than usually reported in the literature[2] and points most probably towards misspecification of the model.

TABLE 5.1: Estimated slope coefficients of the aggregate real exchange rates.

Model	Heterogeneity Constant	Heterogeneity Slope	Cross-dep.	$\hat{\delta}$	Half-life	Homogeneity test constant	Homogeneity test slope	BP-test[3]	CD-test[4]
OLS	No	No	No	0.998	414			6817.61	61.61
				[0.997, 1]	[219, 3667]			(0.00)	(0.00)
FE	Yes	No	No	0.967	21	241.9		7336.96	65.15
				[0.944, 0.99]	[12, 69]	(0.00)		(0.00)	(0.00)
Arellano-Bond	Yes	No	No	0.965	20			7452.94	66.16
				[0.929, 1.001]	[9, ∞]			(0.00)	(0.00)
Blundell-Bond	No	No	No	0.998	427			7407.18	64.55
				[0.995, 1.002]	[135, ∞]			(0.00)	(0.00)
MG	Yes	Yes	No	0.967	20	230.84	616.31	6861.62	63.34
				[0.951, 0.982]	[14, 39]	(0.00)	(0.00)	(0.00)	(0.00)
FE (SURE)	Yes	No	Yes	0.972	24	341.49		7788.82	63.97
				[0.968, 0.976]	[22, 28]	(0.00)		(0.00)	(0.00)
FE (CCE)	Yes	No	Yes	0.966	20	155.75		16758.75	72.62
				[0.946, 0.986]	[13, 48]	(0.00)		(0.00)	(0.00)
MG-CCE	Yes	Yes	Yes	0.905	7	225.67	314.68	5821.79	38.3
				[0.884, 0.926]	[6, 9]	(0.00)	(0.00)	(0.00)	(0.00)
RC	Yes	Yes	Yes	0.972	24			6858.72	62.89
				[0.945, 0.999]	[12, 702]			(0.00)	(0.00)

Notes: Figures in brackets denote standard errors in the case of the estimated slope coefficients and their implied confidence interval in half-life terms. Figures in parenthesis denote the corresponding p-values of the presented test statistics. Standard errors were obtained using on a non-parametric bootstrap procedure with 100 repetitions for the LS-type estimators. For GMM-estimators I report heteroskedasticity- and autocorrelation robust standard errors. [1] GMM-estimations were performed using the Stata package of Roodman (2003). [2] MG-estimations were performed using the Stata package of Eberhardt (2011). [3] Denotes the Breusch-Pagan test statistic. [4] Denotes Pesaran's (2004) CD-statistic.

By moving one step towards a more general model setup I estimated Equation 3.13 by the standard fixed effects (FE) panel estimator. As it is shown in the second row of Table 5.1, allowing for heterogeneous country-specific constants resulted in a significant decrease in the estimated half-lives: the FE-estimator report a shock persistence of 21 months in half-life terms; however, its large confidence interval (12 to 77 months) does not

[2]I consider Rogoff's (1996) consensus view of 3-5 years as the benchmark for my results.

allow do distinguish its results from the consensus view. As a test for the correct model specification I performed a Hausman-test to choose between the FE and the Random Effects setup: here the χ^2_1-statistic of 6.46 allowed to reject the null hypothesis of non-systematic differences between the FE and RE estimators at the 1% confidence level, thus the FE-specification is preferred to the RE.

As I point out is Section 3.2.1, FE estimators produce biased estimates in the dynamic panel context. Therefore I estimated Equation 3.13 by two standard GMM-estimators, those developed by Arellano and Bond (1991) and Blundell and Bond (1998). The Arellano-Bond estimator produceed quantitatively similar results to those of the FE; however, its confidence interval is significantly larger.[3] On the other hand, the Blundell-Bond estimates report a similar persistence as the OLS-estimator – although I implemented the Blundell-Bond estimator as it relies on a wider set of moment conditions than the Arellano-Bond estimator, the fact that in the level equation of the system a homogeneous constant is included makes the estimator rather comparable to the OLS than the FE.

As the last step of my analysis I allowed for heterogeneous dynamics and cross-correlated residuals across countries. I estimated the coefficients with two heterogeneous estimators, out of which the Mean Group does not allow for cross-dependence while the Random Coefficient model does. Therefore, former was also augmented by the CCE approach, as was the standard Fixed Effects estimator. Rows 5-10 of Table 5.1 report the results.

[3]This result is not surprising as GMM-estimators rely on less assumptions and thus smaller information set than the LS-type estimators, hence the larger confidence interval.

Applying either the SURE or the CCE correction method on the standard FE estimators did not change the previously obtained results substantially: the estimated half-lives of 24 and 22 months for FE-SURE and FE-CCE, respectively are now slightly higher than the uncorrected estimates with the estimated confidence intervals lying within those reported in the previous paragraphs. Based on the results obtained by homogeneous estimators I didn't find support for cross dependence bias on the aggregate level; however, the slope homogeneity tests reported in Table 5.1 indicate that the FE-setup might not be the correct model specification.

Turning now to the estimators allowing for heterogeneous adjustment dynamics, the MG estimator did not produce substantially different results from those obtained by the standard panel methods: the point estimate lies remarkably close to the FE and Arellano-Bond estimates (20-21 months in half-life terms in all 3 setups). The estimated confidence intervals are now substantially smaller with the upper bound (35 months) just matching the lowest bound of the consensus view.

In the case of estimators which allow both for heterogeneous slopes and cross-dependent error terms, the random coefficient model produced again similar results as the previous estimation techniques: its point estimate puts the half-lives of the shocks to 2 years while the estimated confidence interval is now rather wide as it ranges from 13 months to 15 years. On the other hand, the MG-CCE estimator gave substantially different results as it put the half-lives as low as 7 months with a quite narrow confidence interval of 6 to 9 months. As latter approach was able to control for the most cross-correlation in the residuals according to both the Breusch-Pagan and the Pesaran CD test statistics, its estimates should be preferred over the RC-model's.

The results presented in this section can be summarized as follows: restricting the country-specific constants to be identical (as in the case of the OLS- and sys-GMM estimators) resulted in half-lives substantially larger than Rogoff's (1996) consensus view. Allowing for heterogeneous constants put the point estimates in a fairly close interval of 20-24 months, which are somewhat below both Rogoff's (1996) consensus view and the results of Imbs et al. (2005*b*); nevertheless, the large confidence intervals do not allow to distinguish them from the 3-5 year range. More general model setups did not result in qualitatively different results; however, the residual diagnostics indicate that they were not able to account for cross-dependence. Being the only exception from the previous statement, the MG-CCE model showed somewhat smaller residual cross-correlation and reported lower half-life estimates as well. Therefore I conclude that on the aggregate level cross-dependence bias plays a more important role than heterogeneity bias.

5.2.2 Sectoral Real Exchange Rates

We now turn to the panel of disaggregated prices of the second level of the COICOP classification. Our sample now includes cross-sections for 12 basket of goods and services which are used by the Eurostat to construct the headline HICP figures we investigated in the previous subsection and, as before, the 26 member states of the European Union excluding Germany. The sample analyzed in this subsection is qualitatively different from the panel of headline price indices in the sense that now the cross-section itself can be split along two further two further dimensions, namely that of the price indices and that of the countries. Therefore, I first investigate the impact of heterogeneous slope dynamics across countries, then I only allow for different different constants and slopes across the different price indices but restrict them to be identical across countries. Finally, the two aspect are combined so that each price index in each county can potentially follow a unique adjustment path.

The analysis of the disaggregate real exchange rate relies on the econometric tools applied in the previous subsection: first I fit the model to the data with the usual fixed effects estimator – again, this estimator neglects dynamic heterogeneity across the clusters and assumes cross-sectional independence across panel members as well. Later I relax latter assumption by including cross-sectional common effects among the explanatory variables as proposed by Pesaran (2006). With this adjustment I expect to obtain estimates robust to the 'cross-correlation bias' discussed in Section 2.2.2.[4] In order to tackle the bias arising from disregarding the heterogeneity of the slope coefficients I perform regressions with both the MG- and the RC-estimators. Finally, I also implement the CCE-correction in the case of the MG models.

Table 5.2 summarizes the estimation results. The first five rows of the table report the results when allowing only for heterogeneity across countries, in the next five rows I show the estimates obtained by clustering the panel across the price indices while the last five rows report results for the case when heterogeneous dynamics is allowed across both dimensions.

Interestingly, when clustering the members of the panel only along the country dimension (as I did in the previous subsection) did not produce substantially different results from

[4]As an alternative treatment of cross-sectional dependence the SURE approach is commonly used in applied research; however, the dimensions of my data set (312 cross sections and a maximum of 204 time periods) make the estimation of seemingly unrelated regressions unfeasible as it can only be applied to panels with a small N relative to T.

Table 5.2: Estimation results, sectoral real exchange rates (COICOP-level 2).

Model	Slope hetero.	Cross-dep.	Cluster[1]	$\hat{\delta}$	Half-life	Homogeneity test Constant	Homogeneity test Slope	BP-test	CD-test
FE	No	No	C	0.968	21	259.06		4.1e05	232.67
				[0.918, 1.018]	[8, ∞]	(0.00)		(0.00)	(0.00)
FE (CCE)	No	Yes	C	0.967	21	252.27		9.6e05	365.32
				[0.914, 1.021]	[8, ∞]	(0.00)		(0.00)	(0.00)
MG[2]	Yes	No	C	0.961	17	720.98	718.23	3.9e05	219.45
				[0.957, 0.964]	[16, 19]	(0.00)	(0.00)	(0.00)	(0.00)
MG[2] (CCE)	Yes	Yes	C	0.956	15	28.89	840.37	3.8e05	206.1
				[0.953, 0.959]	[14, 17]	(0.00)	(0.00)	(0.00)	(0.00)
RC	Yes	Yes	C	0.961	18			3.8e05	208.9
				[0.941, 0.982]	[11, 37]			(0.00)	(0.00)
FE	No	No	S	0.998	343	1.6		3.9e05	202.74
				[0.998, 0.998]	[305, 392]	(0.09)		(0.00)	(0.00)
FE (CCE)	No	Yes	S	0.998	393	1.69		1.1e06	414.74
				[0.998, 0.998]	[345, 456]	(0.07)		(0.00)	(0.00)
MG[2]	Yes	No	S	0.998	343	1.8	1.1e05	3.9e05	202.81
				[0.998, 0.998]	[278, 448]	(0.05)	(0.00)	(0.00)	(0.00)
MG[2] (CCE)	Yes	Yes	S	0.998	395	1.83	1.1e05	4.8e05	293.27
				[0.998, 0.999]	[319, 520]	(0.04)	(0.00)	(0.00)	(0.00)
RC	Yes	Yes	S	0.998	348			3.8e05	206.7
				[0.998, 0.998]	[309, 398]			(0.00)	(0.00)
FE	No	No	CS	0.967	21	21.19		4.1e05	234.32
				[0.958, 0.977]	[16, 29]	(0.00)		(0.00)	(0.00)
FE (CCE)	No	Yes	CS	0.967	20	20.68		9.6e05	364.21
				[0.956, 0.978]	[15, 31]	(0.00)		(0.00)	(0.00)
MG[2]	Yes	No	CS	0.945	12	80.35	80.28	3.8e05	204.85
				[0.94, 0.95]	[11, 14]	(0.00)	(0.00)	(0.00)	(0.00)
MG[2] (CCE)	Yes	Yes	CS	0.906	7	18.63	87.74	3.7e05	203.03
				[0.9, 0.912]	[7, 7]	(0.00)	(0.00)	(0.00)	(0.00)
RC	Yes	Yes	CS	0.953	14			3.8e05	204.81
				[0.94, 0.965]	[11, 19]			(0.00)	(0.00)

Notes: Figures in brackets denote standard errors in the case of the estimated slope coefficients and their implied confidence interval in half-life terms. Figures in parenthesis denote the corresponding p-values of the presented test statistics. [2] C, S and CS denote country-, sector- and country- and sector-specific clustering, respectively. [2] MG-estimations were performed using the Stata package of Eberhardt (2011).

the estimates based on aggregate price indices, at least in terms of the point estimates. The Fixed Effect estimator (both with and without the CCE-correction) reported half-lives of 21 months, just as in the case of the headline HICP data. The estimated confidence intervals are now; however, significantly larger, ranging between 8 months and infinity. The MG- and the RC-models now estimates the half-lives to be somewhat smaller than in the previous case, and, most importantly, the confidence intervals now span over a substantially shorter range. Somewhat surprisingly, the results of the MG-CCE model indicate larger half-lives of the shocks than they did in the case of aggregate price indices (15 months vs. 7 months in the aggregate case) but lie still below the estimates of the other models.

Turning to the case when the cross-sectional units were clustered only across consumption baskets, we can observe remarkably high persistence estimates in case of all estimators. The implied half-lives are close to thirty years for the FE-, MG- and RC-model while the CCE-augmented estimators put the half-lives just below 33 years. Needless to say, these persistence parameters are in strong contradiction with both the previous estimates and the results of the PPP-literature in general and point towards a strong upward bias caused by neglecting the heterogeneity across countries in the model.

Finally, as the last part of the analysis I extended the cross-sectional dimension of the sample across both countries and sectoral price indices thereby allowing for different model parameters across all panel members. In this setup, all coefficients show substantial improvements compared to the previous cases. Although the point estimates of the FE-estimators are still basically identical to those obtained on the sample of aggregate prices and in the case of country-specific clustering, the confidence intervals are now narrower and below the lower bound of the consensus view. The estimates of the RC-model have improved in terms of both lower implied half-lives (14 months now) and narrower confidence intervals (with an upper bound of 19 months). Similarly to the previous cases, the MG-estimator reported the lowest coefficient estimates, 12 and 7 months in half-life terms without and with the CCE-correction, respectively.

Similarly to the case of aggregate prices, the coefficient tests support the need for heterogeneity in essentially all setups while the residual correlation tests also favor heterogeneous model over the standard ones.

The results obtained on the samples of sectoral price indices of the COICOP-levels three and four are reported in Table A.3 and A.4 in the Appendix, respectively. As the most important difference compared to the level-2 estimates one should point out that the estimated coefficients of the FE-estimators dropped substantially in the case of more disaggregated prices. Nevertheless, the coeficient tests indicate that heterogeneous models should still be preferred over FE. For latter, the estimates are basically unchanged compared to the results reported in this chapter, indicating that aggregation bias is mostly observable in the case of the headline price indices and plays a less important role at the disaggregated levels.

The results presented in this section have three important implications. First and foremost, comparing the results of the second and third setups one can find empirical evidence for substantial heterogeneity bias across countries: when allowing only for differences across sectors I found persistence estimates in the magnitude of three decades in terms of half-lives whereas allowing also for country-specific heterogeneity resulted in half-life estimates ranging between 7 and 21 months. Comparing the first and last setups one can observe some additional reduction in the estimated parameters; however, these are by far not as striking as in the case of the first comparison. Thus, heterogeneity bias is mostly present in the cross-country dimension of our data.

Second, by comparing the results of the last setup to those obtained in the case of aggregate prices one cannot observe a striking difference in the point estimates of the models: the FE and FE-CCE estimators consistently report half-lives of 20-21 months, the MG and RC models estimate half-lives of 20-24 months on the aggregate sample and 12-14 months in the case of sectoral prices while the MG-CCE estimator implies half-lives of 7 months in both cases. These results contradict to some extent to those of Imbs et al. (2005*b*) who report a dramatical fall in the average persistence estimates once they account for heterogeneous dynamics. Our results mostly differ in the case of aggregate prices where they obtain half-life estimates well within Rogoff's (1996) consensus view while in our sample even a simple FE estimator put the half-lives to approximately 20 months.

Finally, by analyzing 3 levels of disaggregated prices I found that aggregation bias has the largest impact in the case of the headline price levels since at for the COICOP-levels 2-4, the estimated coefficients of the heterogeneous model were remarkably similar. Of course, most of the actual aggregation is performed between levels 1 and 2, as here 12 indices are merged to form the headline figures, while level-2 and level-3 indices have 3 and 2 input indices from the lower levels on average, respectively.

Chapter 6

Robustness Checks

In order to increase the external validity of my results, I re-estimate the regressions on two additional samples: first I investigate the presence of aggregation bias in real exchange rates based on the UK as the reference country (thereby allowing for more fluctuation in the nominal exchange rates, then I restrict my sample in a way to make it directly comparable to the data set of Imbs et al. (2005*b*). As two additional tests I present two small case studies related to my research: one on the appropriate number of bootstrap repetitions and one on correcting for small sample bias.

6.1 Estimations on the UK Sample

As indicated earlier, although the choice of Germany as the reference country have several advantages, it has one drawback: Germany has been part of the EMU from the very beginnings, resulting in fixed nominal exchange rates for the majority of my sample. In order to assess the robustness of my results with regards nominal exchange rate movements, I carried out the previous analysis on real exchange rates based UK as the numeraire country as well. Tables A.5 and A.6 summarize the results.

On the aggregate level, the FE-estimators reported half-life estimates below one year, however, the confidence intervals are extremely large (e.g. in the case of FE-CCE it ranges between two months and infinity) so that the point estimates cannot be considered reliable. This is confirmed by the slope coefficient tests of the heterogeneous model as well which indicate that FE-model suffer from model misspecification (neglected heterogeneity). The implied half-lives estimated by the heterogeneous model are on the other hand basically unchanged, supporting the validity of our previous findings.

On the sectoral level the fixed effects estimators again report half-lives less than twelve months. The confidence intervals are reassuringly small when allowing for both country- and sectoral effects, however, due to the significant slope homogeneity test I continue to favor the heterogeneous models over the FE-specification. The results of these models remained essentially unchanged compared to the results of the aggregate level, thereby somewhat contradicting my previous results and supporting only the importance of heterogeneity across countries and cross-dependence bias but indicating that allowing for additional sectoral heterogeneity does not result in the reduction of the aggregation bias.

6.2 Estimations on the Imbs et al. (2005*b*) Sample

In Chapter 4 my results were evaluated using two key papers of the literature, Rogoff (1996) and Imbs et al. (2005*b*). Naturally, these studies are not the most appropriate benchmarks in the sense that their results were obtained from samples of different countries, different time periods etc. In order to mitigate the effects of such differences and increase the validity of such comparisons I restricted my sample to the countries and sectors analyzed by Imbs et al. (2005*b*).[1]

Tables A.7 and A.8 in the Appendix summarize the results of the estimations. We can observe several interesting implications. First and foremost, our results for the aggregate real exchange rates for homogeneous estimators are now substantially closer to Imbs et al.'s (2005*b*) estimations and suggest half-lives of 39 months, well in Rogoff's (1996) consensus band. This is a striking difference to the results reported in Section 5.2.1 where even these "naive" estimators reported half-lives of less than two years. As for the estimators correcting for cross-dependence the reported half-lives are somewhat smaller and range around 28-months for the corrected FE models, while the MG-CCE, which again controlled the most for residual cross-correlation according for both tests the estimated half-life is just above one year.

Similarly to our results in Section 5.2.2, the regression estimates for the standard estimators did not change substantially in the case of sectoral price indices. The reported

[1]This subsample can be considered as a direct extension of Imbs et al.'s (2005*b*) data set as it starts in January 1996, just after the last observation of the cited paper. As price indices according to the COICOP-classification are only recorded since 1996, the sectors of the two sample do not coincide exactly but the 19 sectors in Imbs et al. (2005*b*) strongly overlap the 12 sectoral indices of the second level of the COICOP-structure. As for the countries, the coverage is 100%.

half-lives for the cross-correlation corrected estimators dropped; however, substantially, with a point estimate for the MG-CCE estimator of 8 months. Our results confirm that heterogeneity bias is mostly present across countries as extending the cross-sectional dimension across both countries and sectors did not show a substantial reduction in half-lives. Finally, all diagnostic tests confirm the presence of heterogeneous coefficients as well as heterogeneous residual structure in the panel.

Compared to Imbs et al.'s (2005*b*) results I could confirm their conclusion that once allowing for heterogeneity on the sectoral level, the reported half-lives are substantially smaller compared to standard estimators (with strikingly similar half-life estimates for the FE, MG and MG-CCE estimators in the disaggregated case. However, this sample also confirmed my previous finding that not only heterogeneity matters: once allowing for a non-diagonal variance-covariance matrix, even estimates obtained on the aggregate level turn out to be substantially lower than the results obtained by standard estimation procedures.

6.3 Bootstrap Repetitions

In my thesis I often referred to my estimation results in terms of half-lives, emphasizing the importance of the implied confidence intervals of the point estimates. Therefore it is essential to obtain the most accurate standard errors available. Recent literature heavily builds on non-parametric bootstrap methods to obtain these estimates. Following this best practice I calculated the standard errors for the LS-type estimators based on such methods. Due to computational limitations these estimates are based on 100 bootstrap simulations. In order to assess the magnitude of bias resulting from this relatively low number of replications I re-estimated the parameters described in Section 5.2.2 based on 1000 replications as well.

Table A.9 in the Appendix summarizes the results of this small case study. As we can see, the estimated standard errors are somewhat smaller based on 100 replications when the sample was clustered by countries, while grouping by sectors generally led to larger standard errors for the smaller number of replications. Finally, when allowing for both country- and sector-specific clustering, the different numbers of simulations basically led to the same results. The nominal average of the ratios reported in the last column of Table A.9 is exactly one, therefore I conclude that based on this small simulation, 100 bootstrap repetitions proved to be enough for my applied work.

6.4 Small Sample Bias Corrected Estimates

This subsection addresses the problem of small sample bias in the case of the aggregate real exchange rates. As Chen and Engel (2005) note, small sample bias generally leads to estimated half-lives below their true parameter value. Therefore they suggest three alternative procedure to address this problem: 1) the median-unbiased estimator proposed by Andrews (1993); the two-step bootstrap procedure by Kilian (1998); and lastly the recursive demeaning procedure of So and Shin (1999). In this section we present the results of the MG-estimators with and without small-sample bias correction. Due to computational limitations, the small-sample-corrected estimators were only calculated for the COICOP-levels one and two. The below table summarizes the results:

Table 6.1: Small sample bias corrected estimates.

Level	Standard Estimator			So and Shin (1999) Correction		
	Point Estimate	Half-life	CI	Point Estimate	Half-life	CI
1	0.9667	20	[14–39]	0.9953	146	[40–∞]
2	0.9453	12	[11–14]	0.9761	29	[23–38]

As we can see, by performing So and Shin's (1999) correction method the estimated half-lives did increase considerably, with the sectoral estimate now lying at the lower bound of Rogoff's (1996) consensus view while the aggregate estimate being significantly larger than the results of the previous literature. However, the following points should be made: by definition, the small sample bias correction methods were meant to correct estimators which are applied to samples with a limited number of observations, whereas even in the aggregate case the MG-estimator is calculated using 5267 data points. Moreover, according to the results of Chen and Engel (2005), So and Shin's (1999) method results in the largest upward correction of the original estimates. Finally, this analysis is limited to two estimates, therefore the findings should interpreted with the appropriate caution. All in all, due to the vast number of observations in my sample, the downward bias observed in small samples might not play an important role in this case.

Chapter 7

Conclusions

In my thesis I investigate the presence of aggregation bias introduced by Imbs et al. (2005*b*) on a panel of sectoral European price indices. In this setup, aggregation bias can emerge along two dimensions, 1) due to neglecting the heterogeneity of the underlying sectoral or country-specific real exchange rates or 2) when disregarding the heterogeneous covariance-structure of the stochastic terms of the panel members. General intuition supports the presence of such bias in our data, as did the formal tests performed on the estimated models. Therefore, several corrections methods were applied in order to quantify the magnitude of aggregation bias.

My empirical results confirm the presence of aggregation bias, although not necessarily in the form as reported by Imbs et al. (2005*b*): they report half-life estimates obtained by standard panel estimators both for aggregate and sectoral price indices in the range of the consensus view of the literature. However, after allowing for heterogeneous persistence and dependence across cross-sections, they obtain significantly lower estimates of mean reversion, and hence they conclude that aggregation bias does have a substantial positive impact on the estimators which neglect the previously introduced forms of bias.

As the most important difference to the findings of both Imbs et al. (2005*b*) and Rogoff's consensus view, in my analysis I obtained persistence estimates of below 2 years even based on homogeneous panel estimators. These results seem robust and were confirmed on several levels of aggregation as well as on a control sample real exchange rates with the UK as the reference country. However, after restricting my sample to match that of Imbs et al. (2005*b*) and calculating USD-based real exchange rates, I obtained similar results to the cited article, indicating a faster pace of conversion for intra-EU real exchange rates than compared to the US.

Allowing for heterogeneous slopes in the panel as well as correcting for residual cross dependence did result in substantially lower half-life estimates, just as in the case of Imbs et al. (2005*b*). However, while according to their findings aggregation bias plays a role when constructing aggregates based on sectoral data, I found evidence for such bias also on the country level. Interestingly, the preferred MG-CCE estimator did report estimated half-lives of 7 months on both aggregate and sectoral levels, indicating that once allowing for country-specific slopes and controlling for cross-dependence in the error structure, the estimated persistence of the real exchange rates is similar. This was confirmed on both my control samples.

As an additional result of my analysis, I showed that the bias resulting from neglected heterogeneity is substantially larger in the case of countries than for the sectors, indicating that assuming homogeneous adjustment processes for the sectoral real exchange rates within a country does not result in severe aggregation bias; however, it is the cross-country heterogeneity that mostly matters.

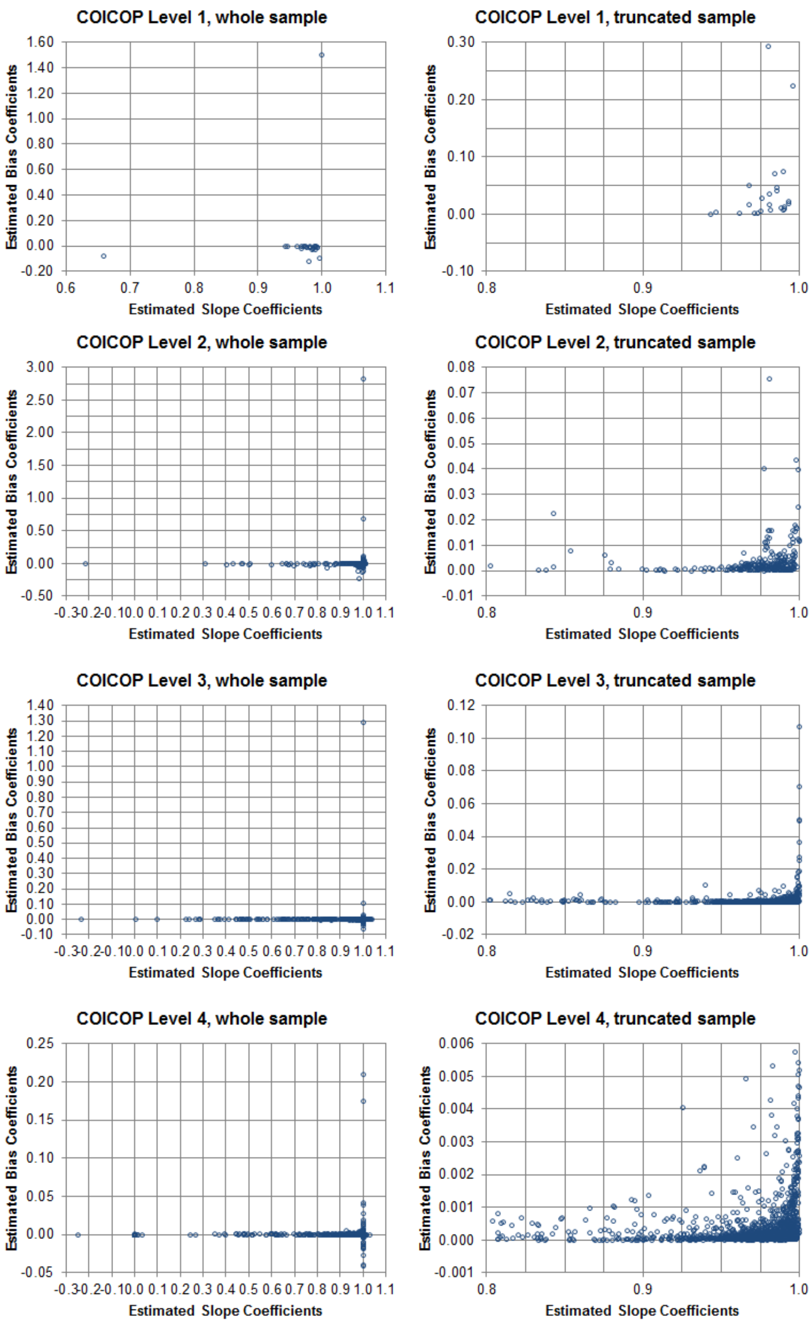
As an important policy implication of my thesis I found support for the stationarity of the real exchange rates at a rate of conversion which is consistent with nominal rigidities thereby providing further evidence against the existence of Rogoff's puzzle. In contrast with most studies in the recent literature, I obtained half-lives of less than 2 years in case of aggregate price indices on all the investigated samples, once I allowed for heterogeneous adjustment dynamics. Second, I found that "intra-European" real exchange rates show a higher pace of conversion than the USD-based rates – this finding underlines the importance of choosing the right numeraire country and supports the choice of a European country over the US when analyzing European real exchange rates.

My research can be extended into several directions. First, due to the extensive dimensions of my sample, the behavior of the real exchange rates could be analyzed on different subsamples by e.g. comparing the persistence of PPP-deviations over different parts of the business cycles. Second, alternative forms of bias suggested by the previous literature (i.e. bias due to neglecting non-linear mean reversion, time aggregation bias) could be also considered and analyzed jointly, with the bias due to heterogeneity in adjustment dynamics and cross-dependent error processes. Finally, my sample could be merged with that of Imbs et al. (2005*b*) to see how the previous results hold over the time span of 30 years. All these remain open quetions of my thesis and shall be addressed by future research.

Appendix A

Figures & Tables

FIGURE A.1: Estimated slope and heterogeneity bias coefficients across sectors and countries.



The figures plot the estimated slope coefficients (horizontal axis) against the estimated bias coefficients (vertical axis) as defined in Equation 2.7. The left charts show the results for all cross-sections while on the right charts we truncated our sample to the cross-sections with estimated slope coefficients between 0.8 and 1.

TABLE A.1: Average number of observations by country and COICOP level.

Country	Levels					Country	Levels				
	1	2	3	4	Total		1	2	3	4	Total
AT	204	204	198	191	194	IT	204	204	199	186	192
BE	204	200	199	184	190	LT	204	204	198	189	193
BG	193	193	186	175	180	LU	204	204	199	191	195
CY	204	204	197	182	188	LV	204	204	192	178	185
CZ	204	204	155	150	157	MT	204	204	190	168	178
DK	204	204	196	194	196	NL	204	204	187	186	188
EE	204	204	179	173	178	PL	204	204	194	198	198
EL	204	204	195	191	194	PT	204	204	198	186	191
ES	204	204	194	182	188	RO	204	204	138	132	141
FI	204	204	200	189	194	SE	204	204	188	187	189
FR	204	204	200	195	197	SI	204	204	156	143	153
HU	204	204	139	136	143	SK	204	204	195	187	191
IE	204	204	189	190	191	UK	204	204	188	178	184
Total	204	203	186	178	183						

TABLE A.2: Estimated slope and heterogeneity bias coefficients across countries (COICOP Level 1).

Country	$\hat{\delta}_c$	$\hat{\sigma}_c^2$	$\hat{\gamma}_c$	Country	$\hat{\delta}_c$	$\hat{\sigma}_c^2$	$\hat{\gamma}_c$
AT	0.9735	0.00013	-0.0006	IT	0.9812	0.00055	-0.0035
BE	0.9467	0.00068	-0.0016	LT	0.9855	0.000235	-0.0196
BG	0.6596	0.018108	-0.0770	LU	0.9934	0.00042	-0.0076
CY	0.9807	0.000119	-0.0074	LV	1.0001	0.000176	1.5039
CZ	0.9894	0.000279	-0.0318	MT	0.9674	0.000184	-0.0069
DK	0.9711	0.00022	-0.0009	NL	0.9750	0.00044	-0.0021
EE	0.9901	0.00034	-0.0042	PL	0.9681	0.000563	-0.0215
EL	0.9811	0.000241	-0.0154	PT	0.9896	0.00027	-0.0032
ES	0.9906	0.00040	-0.0051	RO	0.9800	0.002053	-0.1249
FI	0.9615	0.00031	-0.0010	SE	0.9757	0.000231	-0.0116
FR	0.9430	0.00013	-0.0003	SI	0.9880	0.00046	-0.0046
HU	0.9842	0.000394	-0.0303	SK	0.9854	0.000212	-0.0176
IE	0.9928	0.00056	-0.0094	UK	0.9955	0.000356	-0.0957
$Corr(\hat{\delta}_c, \hat{\gamma}_c) = 0.1360$							

Table A.3: Estimation results, sectoral real exchange rates (COICOP-level 3).

Model	Slope hetero.	Cross-dep.	Cluster[1]	$\hat{\delta}$	Half-life	Homogeneity test Constant	Homogeneity test Slope	BP-test	CD-test
FE	No	No	C	0.94 [0.88, 1]	11 [5, ∞]	534.7 (0.00)		4.5e06 (0.00)	612.24 (0.00)
FE (CCE)	No	Yes	C	0.94 [0.881, 0.998]	11 [5, 342]	524.14 (0.00)		9.1e06 (0.00)	1151.02 (0.00)
MG	Yes	No	C	0.959 [0.955, 0.963]	16 [15, 18]	1440.19 (0.00)	1480.24 (0.00)	3.0e60 (0.00)	493.61 (0.00)
MG (CCE)	Yes	Yes	C	0.957 [0.953, 0.96]	16 [14, 17]	185.13 (0.00)	2782.14 (0.00)	2.8e06 (0.00)	427.46 (0.00)
RC	Yes	Yes	C	0.959 [0.939, 0.979]	17 [11, 33]			3.0e060 (0.00)	492.59 (0.00)
FE	No	No	S	0.999 [0.998, 0.999]	510 [414, 662]	3.52 (0.00)		2.6e06 (0.00)	412.71 (0.00)
FE (CCE)	No	Yes	S	0.999 [0.999, 0.999]	577 [475, 735]	3.64 (0.00)		1.0e07 (0.00)	1310.89 (0.00)
MG	Yes	No	S	0.999 [0.998, 0.999]	495 [409, 627]	1.97 (0.00)	23122.12 (0.00)	2.6e06 (0.00)	413.82 (0.00)
MG (CCE)	Yes	Yes	S	0.999 [0.999, 0.999]	571 [472, 723]	1.88 (0.00)	21927.48 (0.00)	2.7e06 (0.00)	459.64 (0.00)
RC	Yes	Yes	S	0.999 [0.998, 0.999]	547 [447, 706]			2.6e06 (0.00)	413.72 (0.00)
FE	No	No	CS	0.93 [0.916, 0.944]	10 [8, 12]	16.14 (0.00)		5.4e06 (0.00)	674.22 (0.00)
FE (CCE)	No	Yes	CS	0.929 [0.915, 0.942]	9 [8, 12]	15.97 (0.00)		9.2e06 (0.00)	1113.12 (0.00)
MG	Yes	No	CS	0.94 [0.937, 0.943]	11 [11, 12]	66.44 (0.00)	66.61 (0.00)	2.5e06 (0.00)	433.79 (0.00)
MG (CCE)	Yes	Yes	CS	0.91 [0.907, 0.914]	7 [7, 8]	34.49 (0.00)	69.77 (0.00)	2.5e06 (0.00)	427.37 (0.00)
RC	Yes	Yes	CS	0.948 [0.939, 0.957]	13 [11, 16]			2.6e06 (0.00)	433.45 (0.00)

Notes: Figures in brackets denote standard errors in the case of the estimated slope coefficients and their implied confidence interval in half-life terms. Figures in parenthesis denote the corresponding p-values of the presented test statistics. [1] C, S and CS denote country-, sector- and country- and sector-specific clustering, respectively.

TABLE A.4: Estimation results, sectoral real exchange rates (COICOP-level 4).

Model	Slope hetero.	Cross-dep.	Cluster[1]	$\hat{\delta}$	Half-life
FE	No	No	C	0.942 [0.883, 1.002]	12 [6, ∞]
FE (CCE)	No	Yes	C	0.942 [0.879, 1.004]	12 [5, ∞]
MG	Yes	No	C	0.973 [0.971, 0.975]	25 [24, 27]
MG (CCE)	Yes	Yes	C	0.976 [0.974, 0.978]	28 [26, 31]
RC	Yes	Yes	C	0.973 [0.955, 0.991]	25 [15, 81]
FE	No	No	S	0.999 [0.999, 0.999]	573 [514, 647]
FE (CCE)	No	Yes	S	0.999 [0.999, 0.999]	654 [587, 738]
MG	Yes	No	S	0.999 [0.999, 0.999]	557 [479, 666]
MG (CCE)	Yes	Yes	S	0.999 [0.999, 0.999]	633 [551, 743]
RC	Yes	Yes	S	0.999 [0.999, 0.999]	610 [542, 698]
FE	No	No	CS	0.934 [0.924, 0.944]	10 [9, 12]
FE (CCE)	No	Yes	CS	0.932 [0.922, 0.943]	10 [9, 12]
MG	Yes	No	CS	0.953 [0.949, 0.958]	14 [13, 16]
MG (CCE)	Yes	Yes	CS	0.933 [0.929, 0.938]	10 [9, 11]
RC	Yes	Yes	CS	0.960 [0.955, 0.965]	17 [15, 19]

Notes: Figures in brackets denote standard errors in the case of the estimated slope coefficients and their implied confidence interval in half-life terms. Figures in parenthesis denote the corresponding p-values of the presented test statistics. [1] C, S and CS denote country-, sector- and country- and sector-specific clustering, respectively.

TABLE A.5: Results, UK-based aggregate real exchange rates.

Model	Heterogeneity Constant	Heterogeneity Slope	Cross-dep.	$\hat{\delta}$	Half-life	Homogeneity test constant	Homogeneity test slope	BP-test	CD-test
OLS	No	No	No	0.999	888			31205.89	163.38
				[0.998, 1.001]	[317, ∞]			(0.00)	(0.00)
FE	Yes	No	No	0.922	9	26.09		28536.03	155.44
				[0.839, 1.006]	[4, ∞]	(0.00)		(0.00)	(0.00)
MG	Yes	Yes	No	0.966	20	160.26	177.76	30339.97	158.78
				[0.947, 0.985]	[13, 46]	(0.00)	(0.00)	(0.00)	(0.00)
FE (SURE)	Yes	No	Yes	0.956	16	209.21		27167.33	155.9
				[0.95, 0.963]	[14, 18]	(0.00)		(0.00)	(0.00)
FE (CCE)	Yes	No	Yes	0.881	5	40.77		11023.27	42.54
				[0.744, 1.019]	[2, ∞]	(0.00)		(0.00)	(0.00)
MG (CCE)	Yes	Yes	Yes	0.935	10	11.82	106.19	11843.89	65.98
				[0.916, 0.955]	[8, 15]	(0.00)	(0.00)	(0.00)	(0.00)
RC	Yes	Yes	Yes	0.969	22			28463.92	148.28
				[0.932, 1.005]	[10, ∞]			(0.00)	(0.00)

Notes: Figures in brackets denote standard errors in the case of the estimated slope coefficients and their implied confidence interval in half-life terms. Figures in parenthesis denote the corresponding p-values of the presented test statistics.

Table A.6: Results, UK-based sectoral real exchange rates.

Model	Slope hetero.	Cross-dep.	Cluster[1]	$\hat{\delta}$	Half-life	Homogeneity test Constant	Homogeneity test Slope	BP-test	CD-test
FE	No	No	C	0.9454	12	214.87		2.7e06	1433.35
				[0.8775, 1.0133]	[5, ∞]	(0.00)		(0.00)	(0.00)
FE (CCE)	No	Yes	C	0.9332	10	263.07		6.9e05	306.37
				[0.8657, 1.0007]	[5, ∞]	(0.00)		(0.00)	(0.00)
MG	Yes	No	C	0.973	25	917.19	1386.37	2.7e06	1429.76
				[0.9698, 0.9762]	[23, 29]	(0.00)	(0.00)	(0.00)	(0.00)
MG (CCE)	Yes	Yes	C	0.9708	23	85.38	964.69	2.7e06	1424.37
				[0.9678, 0.9737]	[21, 26]	(0.00)	(0.00)	(0.00)	(0.00)
RC	Yes	Yes	C	0.9733	26			27e06	1430.85
				[0.9505, 0.9961]	[14, 179]			(0.00)	(0.00)
FE	No	No	S	0.9991	751	4.72		28e06	1470.37
				[0.9989, 0.9993]	[609, 979]	(0.00)		(0.00)	(0.00)
FE (CCE)	No	Yes	S	0.9994	1067	5.63		8.0e06	423.05
				[0.9992, 0.9995]	[882, 1352]	(0.00)		(0.00)	(0.00)
MG	Yes	No	S	0.9991	753	5	1.1e05	2.8e06	1469.81
				[0.9986, 0.9995]	[506, 1472]	(0.00)	(0.00)	(0.00)	(0.00)
MG (CCE)	Yes	Yes	S	0.9993	1064	0.22	9.7e04	2.7e06	1452.33
				[0.999, 0.9997]	[670, 2584]	(1)	(0.00)	(0.00)	(0.00)
RC	Yes	Yes	S	0.9991	781			2.8e06	1470.02
				[0.9989, 0.9993]	[623, 1046]			(0.00)	(0.00)
FE	No	No	CS	0.9429	12	18.07		2.7e06	1431.17
				[0.9244, 0.9613]	[9, 18]	(0.00)		(0.00)	(0.00)
FE (CCE)	No	Yes	CS	0.9279	9	22.71		7.0e06	294.71
				[0.901, 0.9549]	[7, 15]	(0.00)		(0.00)	(0.00)
MG	Yes	No	CS	0.9671	21	91.67	82.39	2.7e06	1411.51
				[0.9625, 0.9716]	[18, 24]	(0.00)	(0.00)	(0.00)	(0.00)
MG (CCE)	Yes	Yes	CS	0.939	11	12.22	78.89	2.6e06	1394.6
				[0.9344, 0.9436]	[10, 12]	(0.00)	(0.00)	(0.00)	(0.00)
RC	Yes	Yes	CS	0.9709	23			2.7e06	1419.56
				[0.9626, 0.9792]	[18, 33]			(0.00)	(0.00)

Notes: Figures in brackets denote standard errors in the case of the estimated slope coefficients and their implied confidence interval in half-life terms. Figures in parenthesis denote the corresponding p-values of the presented test statistics. [1] C, S and CS denote country-, sector- and country- and sector-specific clustering, respectively.

Table A.7: Results, aggregate real exchange rates, "Imbs et al. (2005*b*) sample".

Model	Heterogeneity Constant	Heterogeneity Slope	Cross-dep.	$\hat{\delta}$	Half-life	Homogeneity test constant	Homogeneity test slope	BP-test	CD-test
OLS	No	No	No	1.000	3804			7596.37	86.8
				[0.999, 1.001]	[738, ∞]			(0.00)	(0.00)
FE	Yes	No	No	0.982	39	4.9		7678.64	87.33
				[0.976, 0.988]	[29, 59]	(0.00)		(0.00)	(0.00)
Arellano-Bond	Yes	No	No	0.982	39			7678.69	87.33
				[0.972, 0.993]	[24, 96]			(0.00)	(0.00)
Blundell-Bond	No	No	No	0.998	3804			7596.37	86.80
				[0.992, 1.0004]	[911, ∞]			(0.00)	(0.00)
MG	Yes	Yes	No	0.984	44	5.27	160.2	7643.9	87.62
				[0.976, 0.992]	[29, 89]	(0.00)	(0.00)	(0.00)	(0.00)
FE (SURE)	Yes	No	Yes	0.977	30	135.18		7678.64	87.82
				[0.972, 0.982]	[24, 38]	(0.00)		(0.00)	(0.00)
FE (CCE)	Yes	No	Yes	0.976	28	32.2		1415.17	10.31
				[0.968, 0.983]	[21, 40]	(0.00)		(0.00)	(0.00)
MG (SURE)	Yes	Yes	Yes	0.969	22	18.64	118.61	7643.9	87.11
				[0.5, 0.5]	[1, 1]	(0.05)	(0.00)	(0.00)	(0.00)
MG (CCE)	Yes	Yes	Yes	0.951	14	6.71	165.83	1290.65	8.67
				[0.939, 0.963]	[11, 18]	(0.00)	(0.00)	(0.00)	(0.00)
RC	Yes	Yes	Yes	0.988	57			7660.41	87.22
				[0.982, 0.994]	[39, 110]			(0.00)	(0.00)

Notes: Figures in brackets denote standard errors in the case of the estimated slope coefficients and their implied confidence interval in half-life terms. Figures in parenthesis denote the corresponding p-values of the presented test statistics.

TABLE A.8: Results, sectoral real exchange rates, "Imbs et al. (2005*b*) sample".

Model	Slope hetero.	Cross-dep.	Cluster[1]	$\hat{\delta}$	Half-life	Homogeneity test Constant	Homogeneity test Slope	BP-test	CD-test
FE	No	No	C	0.98	34	51.35		9.1e05	933.16
				[0.9716, 0.9884]	[24, 59]	(0.00)		(0.00)	(0.00)
FE (CCE)	No	Yes	C	0.9714	24	114.24		1.2e05	59.66
				[0.9548, 0.9881]	[15, 58]	(0.00)		(0.00)	(0.00)
MG	Yes	No	C	0.9802	35	45.05	1475.87	9.1e05	928.98
				[0.9779, 0.9825]	[31, 39]	(0.00)	(0.00)	(0.00)	(0.00)
MG (CCE)	Yes	Yes	C	0.9569	16	24.37	197.16	8.9e05	920.94
				[0.9518, 0.962]	[14, 18]	(0.00)	(0.00)	(0.00)	(0.00)
RC	Yes	Yes	C	0.9812	37			9.1e05	929.66
				[0.9766, 0.9858]	[29, 48]			(0.00)	(0.00)
FE	No	No	S	0.9998	2804	1.74		9.0e050	927.75
				[0.9997, 0.9998]	[2040, 4480]	(0.06)		(0.00)	(0.00)
FE (CCE)	No	Yes	S	0.9999	4954	3.96		1.2e05	64.27
				[0.9998, 0.9999]	[3237, 10550]	(0.00)		(0.00)	(0.00)
MG	Yes	No	S	0.9998	2799	1.58	11e05	9.0e050	927.74
				[0.9995, 1]	[1338, ∞]	(0.1)	(0.00)	(0.00)	(0.00)
MG (CCE)	Yes	Yes	S	0.9999	5033	0.33	98414.53	9.0e050	926.96
				[0.9997, 1]	[2411, ∞]	(0.98)	(0.00)	(0.00)	(0.00)
RC	Yes	Yes	S	0.9998	3302			9.0e050	927.74
				[0.9997, 0.9998]	[2686, 4284]			(0.00)	(0.00)
FE	No	No	CS	0.9796	34	4.18		9.1e05	933.19
				[0.9749, 0.9843]	[27, 44]	(0.00)		(0.00)	(0.00)
FE (CCE)	No	Yes	CS	0.9703	23	9.48		1.2e05	59.41
				[0.9605, 0.9802]	[17, 35]	(0.00)		(0.00)	(0.00)
MG	Yes	No	CS	0.9737	26	8.05	15.61	9.1e05	930.21
				[0.9709, 0.9765]	[23, 29]	(0.00)	(0.00)	(0.00)	(0.00)
MG (CCE)	Yes	Yes	CS	0.9243	9	19.97	24.5	9.0e050	926.22
				[0.9196, 0.929]	[8, 9]	(0.00)	(0.00)	(0.00)	(0.00)
RC	Yes	Yes	CS	0.98	34			9.1e05	930.5
				[0.9728, 0.9872]	[25, 54]			(0.00)	(0.00)

Notes: Figures in brackets denote standard errors in the case of the estimated slope coefficients and their implied confidence interval in half-life terms. Figures in parenthesis denote the corresponding p-values of the presented test statistics. [1] C, S and CS denote country-, sector- and country- and sector-specific clustering, respectively.

Table A.9: Comparison of bootstrapped standard errors.

Model	Cluster	100 reps		1000 reps		Std. Error
		Std. Error	Conf. Int.	Std. Error	Conf. Int.	Ratio
FE	C	0.0254	[0.9182–1.0178]	0.0299	[0.9094–1.0266]	1.18
FE (CCE)	C	0.0274	[0.9137–1.0212]	0.0331	[0.9025–1.0324]	1.21
MG	C	0.0018	[0.9572–0.9643]	0.0018	[0.9572–0.9643]	1.00
MG (CCE)	C	0.0099	[0.9527–0.9592]	0.0100	[0.9526–0.9593]	1.01
RC	C	0.0102	[0.9414–0.9815]	0.0103	[0.9414–0.9816]	1.00
FE	S	0.0001	[0.9977–0.9982]	0.0001	[0.9978–0.9982]	0.90
FE (CCE)	S	0.0001	[0.998–0.9985]	0.0001	[0.998–0.9984]	0.82
MG	S	0.0002	[0.9975–0.9985]	0.0002	[0.9975–0.9984]	0.98
MG (CCE)	S	0.0001	[0.9978–0.9987]	0.0001	[0.9979–0.9986]	0.98
RC	S	0.0001	[0.9978–0.9983]	0.0001	[0.9978–0.9982]	0.94
FE	CS	0.0047	[0.9583–0.9765]	0.0047	[0.9582–0.9766]	1.01
FE (CCE)	CS	0.0057	[0.9556–0.9778]	0.0052	[0.9566–0.9768]	0.91
MG	CS	0.0069	[0.9402–0.9504]	0.0069	[0.9409–0.9497]	1.00
MG (CCE)	CS	0.0085	[0.9004–0.9117]	0.0085	[0.9006–0.9115]	1.00
RC	CS	0.0064	[0.9400–0.9650]	0.0068	[0.9392–0.9658]	1.06

Notes: The table shows the estimated standard errors based on different bootstrap replications for the results presented in Section 5.2.2.

Bibliography

Abuaf, Niso, and Philippe Jorion. 1990. "Purchasing Power Parity in the Long Run." *The Journal of Finance*, 45(1): pp. 157–174.

Andrews, Donald W K. 1993. "Exactly Median-Unbiased Estimation of First Order Autoregressive/Unit Root Models." *Econometrica*, 61(1): 139–65.

Arellano, Manuel, and Stephen Bond. 1991. "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations." *Review of Economic Studies*, 58(2): 277–97.

Blundell, Richard, and Stephen Bond. 1998. "Initial conditions and moment restrictions in dynamic panel data models." *Journal of Econometrics*, 87(1): 115–143.

Breitung, Jrg, and Samarjit Das. 2005. "Panel unit root tests under cross-sectional dependence." *Statistica Neerlandica*, 59(4): 414–433.

Breusch, T S, and A R Pagan. 1980. "The Lagrange Multiplier Test and Its Applications to Model Specification in Econometrics." *Review of Economic Studies*, 47(1): 239–53.

Cassel, G. 1918. "Abnormal Deviations in International Exchanges." *The Economic Journal*, 28(112): pp. 413–415.

Chen, Shiu-Sheng, and Charles Engel. 2005. "Does 'Aggregation Bias' Explain The Ppp Puzzle?" *Pacific Economic Review*, 10(1): 49–72.

Crucini, Mario J., and Mototsugu Shintani. 2008. "Persistence in law of one price deviations: Evidence from micro-data." *Journal of Monetary Economics*, 55(3): 629–644.

Dickey, David A., and Wayne A. Fuller. 1979. "Distribution of the Estimators for Autoregressive Time Series With a Unit Root." *Journal of the American Statistical Association*, 74(366): pp. 427–431.

Eberhardt, Markus. 2011. "XTMG: Stata module to estimate panel time series models with heterogeneous slopes." *Statistical Software Components, Boston College Department of Economics.*

Eurostat. 1996-2013*a*. “Conversion factors for euro fixed series into euro/ECU - monthly data (ert_bil_conv_m).” Hyperlink available upon request. Retrieved on April 21, 2013.

Eurostat. 1996-2013*b*. “Euro/ECU exchange rates - monthly data (ert_bil_eur_m).” *http://epp.eurostat.ec.europa.eu/NavTree_prod/everybody/BulkDownloadListing?sort=1&downfile=data%2Fert_bil_eur_m.tsv.gz*, Retrieved on April 21, 2013.

Eurostat. 1996-2013*c*. “HICP (2005 = 100) - monthly data (index) (prc_hicp_midx).” *http://epp.eurostat.ec.europa.eu/NavTree_prod/everybody/BulkDownloadListing?sort=1&downfile=data%2Fprc_hicp_cind.tsv.gz*, Retrieved on April 21, 2013.

Frankel, Jeffrey A., and Andrew K. Rose. 1996. “A panel project on purchasing power parity: Mean reversion within and between countries.” *Journal of International Economics*, 40(1-2): 209–224.

Frenkel, Jacob A. 1980. “The Collapse of Purchasing Power Parities during the 1970s.” National Bureau of Economic Research Working Paper 569.

Froot, Kenneth A., and Kenneth Rogoff. 1995. “Perspectives on PPP and long-run real exchange rates.” In *Handbook of International Economics*. Vol. 3 of *Handbook of International Economics*, , ed. G. M. Grossman and K. Rogoff, Chapter 32, 1647–1688. Elsevier.

Froot, Kenneth A., Michael Kim, and Kenneth Rogoff. 1995. “The Law of One Price Over 700 Years.” National Bureau of Economic Research, Inc NBER Working Papers 5132.

Goldberg, Pinelopi K., and Frank Verboven. 2005. “Market integration and convergence to the Law of One Price: evidence from the European car market.” *Journal of International Economics*, 65(1): 49–73.

Hansen, Bruce E. 1999. “The Grid Bootstrap And The Autoregressive Model.” *The Review of Economics and Statistics*, 81(4): 594–607.

Haskel, Jonathan, and Holger Wolf. 2001. “The Law of One Price–A Case Study.” *Scandinavian Journal of Economics*, 103(4): 545–58.

Imbs, Jean, Haroon Mumtaz, Morten O. Ravn, and Hlne Rey. 2005*a*. “"Aggregation Bias" DOES Explain the PPP Puzzle.” National Bureau of Economic Research, Inc NBER Working Papers 11607.

Imbs, Jean, Haroon Mumtaz, Morten Ravn, and Hlne Rey. 2005*b*. “PPP Strikes Back: Aggregation and the Real Exchange Rate.” *The Quarterly Journal of Economics*, 120(1): 1–43.

Imbs, Jean, Haroon Mumtaz, Morton O. Ravn, and Helene Rey. 2002. “PPP Strikes Back: Aggregation and the Real Exchange Rate.” National Bureau of Economic Research, Inc NBER Working Papers 9372.

Im, Kyung So, M. Hashem Pesaran, and Yongcheol Shin. 2003. “Testing for unit roots in heterogeneous panels.” *Journal of Econometrics*, 115(1): 53–74.

Isard, Peter. 1977. “How Far Can We Push the "Law of One Price"?” *American Economic Review*, 67(5): 942–48.

Kilian, Lutz. 1998. “Small-Sample Confidence Intervals For Impulse Response Functions.” *The Review of Economics and Statistics*, 80(2): 218–230.

Levin, Andrew, Chien-Fu Lin, and Chia-Shang James Chu. 2002. “Unit root tests in panel data: asymptotic and finite-sample properties.” *Journal of Econometrics*, 108(1): 1–24.

Lewandowski, Piotr. 2006. “PESCADF: Stata module to perform Pesaran’s CADF panel unit root test in presence of cross section dependence.” *Statistical Software Components, Boston College Department of Economics.*

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

Parsley, DavidC., and Shang-Jin Wei. 2007. “A Prism into the PPP Puzzles: The Micro-Foundations of Big Mac Real Exchange Rates.” *Economic Journal*, 117(523): 1336–1356.

Pesaran, M.H. 2004. “General Diagnostic Tests for Cross Section Dependence in Panels.” Faculty of Economics, University of Cambridge Cambridge Working Papers in Economics 0435.

Pesaran, M. Hashem. 2006. “Estimation and Inference in Large Heterogeneous Panels with a Multifactor Error Structure.” *Econometrica*, 74(4): 967–1012.

Pesaran, M. Hashem. 2007. “A simple panel unit root test in the presence of cross-section dependence.” *Journal of Applied Econometrics*, 22(2): 265–312.

Pesaran, M. Hashem, and Ron Smith. 1995. “Estimating long-run relationships from dynamic heterogeneous panels.” *Journal of Econometrics*, 68(1): 79–113.

Rogoff, Kenneth. 1996. “The Purchasing Power Parity Puzzle.” *Journal of Economic Literature*, 34(2): 647–668.

Roodman, David. 2003. “XTABOND2: Stata module to extend xtabond dynamic panel data estimator.” *Statistical Software Components, Boston College Department of Economics.*

So, Beong Soo, and Dong Wan Shin. 1999. "Recursive mean adjustment in time-series inferences." *Statistics & Probability Letters*, 43(1): 65–73.

Swamy, P A V B. 1970. "Efficient Inference in a Random Coefficient Regression Model." *Econometrica*, 38(2): 311–23.

Taylor, Alan M. 2001. "Potential Pitfalls for the Purchasing-Power-Parity Puzzle? Sampling and Specification Biases in Mean-Reversion Tests of the Law of One Price." *Econometrica*, 69(2): 473–98.

Taylor, Alan M. 2002. "A Century Of Purchasing-Power Parity." *The Review of Economics and Statistics*, 84(1): 139–150.

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

Zellner, Arnold. 1962. "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias." *Journal of the American Statistical Association*, 57(298): pp. 348–368.

Zellner, Arnold. 1963. "Estimators for Seemingly Unrelated Regression Equations: Some Exact Finite Sample Results." *Journal of the American Statistical Association*, 58(304): pp. 977–992.

Zellner, Arnold, and David S. Huang. 1962. "Further Properties of Efficient Estimators for Seemingly Unrelated Regression Equations." *International Economic Review*, 3(3): pp. 300–313.