Federal Reserve Bank of Dallas
Globalization and Monetary Policy Institute
Working Paper No. 138
http://www.dallasfed.org/assets/documents/institute/wpapers/2013/0138.pdf

Spatial Considerations on the PPP Debate*

Michele Ca' Zorzi
European Central Bank

Alexander Chudik
Federal Reserve Bank of Dallas
CAFE and CIMF

January 2013

Abstract

This paper studies the influence of aggregating across space when (i) testing the PPP theory or more generally pair-wise cointegration and (ii) evaluating the PPP puzzle. Our contribution is threefold: we show that aggregating foreign data and applying an ADF test may lead to erroneously reject the PPP hypothesis. We then show, on the basis of theoretical arguments as well as Monte Carlo experiments, that a sizable bias in the estimates of half-life deviations to PPP may be due to the effect of aggregation across space. We finally illustrate empirically the importance of spatial considerations when estimating the speed of price convergence among euro area countries.

JEL codes: C23, F41

* Michele Ca'Zorzi, European Central Bank, Kaiserstrasse 29, 60311 Frankfurt am Main, Germany. +49-69-1344-6563. michele.cazorzi@ecb.europa.eu. Alexander Chudik, Research Department, Federal Reserve Bank of Dallas, 2200 N. Pearl Street, Dallas, TX 75201. 214-922-5769. alexander.chudik@dal.frb.org. The views in this paper are those of the authors and do not necessarily reflect the views of the European Central Bank, the Federal Reserve Bank of Dallas or the Federal Reserve System.

1 Introduction

The hypothesis of PPP convergence is one of the oldest and yet contentious in economics (Taylor and Taylor (2004)). It goes back to the simple and appealing economic theory that international arbitrage in goods market should ensure the same price for the same basket of goods across countries (absolute PPP). More popular, however, is a weaker form of the same theory, which allows for the relative cost of two identical baskets to be constant (relative PPP). Particularly influential was the seminal paper by Rogoff (1996), which pointed to PPP as one of the few areas of exchange rate economics supported empirically. More recently Taylor and Taylor (2004) argued that we are back to the pre-1970s consensus when long-run PPP was a widely accepted theoretical proposition.[1] It is therefore not surprising that the recent debate has concentrated less on the validity of PPP than on Rogoff's puzzle, i.e. the difficulty of reconciling the estimated slow convergence to PPP with nominal rigidities that are realistic.[2]

This literature has identified several underlying factors that may bias the estimated half-life statistics of deviations from PPP convergence. Following Nickell (1981) it is known that the estimated dynamic lag coefficient in time series and panel regressions is biased down. However, in the context of the PPP debate this would reinforce rather than explain the puzzle. Imbs et al. (2005) attempted to explain the puzzle by stressing how half-life estimates might be biased up due to the aggregation bias of heterogenous products. Chambers (2005) has quantified the implications of the bias that comes from the temporal aggregation of price data. Chi-Young et al. (2006) addressed the three above mentioned sources of bias in a panel data setting. Finally, Pesaran et al. (2009) argued that one should use a pair-wise approach to avoid the dependence on the choice of a numeraire country.[3]

Our paper adds to this literature by showing the key role of spatial considerations when testing the PPP hypothesis and estimating the speed of PPP convergence. On the issue of testing the PPP hypothesis we argue that, even though cointegration would be a natural conceptual framework to define price convergence, cointegration theory has not yet been developed to deal with a large

[1] For a recent analysis supportive of the PPP model relatively to the random walk in terms of forecasting performance see Ca' Zorzi and Rubaszek (2012).

[2] For theoretical explanations of how to de-link the persistence of the real exchange rate and the stickiness of nominal prices, see the discussion in Engel et al. (2007) and Benigno and Benigno (2008).

[3] They also argue that using a pair-wise approach country pairs that are stationary can be, under certain conditions, estimated consistently.

(or even moderate) number of variables as required for PPP applications. We also prove that the short-cut of evaluating PPP in two steps, i.e. aggregating foreign prices as a first step and then running a unit root test, could lead to wrong conclusions even under large asymptotics, i.e. when both N and T are large.[4]

On the speed of convergence to PPP, we show that, depending on how individual economies are ordered across space, the aggregation of foreign variables could influence sizably the results, as half-life estimates could range between values close to the truth and infinity.

The structure of the remainder of this paper is as follows. Section 2 presents the theoretical arguments, starting with the definition of the global convergence hypothesis. Section 3 develops two Monte Carlo examples to illustrate the importance of spatial considerations for the half-life statistics debate on PPP. Section 4 investigates the empirical relevance of how foreign data is aggregated for the case of price convergence in the euro area, showing the sensitivity of half-lives estimates to the choice of weights. The final section provides our main conclusions while proofs are presented in the Appendix.

2 Testing convergence

Let us define the notion of international global convergence formally and then apply this conceptual framework to PPP. Let $\mathbf{x}_t = (x_{1t}, x_{2t}, ..., x_{Nt})'$ be $N \times 1$ vector of endogenous variables, where x_{it} is endogenous variable for country $i = 1, 2, ..., N$ at time $t = 1, 2, ..., T$. The null hypothesis that there is international global convergence is defined as follows:

$$\text{Convergence Hypothesis: } x_{it} - x_{jt} \text{ is stationary for all } i \neq j;\ i, j \in \{1, 2, ..., N\}. \tag{1}$$

If x_{it} represents the price level of country i (expressed in a common numeraire), the null hypothesis states that relative PPP holds over the long run for all country-pairs.[5]

The size of the cross section, N, matters from a methodological perspective:

i) if N is small (fixed) one can perform the traditional Johansen cointegration analysis by

[4] The same applies if a pair-wise approach is adopted.

[5] If x_{it} denotes the country-specific real output than the null hypothesis is that of relative output convergence across countries (the catching-up hypothesis).

estimating a VAR(p) model for $\mathbf{x}_t$,

$$\mathbf{x}_t = \sum_{\ell=1}^{p} \mathbf{\Phi}_\ell \mathbf{x}_{t-\ell} + \mathbf{u}_t, \tag{2}$$

where $\mathbf{\Phi}_\ell$ for $\ell = 1, 2, ..., p$ are $N \times N$ matrices of unknown reduced form coefficients and then conduct the standard rank reduction tests. If $\mathbf{x}_t$ is integrated of order one, the PPP hypothesis implies that there are $N-1$ cointegrating relations, i.e. $rank\,(\mathbf{\Pi}) = N-1$, where $\mathbf{\Pi} = \sum_{\ell=1}^{p} \mathbf{\Phi}_\ell - \mathbf{I}_N$ and one stochastic trend in $\mathbf{x}_t$.

ii) If N is large the traditional Johansen (1991) approach can not be applied due to the high dimensionality of the VAR(p) model in $\mathbf{x}_t$,[6] and the literature offers limited guidance on how to proceed.

In light of these difficulties, a typical short-cut consists in aggregating the foreign variables first and then applying a unit root test country by country. The theoretical motivation for this is that, given the null hypothesis (1), *any* relative variable

$$z_{wit} = x_{it} - \overline{x}_{wit}, \tag{3}$$

must be stationary, where $\overline{x}_{wit} = \mathbf{w}_i' \mathbf{x}_t = \sum_{j=1}^{N} w_{ij} x_{jt}$ is a weighted cross section average with the weights in the vector $\mathbf{w}_i = (w_{i1}, w_{i2}, ..., w_{iN})'$ satisfying the normalization condition

$$\sum_{j=1}^{N} w_{ij} = 1 \text{ for all } i. \tag{4}$$

The reverse implication also holds assuming certain conditions on the weights. The sufficient and necessary conditions are described in the following proposition.

Proposition 1 *Let* $\mathbf{W}$ *be arbitrary* $N \times N$ *matrix of weights satisfying the normalization condition (4),*

$$w_{ii} = 0, \text{ for } i = 1, 2, .., N, \tag{5}$$

and

$$rank\,(\mathbf{W}) = N - 1. \tag{6}$$

[6] For commonly available time dimensions of macroeconomic datasets, Johansen (1991) cointegration analysis is typically applied to no more than 6-7 endogenous variables.

Then, $x_{it} - x_{jt}$ is stationary for any pair of countries $i, j = 1, 2, ..., N$ if and only if $(\mathbf{I} - \mathbf{W})\mathbf{x}_t$ is stationary.

The proof is relegated to the Appendix. The rank condition (6) has the intuitive interpretation that no country (or a group of countries) is isolated from the rest of the world (Chudik and Straub, 2011).

To asses if relative PPP holds for a given country it is common to conduct a standard augmented Dickey-Fuller (ADF) tests by estimating the following regression:[7]

$$\Delta z_{wit} = \alpha_{wi} z_{wi,t-1} + \sum_{\ell=1}^{p_i} \beta_{wi\ell} \Delta z_{wi,t-\ell} + u_{wit}. \tag{7}$$

The choice of the weights may, however, affect the conclusions. We show this by designing an example where the ADF test fails to detect PPP even when it holds for all currency pairs.

Example 1 *Let*

$$\Delta x_{it} = u_{it},\ \textit{for } i = 1,\ \textit{and } t = 1, 2, ..., T, \tag{8}$$

and

$$\Delta x_{it} = -\alpha \left(x_{i,t-1} - x_{i-1,t-1} \right) + u_{it},\ \textit{for } i = 2, 3, ..., N,\ \textit{and } t = 1, 2, ..., T, \tag{9}$$

where $0 < \alpha < 1$, innovations, u_{it}, are independently and identically distributed (IID) with mean 0 and variance $0 < \sigma_u^2 < K$, and the starting values collected in the vector $\mathbf{x}_0 = (x_{1,0}, x_{2,0}, ..., x_{N,0})'$ are independent of u_{it} for any $i = 1, 2, ..., N$ and any $t = 1, 2, ..., T$ and distributed such that $\|Var(\mathbf{x}_0)\|_r < K$, where $\|.\|_r$ denotes the maximum absolute row sum matrix norm. There are $N - 1$ cointegrating relationships and 1 stochastic trend and all pairs of countries cointegrate. Suppose the vector $\mathbf{w}_i$ satisfies the following granularity conditions:

$$\begin{aligned} \|\mathbf{w}_i\| &= O\left(N^{-\frac{1}{2}}\right),\ \textit{uniformly in } i, & (10) \\ \frac{w_{ij}}{\|\mathbf{w}_i\|} &= O\left(N^{-\frac{1}{2}}\right)\ \textit{uniformly in } i \textit{ and } j. & (11) \end{aligned}$$

[7] α_{wi} depends also on the selection of the number of lags p, but we do not show this dependence explicitly to simplify the notations. Optimal selection criteria for the selection of p has been discussed in Ng and Perron (2001).

Consider the ADF regression applied to $z_{wit} = x_{it} - \overline{x}_{wit}$*,*

$$\Delta z_{wit} = -\alpha_{zwi} z_{wi,t-1} + \sum_{\ell=1}^{p_i} \psi_{zwi} \Delta z_{wi,t-\ell} + e_{zit}, \tag{12}$$

where p_i *is a truncation lag. The relative variable* z_{wit} *does not contain a unit root. Nevertheless, Lemma A.1 in Appendix establishes that* $Var\left(\overline{x}_{wit}\right)$ *is bounded by* $O\left(t^2/N\right)$*, which implies*

$$z_{wit} - x_{it} \overset{q.m.}{\rightarrow} 0 \text{ as } N, T \rightarrow \infty \text{ such that } T^2/N \rightarrow 0.$$

The process z_{it} *is arbitrarily close to the unit root process* x_{it} *under the large* N, T *asymptotics above. As a consequence* $\alpha_{zwi} \rightarrow 0$ *and the ADF test based on (12) would not reject the null hypothesis of unit root (under the asymptotics above) for any arbitrary weights* $\mathbf{w}$ *satisfying the granularity conditions and any* $i = 1, 2, ..., N$*.*

3 Half-lives estimates

In this section we will show that spatial effects may critically affect estimates of half-life deviations to PPP. To illustrate this point, let us consider two data generating processes (DGPs) that we shall call henceforth the "star" and the "line" models. They are designed so that price convergence holds for all country pairs and, hence, in effective terms. The two models differ for the way countries are connected: in the star model the connection takes place through the centre country, while in the line model countries are ordered along a line. Let us also denote the half-life statistics of the mean reversion for the relative variable z_{wit} as h_{wi} :

$$h_{wi} = \frac{\ln(0.5)}{\ln(1 + \alpha_{wi})}, \tag{13}$$

Note that h_{wi} depends on the vector $\mathbf{w}_i$, which is in general *unknown*. For both models we shall postulate that half of the convergence is completed within ten months, i.e. the process of convergence is relatively fast.

3.1 Data generating processes

More formally the two DGPs are defined as follows:

DGP1 (Star model)

$$\Delta x_{it} = u_{it}, \text{ for } i = 1 \tag{14}$$

$$\Delta x_{it} = -\alpha \left(x_{i,t-1} - x_{1,t-1}\right) + u_{it}, \text{ for } i = 2, 3..., N, \tag{15}$$

DGP2 (Line model) x_{it}, for $i = 1, 2, ..., N$ is generated according to equations (8)-(9).

In both DGPs, $u_{it} \sim IIDN\,(0, 1)$ and observations are generated for $t = -19, -18, ..., 0, 1, 2, ..., T$ with the starting values $\mathbf{x}_{-20} = 0$. The first 20 replications are discarded. We set $\alpha = \exp\left[\ln(0.5)/10\right] - 1$, which corresponds to 10 month half-live, assuming a monthly frequency. [8] The number of time periods is set to $T = 600$, which corresponds to 50 years of artificial data, the number of cross section units considered is $N = \{8, 12, 16, 20, 100\}$, and the number of Monte Carlo replications is $R_{mc} = 2000$. Our designed set-up has the following characteristics: (i) relative convergence holds for all pair of countries, (ii) we have half-century of data and (iii) the time required to achieve half of the adjustment is 10 months. We randomly generate $R_w = 10^3$ weights matrices for each of R_{mc} Monte Carlo replications.[9] We then compute the histogram of the estimates $\widehat{\alpha}_{wi}$ in the ADF regression (7) with intercepts included and the lags set equal to 8 (the integer part of $T^{1/3}$). We also compute the median of half-lives statistics, denoted $\widehat{h}_i^{med}$.

3.2 Monte Carlo findings

Figure 1 plots the histograms of $\widehat{\alpha}_{wi}$ (left charts) together with the histograms of the corresponding half-lives $\hat{h}_{wi}$ (right charts).

Panel A shows the findings for $N = 8$. For the star model the histogram of $\widehat{\alpha}_{wi}$ and $\hat{h}_{wi}$ are centered close to the true values of $\alpha \approx -0.067$ and $\hat{h}_{wi}$ is close to 10 months with a small time series bias. For the line model the histogram of $\widehat{\alpha}_{wi}$ peaks at around half of the true value while the histogram of $\hat{h}_{wi}$ peaks at thirty months, i.e. three times higher than the true value.

[8] Introducing a weak cross sectional dependence of the innovations (as defined in Chudik et al. (2011)) is unlikely to affect results sizably.

[9] We first generate $\varsigma_{ij}^{(r_w)} \sim IIDU\,[0, 1]$, and then set $w_{ii}^{(r_w)} = 0$ and $w_{ij}^{(r_w)} = \varsigma_i^{(r_w)} / \sum_{j=1, j\neq i}^{N} \varsigma_{ij}^{(r_w)}$, for $i \neq j$, $i, j = 1, 2, ..., N$ and for $r_w = 1, 2, ..., R_w$.

Panel B shows how the findings are affected when N is equal to 100. For the star model both histograms continue to be centered close to the true values. For the line model the peak of the histogram for $\widehat{\alpha}_{wi}$ shifts further away from the true value. The histogram of $\hat{h}_{wi}$ instead flattens out, as the maximum $\hat{h}_{wi}$ is an increasing function of the size of N.[10] For the line model the extent of underestimating the speed of adjustment therefore tends to grow with N.

To sum up, the experiments presented here could be considered as polar cases. For the star model the spatial aggregation bias is small and the choice of weights should not play a critical role. For the line model the bias of the estimated half-life statistics ranges depends on the weights and the size of N. Similar conclusions are reached by looking at Table 1, which presents the mean of $\widehat{h}_i^{med}$ for all values of N considered. The median half-life statistics stays unchanged in the star model, whereas it increases with N in the line model reaching about 48 months when $N = 100$.

4 Empirical Illustration

In this section we develop an empirical application for the 17 euro area countries to evaluate the sensitivity of the half-life estimates to the choice of weights. We briefly describe the data in Subsection 4.1, summarize the methodology in Subsection 4.2, and present the results in Subsection 4.3.

4.1 Data

We collect from the IMF International Financial Statistics (IFS) consumer prices and nominal exchange rates data. We adjust consumer price data for seasonality, using the same techniques described in the supplement to Pesaran et al. (2009).[11] Our focus is on the euro area member states for the period January 1997 until February 2011. Altogether we compile $T = 170$ observations per each of the $N = 17$ countries.

4.2 Methodology

We run a set of standard ADF regressions after generating a whole range of alternative weights. More specifically, for each on the individual economies, we randomly generate $R_W = 10^4$ country-

[10] Figure 1 truncates the histograms at 150 months and 15% of draws would point to even higher half-lives measures.

[11] This supplement is available at http://www.econ.cam.ac.uk/faculty/pesaran/SupplementMarch06.pdf.

specific weights and estimate the corresponding half-life statistics as above.[12] The number of lags in the ADF regression (7) is set equal to the integer part of $\left[T^{1/3}\right]$ (5 in this case) and intercepts are included in all regressions.

4.3 Estimation results

Our findings are summarized in Figure 2, which presents histograms of the estimated half-life statistics truncated at 10 years (120 months). We complement the analysis with two additional statistics, κ, i.e. the fraction of estimated half-life statistics larger than 10 years and $\widehat{h}_{\min}$, i.e. the lowest estimated half-life statistics. For several countries there is a large fraction of the half-life estimates that is higher than 10 years. At the same time in most cases $\widehat{h}_{\min}$ is below twelve months. The wide range of estimates highlights the sensitivity of the results to spatial considerations. The bias associated to half-life estimates could therefore be very large.

5 Conclusions

The presence and speed of international price convergence is a subject that has always fascinated economists. The key contributions of this paper are twofold. The first is to show that while, in principle, one may formalize the testing of PPP as a co-integration problem, the Johansen approach cannot be applied directly since the cross dimension N is too large. Taking the short-cut of aggregating foreign variables and applying then an ADF test could lead to falsely rejecting the PPP hypothesis (assuming both N and T are large) even when it holds. The second contribution of this paper is to show how large the range of of half-life estimates might be because of spatial considerations. Our conclusion is that the bias from aggregating foreign prices across countries could be potentially very important and should be evaluated together with other sources of bias already discussed in the literature.

[12] Weights are randomly generated in the same way as in the Monte Carlo section.

Table 1: Monte Carlo findings for the median half-life statistics.

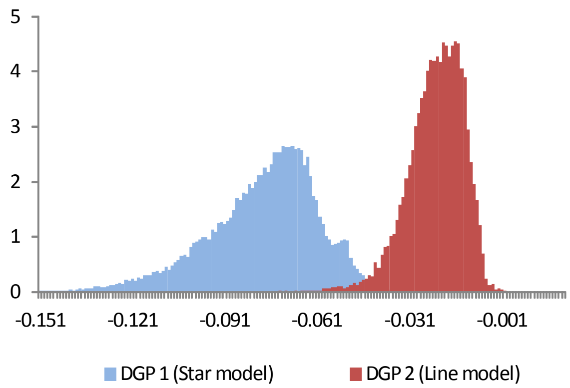
N :	8	12	16	20	100
DGP1 (Star model)	9.22	9.23	9.28	9.23	9.22
DGP2 (Line model)	21.34	29.57	37.89	43.89	48.17

Notes: This table reports $R_{mc}^{-1}\sum_{r=1}^{R_{mc}}\widehat{h}_{i,(r)}^{med}$, where $\widehat{h}_{i,(r)}^{med}$ is the median of $R_w = 10^4$ estimates of half-lives in the r-th Monte Carlo replication.

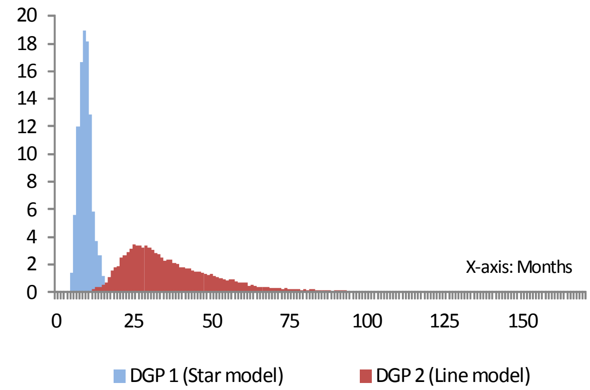
Figure 1: Histogram of $\widehat{\alpha}_{wi}$ (left charts) and the estimates of half-lives, $\widehat{h}_{wi}$ (right charts).

Panel A: Experiments with $N = 8$.

Histogram of $\widehat{\alpha}_{wi}$

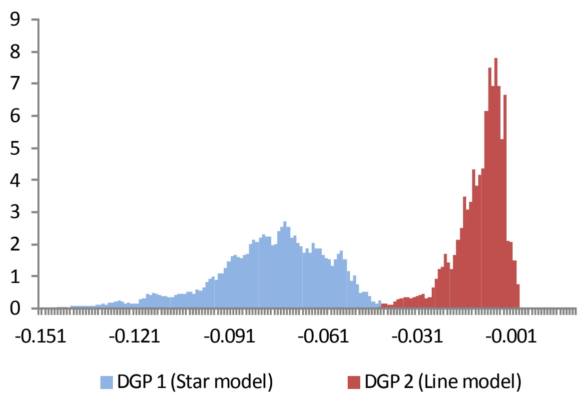


Histogram of half-lives $\widehat{h}_{wi}$(months)

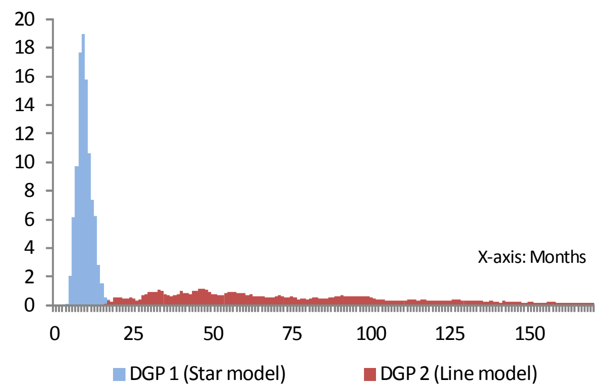


Panel B: Experiments with $N = 100$.

Histogram of $\widehat{\alpha}_{wi}$

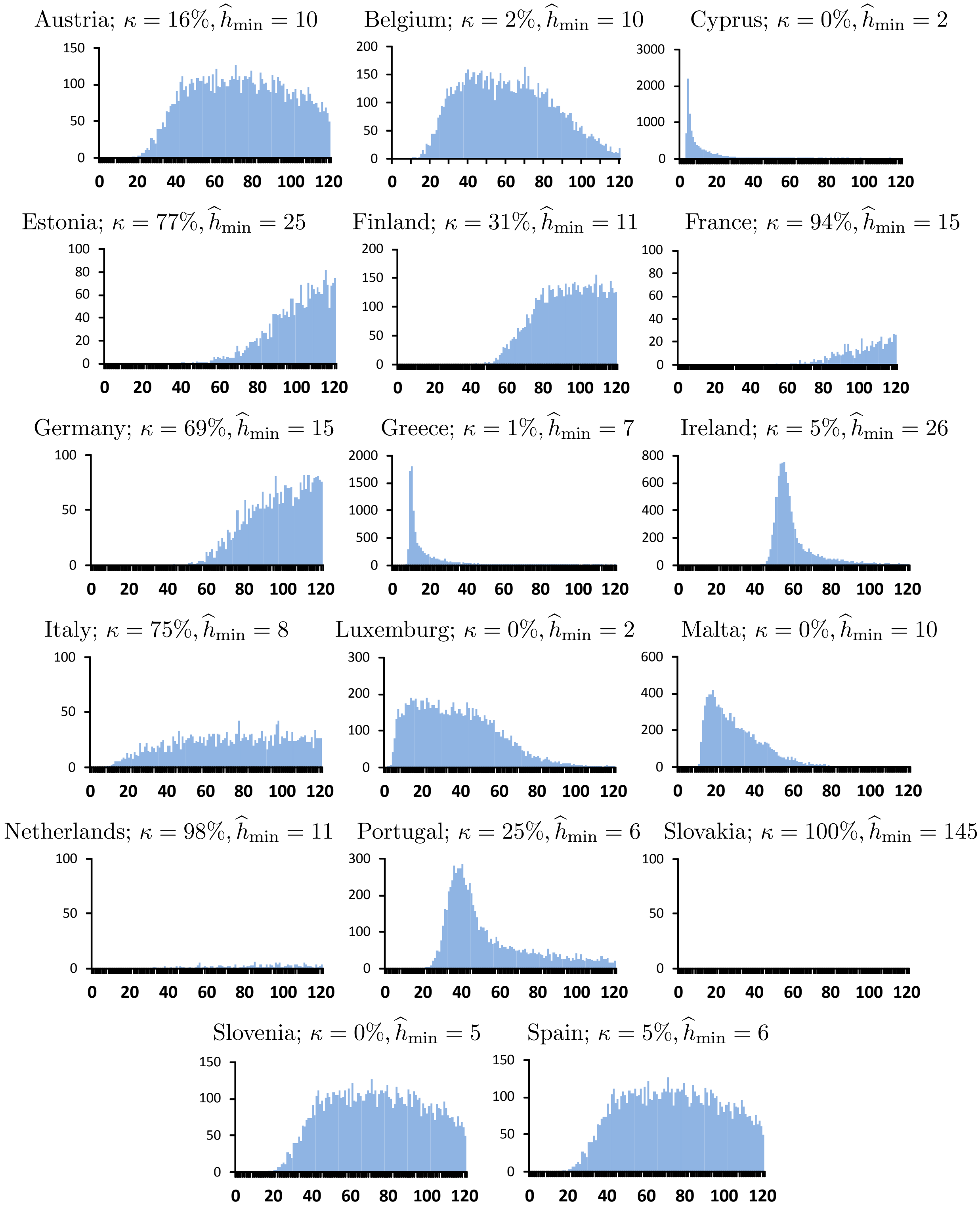


Histogram of half-lives $\widehat{h}_{wi}$(months)



Notes: $\alpha = \ln(0.5)/10$, which corresponds with 10 month half-live, and the number of time periods is $T = 600$, which is 50 years of artificial data. Further details of the MC design are described in Section 3.

Figure 2: Histogram of the estimates of half-lives in the case of price convergence in Euro Area



Notes: Constant κ shows fraction of estimates with half-live statistics larger than 10 years. $\widehat{h}_{\min}$ is the minimum estimated half-live statistics.

References

Benigno, G. and P. Benigno (2008). Exchange rate determination under interest rate rules. *Journal of International Money and Finance 27*, 971–993.

Ca' Zorzi, M. and M. Rubaszek (2012). Real exchange rate forecasting: A calibrated half-life PPP model can beat the random walk. National Bank of Poland Economic Institute Working Paper No. 123.

Chambers, M. (2005). The purchasing power parity puzzle, temporal aggregation, and half-life estimation. *Economics Letters 86*, 193–198.

Chi-Young, C., N. C. Mark, and D. Sul (2006). Unbiased estimation of the half-life to PPP convergence in panel data. *Journal of Money, Credit and Banking 38*, 921–938.

Chudik, A., M. H. Pesaran, and E. Tosetti (2011). Weak and strong cross section dependence and estimation of large panels. *Econometrics Journal 14*, C45–C90.

Chudik, A. and R. Straub (2011). Size, opennes and macroeconomic interdependence. Federal Reserve bank of Dallas, Globalization and Monetary Policy Institute Working Paper No. 103, December 2011.

Engel, C., N. C. Mark, and K. D. West (2007). Exchange rate models are not as bad as you think. NBER Working Paper No. 13318.

Imbs, J., H. Mumtaz, M. Ravn, and H. Rey (2005). PPP strikes back: Aggregation and the real exchange rate. *Quarterly Journal of Economics 120*, 1–43.

Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in gaussian vector autoregressive models. *Econometrica 59*, 1551–1580.

Ng, S. and P. Perron (2001). Lag length selection and the construction of unit root tests with good size and power. *Econometrica 69*, 1519–1554.

Nickell, S. (1981). Biases in dynamic models with fixed effects. *Econometrica 49*, 1417–1426.

Pesaran, M. H., R. P. Smith, T. Yamagata, and L. Hvozdyk (2009). Pairwise tests of purchasing power parity. *Econometric Reviews 28*, 495–521.

Rogoff, K. S. (1996). The purchasing power parity puzzle. *Journal of Economic Literature 34*, 647–668.

Taylor, A. M. and M. P. Taylor (2004). The purchasing power parity debate. *Journal of Economic Perspectives 18*, 135–158.

A Mathematical Appendix

Proof of Proposition 1. ($\Rightarrow$) Let $z_{ijt} = x_{it} - x_{jt}$ for $i, j = 1, 2, ..., N$. Stationarity of z_{ijt} for all $i \neq j$, $i, j = 1, 2, ..., N$, implies that a linear combination $z_{wit} = \sum_{j=1}^{N} w_{ij} z_{ijt}$ is stationary for all $i = 1, 2, ..., N$, where w_{ij} are the individual elements in the row i of $\mathbf{W}$ satisfying the normalization condition (4) and the condition (5), regardless whether the rank condition (6) hold. Hence $\mathbf{z}_{wt} = (z_{w1t}, z_{w2t}, ..., z_{wNt})' = (\mathbf{I} - \mathbf{W})\,\mathbf{x}_t$ is stationary, as desired.

($\Leftarrow$) Now consider the reverse implication and suppose that $\mathbf{z}_{wt} = (\mathbf{I} - \mathbf{W})\,\mathbf{x}_t$ is stationary for an $N \times N$ matrix of weights $\mathbf{W}$ satisfying the normalization condition (4), condition (5) and the rank condition (6). Let us partition $\mathbf{z}_{wt} = (z_{w1t}, \tilde{\mathbf{z}}_{w,-1,t})$, where $\tilde{\mathbf{z}}_{w,-1,t} = (z_{w2t},, z_{w3t}, ..., z_{wNt})'$ is an $N - 1 \times 1$ vector of stationary variables,

$$\mathbf{A} = (\mathbf{I} - \mathbf{W}) = \left(\begin{array}{c|c} a_{00} & \mathbf{a}'_{1\cdot} \\ \hline \mathbf{a}_{\cdot 1} & \mathbf{A}_{22} \end{array} \right),$$

and also define $N - 1 \times 1$ vector of relative variables with respect to country 1 taken as a numeraire $\mathbf{r}_{1t} = (x_{2t} - x_{1t}, x_{3t} - x_{1t}, ..., x_{Nt} - x_{1t})'$. First we show that $rank\,(\mathbf{A}_{22}) = N - 1$ by contradiction. Suppose for moment that $\mathbf{A}_{22}$ is rank deficient. Then there exists $N - 1 \times 1$ vector $\boldsymbol{\vartheta}_2 \neq 0$ such that $\mathbf{A}_{22}\boldsymbol{\vartheta}_2 = 0$. But then

$$\mathbf{A} \begin{pmatrix} 0 \\ \boldsymbol{\vartheta}_2 \end{pmatrix} = 0,$$

in addition to

$$\mathbf{A}\boldsymbol{\tau} = (\mathbf{I} - \mathbf{W})\,\boldsymbol{\tau} = \boldsymbol{\tau} - \mathbf{W}\boldsymbol{\tau} = \mathbf{0},$$

which contradicts (6). This establish $\mathbf{A}_{22}$ has full rank. Next, noting that

$$\mathbf{A}\boldsymbol{\tau} x_{1t} = 0,$$

we have

$$\mathbf{z}_{wt} = \begin{pmatrix} z_{w1t} \\ \tilde{\mathbf{z}}_{w,-1,t} \end{pmatrix} = \mathbf{A}\mathbf{x}_t = \mathbf{A}\,(\mathbf{x}_t - \boldsymbol{\tau} x_{1t}) = \left(\begin{array}{c|c} a_{00} & \mathbf{a}'_{1\cdot} \\ \hline \mathbf{a}_{\cdot 1} & \mathbf{A}_{22} \end{array} \right) \begin{pmatrix} 0 \\ \mathbf{r}_{1t} \end{pmatrix}.$$

The full rank of $\mathbf{A}_{22}$ implies $\mathbf{A}_{22}^{-1}$ exists and

$$\mathbf{r}_{1t} = \mathbf{A}_{22}^{-1}\tilde{\mathbf{z}}_{w,-1,t}.$$

But $\tilde{\mathbf{z}}_{w,-1,t}$ is a stationary process and therefore $\mathbf{r}_{1t}$ is stationary as well. Since all relative variables are stationary when using the numeraire x_{1t} it follows that for any other numeraire j, we have $z_{ijt} = x_{it} - x_{jt} = (x_{it} - x_{1t}) - (x_{jt} - x_{1t})$ is stationary as well. This completes the proof. ■

Lemma A.1 *Let x_{it}, for $i = 1, 2, ..., N$, and $t = 1, 2, ..., T$ be given by Example 1. Then there exists a constant $K < \infty$, which does not depend on N and t and such that*

$$Var\left(\sum_{j=1}^{N} w_{ij}x_{jt}\right) \leq K\frac{t^2}{N},$$

for any $N = 1, 2, ..$ and any t.

Proof. We have

$$Var\left(\sum_{j=1}^{N} w_{ij}x_{jt}\right) = \mathbf{w}_i'Var\left(\mathbf{x}_t\right)\mathbf{w}_i, \tag{A.1}$$

for any $i = 1, 2, ..., N$, where $\mathbf{w}_i = (w_{i1}, w_{i2}, ..., w_{iN})'$, and

$$Var\left(\mathbf{x}_t\right) = \sum_{\ell=0}^{t-1} \mathbf{\Phi}^{\ell}E\left(\mathbf{u}_{t-\ell}\mathbf{u}_{t-\ell}'\right)\mathbf{\Phi}^{\ell\prime} + Var\left(\mathbf{x}_0\right), \tag{A.2}$$

in which

$$\mathbf{\Phi} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ \alpha & 1-\alpha & 0 & \cdots & 0 \\ 0 & \alpha & 1-\alpha & & 0 \\ \vdots & & \ddots & \ddots & \\ 0 & 0 & & \alpha & 1-\alpha \end{pmatrix}. \tag{A.3}$$

Using the assumptions on the reduced form errors in Example 1, we have $E\left(\mathbf{u}_{t-\ell}\mathbf{u}_{t-\ell}'\right) = \sigma_u^2\mathbf{I}_N$, and (A.2) simplifies to

$$Var\left(\mathbf{x}_t\right) = \sigma_u^2\sum_{\ell=0}^{t-1} \mathbf{\Phi}^{\ell}\mathbf{\Phi}^{\ell\prime} + Var\left(\mathbf{x}_0\right). \tag{A.4}$$

Substituting (A.4) for $Var(\mathbf{x}_t)$ in (A.1) and taking the maximum absolute row sum matrix norm gives the following upper bound

$$\begin{aligned} Var\left(\sum_{j=1}^{N} w_{ij}x_{jt}\right) &= \left\|\mathbf{w}_i' Var(\mathbf{x}_t)\mathbf{w}_i\right\|_r, \\ &\leq \left\|\mathbf{w}_i'\right\|_r \left\|Var(\mathbf{x}_t)\right\|_r \left\|\mathbf{w}_i\right\|_r \\ &\leq \left\|\mathbf{w}_i\right\|_c \left\|\mathbf{w}_i\right\|_r \left[\sigma_u^2 \sum_{\ell=0}^{t-1} \left\|\mathbf{\Phi}^\ell\right\|_r \left\|\mathbf{\Phi}^\ell\right\|_c + \left\|Var(\mathbf{x}_0)\right\|_r\right], \end{aligned}$$

where $\|.\|_r$ denotes the maximum absolute row sum matrix norm, $\|.\|_c$ denotes the maximum absolute column sum matrix norm and both inequalities follows from the submultiplicative property of matrix norms. Noting that the granularity conditions (10)-(11) imply

$$\begin{aligned} \left\|\mathbf{w}_i\right\|_c &\leq K \text{ (uniformly in } i\text{)}, \\ \left\|\mathbf{w}_i\right\|_r &\leq \frac{K}{N} \text{ (uniformly in } i\text{)}, \end{aligned}$$

and also noting that $\sigma_u^2 < K$ and $\left\|Var(\mathbf{x}_0)\right\|_r < K$ as assumed in in Example 1, we obtain the following upper bound on the variance of cross section averages,

$$Var\left(\sum_{j=1}^{N} w_{ij}x_{jt}\right) \leq \frac{K}{N}\sum_{\ell=0}^{t-1}\left\|\mathbf{\Phi}^\ell\right\|_r \left\|\mathbf{\Phi}^\ell\right\|_c + \frac{K}{N}, \tag{A.5}$$

for some constant $0 < K < \infty$ that does not depend on i, t nor on N. It is useful to split the matrix $\mathbf{\Phi}^\ell$ into the first column, denoted as $\boldsymbol{\phi}_a^{(\ell)}$, and the remaining submatrix where the first column is replaced by zero vector, denoted as $\mathbf{\Phi}_b^{(\ell)}$, so that

$$\mathbf{\Phi}^\ell = \boldsymbol{\phi}_a^{(\ell)}\mathbf{s}_1' + \mathbf{\Phi}_b^{(\ell)},$$

where $\mathbf{s}_1 = (1, 0, 0, ..., 0)'$. Using the definition of $\mathbf{\Phi}$ in (A.3), we obtain the following upper bounds on $\boldsymbol{\phi}_a^{(\ell)}$ and $\mathbf{\Phi}_b^{(\ell)}$,

$$\left\|\boldsymbol{\phi}_a^{(\ell)}\right\|_r \leq 1, \ \left\|\boldsymbol{\phi}_a^{(\ell)}\right\|_c \leq \ell, \tag{A.6}$$

and

$$\max\left\{\left\|\boldsymbol{\Phi}_b^{(\ell)}\right\|_r, \left\|\boldsymbol{\Phi}_b^{(\ell)}\right\|_c\right\} \leq \sum_{n=0}^{\ell} \begin{pmatrix} \ell \\ n \end{pmatrix} \alpha^n (1-\alpha)^{\ell-n} = 1, \tag{A.7}$$

for $\ell = 1, 2, ...$ Note that the triangle inequality imply $\left\|\boldsymbol{\Phi}^{\ell}\right\| = \left\|\boldsymbol{\phi}_a^{(\ell)}\mathbf{s}_1' + \boldsymbol{\Phi}_b^{(\ell)}\right\| \leq \left\|\boldsymbol{\phi}_a^{(\ell)}\mathbf{s}_1'\right\| + \left\|\boldsymbol{\Phi}_b^{(\ell)}\right\|$ for any matrix norm $\|.\|$. Therefore, using the upper bounds (A.6)-(A.7) in (A.5) and also noting that $\sum_{\ell=1}^{t} \ell = t(t-1)/2$ yields

$$Var\left(\sum_{j=1}^{N} w_{ij} x_{jt}\right) \leq \frac{K}{N} t^2,$$

as desired. ■