

Forthcoming, *Empirical Economics*

The Purchasing Power Parity Puzzle is Worse Than You Think

Christian J. Murray and David H. Papell

University of Houston

Abstract

The “purchasing power parity puzzle” is the difficulty of reconciling very high short-term volatility of real exchange rates with very slow rates of mean reversion. The strongest evidence of slow mean reversion comes from least squares estimates of first-order autoregressive models of the long-horizon dollar-sterling real exchange rate. Using median-unbiased estimation methods, we show that these methods *underestimate* the half-lives of PPP deviations, and thus *overestimate* the speed of mean reversion. When the specification is amended to allow for serial correlation, the speed of mean reversion falls even further. This makes resolution of the purchasing power parity puzzle more problematic.

Correspondence to:

Department of Economics, University of Houston, Houston, TX 77204-5019

Christian Murray, phone: (713) 743-3835, fax: (713) 743-3798, e-mail: cmurray@mail.uh.edu

David Papell, phone: (713) 743-3807, fax: (713) 743-3798, e-mail: dpapell@uh.edu

We thank Lutz Kilian, James Lothian, Mark Taylor, and two anonymous referees for helpful comments and suggestions.

1. Introduction

Rogoff's (1996) "purchasing power parity puzzle," the difficulty of reconciling very high short-term volatility of real exchange rates with very slow rates of mean reversion, has provided the catalyst for much recent work on PPP. The strongest evidence of slow rates of mean reversion comes from the long-horizon dollar-sterling real exchange rate. Lothian and Taylor (1996), using two centuries of dollar-sterling and franc-sterling real exchange rate data, find evidence of mean-reverting behavior for both rates, with half-lives of PPP deviations of just under 6 years for dollar-sterling and 3 years for franc-sterling. The Lothian and Taylor results for the dollar-sterling exchange rate, where PPP deviations damp out at a rate of only 10% per year, constitute the slowest speed of reversion cited by Rogoff. In other words, the dollar-sterling rate provides the upper bound of Rogoff's remarkable consensus of 3 – 5 year half-lives.

In this paper, we argue that the interpretation of five years as an upper bound for the dollar-sterling half-life is unwarranted, and is indeed closer to a lower bound. We show that Lothian and Taylor (LT) underestimate the half-lives of PPP deviations, and thus overestimate the speed of mean reversion, for the dollar-sterling rate. The first part of our analysis is purely statistical. LT conduct unit root tests and calculate half-lives using a first order autoregressive specification for the real exchange rate. Estimates of persistent AR(1) processes are biased downward, with the bias becoming larger as the degree of persistence increases. Using Andrews' (1993) exactly median-unbiased estimation methods, we find that the half-life for the dollar-sterling real exchange increases from 5.78 to 6.58 years. The point estimate of the half-life, however, is an incomplete measure of the information from the autoregression. We calculate 95% confidence intervals for the median-unbiased half-life, and find a lower bound of 3.72 years and an upper bound of 34.31 years.

LT rely on visual examination of the correlogram and partial correlogram in their choice of an AR(1) process. Using the most recent methods for lag selection in unit root tests, the general-to-specific (GS) criterion in Hall (1994) and Ng and Perron (1995) and the modified Akaike information criterion (MAIC) of Ng and Perron (2001), we find that the AR(1) specification for the dollar-sterling series needs to be augmented by either five (GS) or seven (MAIC) first differences to account for serial correlation.

With these specifications, we need to use the approximately median-unbiased estimation methods of Andrews and Chen (1994). With higher order AR models, half-lives can be calculated from either the sum of the AR parameters or the impulse response function. The half-life from the point estimate of the sum of the AR parameters for the dollar-sterling real exchange rate is 11.20 years (GS) and 16.98 years (MAIC). The half-life calculated from the impulse response function is 4.58 years (GS) and 11.23 years (MAIC). With one exception, which we will show below is a poor measure of persistence, all of these half-lives are much larger than those calculated from the AR(1) specification. The lower bounds of the 95% confidence intervals range from 2.66 to 4.98 years, and the upper bounds are all infinite.¹

Rogoff (1996) describes a rate of 15% per year at which deviations from PPP die out as “glacial.” Our results from the AR(1) specification of the dollar-sterling rate, which exactly replicates the regression reported in LT, are that we are 95% sure that PPP deviations die out at a rate between 2% and 18% per year. With the higher-order AR models, we can only say that we are 95% sure that PPP deviations die out at a rate between 0% and 15% per year (GS) or 0% and 13% per year (MAIC). The conclusion is the same for both specifications. An “extremely slow rate at which shocks appear to damp out” of 15% per year, which forms the basis of the PPP

puzzle, is better interpreted as a maximum than as a median of the estimated speed of mean reversion for the long-horizon dollar-sterling real exchange rate.

In addition, we demonstrate that for higher order AR models, the standard measure of persistence, the “half-life,” is incomplete and potentially misleading. The choice of “one-half” is arbitrary, and has important implications for the measured persistence of the dollar-sterling rate. For one specification chosen by a commonly used model selection criterion, the half-life calculated from the impulse response function seriously understates the speed of mean reversion for the dollar-sterling rate. We therefore argue against blindly calculating half-lives, and advocate inspecting the entire impulse response function.

2. Median-Unbiased Estimates and Confidence Intervals

The data used by LT are the dollar-sterling real exchange rate (1791-1990) and the franc-sterling real exchange rate (1803-1990). LT estimate the following Dickey-Fuller regression:

$$q_t = c + \alpha q_{t-1} + u_t, \quad (1)$$

where q_t is the sterling denominated real exchange rate.² The half-life in equation (1), defined as the number of periods required for a unit shock to dissipate by one half, is calculated as $\ln(0.5)/\ln(\alpha)$. Based on the least squares estimates of α , in equation (1), LT find half-lives of deviations from PPP to be 5.78 years and 2.73 years for the dollar-sterling and franc-sterling exchange rates respectively.³

A problem with the least squares estimator of α in equation (1) is that it is both downward mean-biased and downward median-biased. As such, the least squares estimates of

¹ Murray and Papell (2002) use median-unbiased estimation techniques to investigate the behavior of annual century long and quarterly post-1973 real exchange rates.

² The implications for unit roots and mean reversion of including a time trend in (1) are discussed by Cuddington and Liang (2000) and Lothian and Taylor (2000).

the half-life, $\ln(0.5)/\ln(\alpha_{LS})$, understate the persistence of deviations from PPP. We correct for the bias in the least squares estimates by calculating Andrews' (1993) exactly median-unbiased estimator of α . We prefer median-unbiased estimators to mean-unbiased estimators in this context. Since median-unbiasedness is preserved under monotonic transformations, this procedure gives us a median-unbiased estimator of the half-life as well. In addition, the coverage probability of confidence intervals is preserved under monotonic transformations. Neither of these properties hold for mean-unbiased estimators.⁴

For the AR(1) specification of the dollar-sterling rate, we report the median-unbiased estimate of the half-life, and the 95% confidence interval, in the first row of Table 1. The median-unbiased half-lives are higher than their least squares counterparts. Correcting the bias raises the estimate of the half-life from 5.78 to 6.58 years for the dollar-sterling rate. This alone puts the half-life well above Rogoff's 3-5 year "remarkable consensus". In addition, the point estimates provide an incomplete picture of the persistence of deviations from PPP. The 95% confidence interval for the dollar-sterling half-life is [3.72, 34.31]. This is equivalent to the statement that we are 95% certain that shocks to the dollar-sterling rate decay at a rate between 2% and 18% per year. Recall that Rogoff (1996) describes a 15% rate of decay as "glacial." Our results for the dollar-sterling rate are consistent with a much slower rate of convergence to PPP.

The results for the franc-sterling rate are not as dramatic. Correcting the bias raises the estimate of the half-life from 2.73 to 2.94 years, just below Rogoff's consensus. The 95% confidence interval for the franc-sterling half-life is [1.94, 5.95] years. In other words, we are

³ Rogoff (1996) ascribes half-lives in LT of 4.7 and 2.5 years. These, however, are from the autoregressive term in an ARMA (1,1) model, not from the DF regression.

⁴ See Andrews (1993) and Murray and Papell (2002) for further discussion.

95% certain that shocks to the franc-sterling rate decay at a rate between 11% and 30% per year, not nearly as sluggish as the dollar-sterling rate. This is in accord with Andrews' result that the bias in the least squares estimates is less severe the further the estimates are from unity.

The AR(1) specification in the Dickey-Fuller regression in equation (1) is only justified if the error terms are serially uncorrelated. Based on data dependent model selection criteria, we find this assumption appropriate for the franc-sterling rate, but untenable for the dollar-sterling rate. To allow for serial correlation, we estimate the following Augmented Dickey-Fuller regression:

$$q_t = c + \alpha q_{t-1} + \sum_{i=1}^k \psi_i \Delta q_{t-i} + u_t. \quad (2)$$

We choose the lag length, k , via the general-to-specific (GS) criterion of Hall (1994) and Ng and Perron (1995) and the modified Akaike information criterion (MAIC) of Ng and Perron (2001). The lag length chosen is 5 (GS) and 7 (MAIC) for the dollar-sterling rate and 0 (both) for the franc-sterling rate. We note that, in contrast to LT, when we allow for serial correlation in the errors we are unable to reject the null hypothesis that PPP does not hold for the dollar-sterling real exchange rate using either GS or MAIC.⁵

When estimating an Augmented Dickey-Fuller regression, the median-unbiased estimates of α are no longer exact, but have to be approximated. We use the techniques of Andrews and Chen (1994) to calculate approximately median-unbiased estimates of α in equation (2), as well as 95% confidence intervals, for the dollar-sterling rate.⁶ Since the parameter α in equation (2) is the sum of the coefficients in the underlying AR model, one method that has been used for

⁵ This result (with a lag length of 5) is reported in Hegwood and Papell (1998). James Lothian has pointed out to us that, for the dollar-sterling real exchange rate, the unit root null is rejected (at the 5% level) with 0 to 4 lags, but not rejected (at the 10% level) with 5 to 8 lags.

calculating the half-life is to again use the formula $\ln(0.5)/\ln(\alpha)$. We report point estimates and 95% confidence intervals for median-unbiased estimates of $\ln(0.5)/\ln(\alpha)$ in Table 1, for $k = 0 - 8$. The point estimate of the half-life based on α is 11.20 years (GS) and 16.98 years (MAIC), much larger than we find for the AR(1) case. The 95% confidence interval is $[4.27, \infty)$ for GS and $[4.98, \infty)$ for MAIC. The lower bounds of the confidence intervals are consistent with PPP deviations that decay at a rate of 15% and 13% per year for GS and MAIC respectively. These lower bounds are consistent with speeds of adjustment equal to or slower than Rogoff's "glacial" estimate. The upper bounds of the confidence intervals are consistent with PPP deviations not decaying at all.

We also note that in Table 1 there is a perfect correspondence between the outcome of the unit root test and the upper bound of the half-life confidence interval. For lags 0 – 4, where the unit root null is rejected at the 5% level, the upper bounds of the 95% half-life confidence interval are finite. For lags 5 – 8, where we fail to reject the unit root null at the 5% level, the upper bounds of the 95% half-life confidence interval are infinite.

The half-life calculated from the value of α assumes that shocks to real exchange rates decay at a constant rate. While this is appropriate for an AR(1) process ($k = 0$), in general shocks to a higher order autoregression will not decay at a constant rate. In light of this, Cheung and Lai (2000) recommend calculating the half-life directly from the impulse response function. The half-life is defined as the number of years required for deviations from PPP to subside permanently below one half in response to a unit shock.⁷

⁶ Since the lag length for the franc-sterling rate is chosen to be 0 using either criterion, there is no reason to use these methods.

⁷ For a first order autoregression, half-lives calculated from α are the same as half-lives calculated from the impulse response function.

We now turn to half-lives based on the impulse response function, which are also reported in Table 1, along with their 95% confidence intervals.⁸ The half-life calculated from the impulse response function is 4.58 years (GS) and 11.23 years (MAIC). The 95% confidence interval is $[3.10, \infty)$ for GS and $[2.66, \infty)$ for MAIC. These confidence intervals are consistent with nearly the entire admissible range of half-lives. While the point estimate based on the impulse response function for the value of k chosen by the MAIC is in accord with the point estimate of the half-life based on α , the point estimate for GS appears to have substantially lowered our estimate of the persistence of shocks to the dollar-sterling rate. However, inspection of the impulse response function reveals that this estimate is misleading, and demonstrates the potential arbitrariness of the “half-life” as a measure of persistence.

Figure 1 plots the first 100 terms of the impulse response function along with its 95% confidence interval for $k = 1 - 8$. We include a horizontal line at one half for reference. We also include the half-life based on the median-unbiased estimate of α , which assumes that shocks die out at a constant rate. For $k = 1 - 4$, the rate of decay of the impulse response function is nearly constant.⁹ For this reason, calculating the half-life based either on α or the impulse response function leads to qualitatively similar point estimates, although the confidence intervals for the impulse response functions are tighter than the confidence intervals based directly on α .

We see that the rate of decay of the impulse response function for lags $k = 5 - 8$ is clearly not constant, so that the half-life based on α can be misleading. For $k = 7$, the value chosen by the MAIC, the impulse response falls below one half at 4.39 years and reaches a minimum of

⁸ When $k > 0$, the median-unbiased estimates of α and the half-life are approximate. However, simulation evidence has shown that the approximation is quite accurate. See Lopez, Murray, and Papell (2004) for further details.

0.263 at 7 years. It then begins to increase, reaching a value of 0.501 at 11 years, and then eventually decaying to zero. The results for $k = 6$ and 8 are qualitatively similar.

The fact that shocks do not die out at a constant rate, however, makes the half-life based on the impulse response function a poor measure of persistence when $k = 5$, the value chosen by the GS criterion. In particular, it is the “half” in the “half-life” which yields such a small estimate of persistence. The impulse response function falls below one-half at 4.58 years and reaches a minimum of 0.396 at 5 years. However, it then begins to increase, reaching a value of 0.485 at 10 years, and then eventually decaying to zero. This low measure of persistence in this particular case depends on the arbitrary choice of “one-half.”

Cheung and Lai (2000) argue that, if the impulse response function is hump-shaped, rising for a few periods before falling, it is preferable to calculate half-lives from the impulse response function rather than from the value of α . The type of non-monotonic impulse response function in this paper is different. The initial movement is always downward, and the non-monotonic part occurs later. In this particular case, reporting half-lives based on the impulse response function is arguably more inappropriate than reporting half-lives based on α . When shocks do not decay at a constant rate, it is advisable to view the entire impulse response function rather than looking at any particular measure of the half-life.

3. Conclusions

Lothian and Taylor (1996), using two centuries of data, provide some of the most compelling evidence for the consensus that long-horizon real exchange rates are mean-reverting but the speed of reversion is extremely slow. Their result that the half-life for PPP deviations is

⁹ Since the half-life based on α and the impulse response function are nearly indistinguishable for $k = 1 - 4$, we do not plot the former for this range of k .

5.78 years for the dollar-sterling real exchange rate is often cited in support of Rogoff's "purchasing power parity puzzle".

The message of this paper is that, although the LT half-life for dollar-sterling represents the *slowest* speed of reversion cited by Rogoff, it seriously *overestimates* the speed of reversion in the data. Using Andrews' (1993) median-unbiased estimation method for the AR(1) regression reported by LT, the half-life increases to 6.58 years. Furthermore, the 95% confidence interval for the dollar-sterling half-life is [3.72, 34.31]. The speed of mean reversion implied by the *lower* bound of the confidence interval, 18% per year, is only slightly faster than a speed of 15% per year that Rogoff called "glacial". If 15% per year is "glacial", the speed of mean reversion implied by the *upper* bound of the confidence interval, 2% per year, can arguably be called "geologic".

When the LT specification is amended to allow for serial correlation, the speed of mean reversion falls even further. Using the most recent methods for lag selection in unit root tests and inspecting the entire impulse response function, the point estimates approximately double. We cannot even say with 95% certainty that shocks to the dollar-sterling real exchange rate ever decay, as the upper bound of the confidence intervals are infinite. We conclude that the disparity between the persistence in real exchange rates and what can be explained by models with nominal rigidities is greater than Rogoff originally suggested. In other words, the PPP puzzle is worse than you think.

References

- Andrews, D.W.K, 1993, "Exactly Median-Unbiased Estimation of First Order Autoregressive/Unit Root Models," *Econometrica* 61: 139-165.
- Andrews, D.W.K., and H.-Y. Chen, 1994, "Approximately Median-Unbiased Estimation of Autoregressive Models." *Journal of Business and Economic Statistics* 12: 187-204.
- Cheung, Y.-W. and K.S. Lai, 2000, "On the Purchasing Power Parity Puzzle," *Journal of International Economics* 52: 321-330.
- Cuddington, J.T. and H. Liang, 2000, "Purchasing Power Parity Over Two Centuries?" *Journal of International Money and Finance* 19: 753-757.
- Hall, A., 1994, "Testing for a Unit Root in Time Series with Pretest Data-Based Model Selection," *Journal of Business and Economic Statistics* 12: 461-70.
- Hegwood, N. and D.H. Papell, 1998, "Quasi Purchasing Power Parity." *International Journal of Finance and Economics* 3: 279-289.
- Lopez, C., C.J. Murray, and D.H. Papell, 2004, "More Powerful Unit Root Tests and the PPP Puzzle," working paper, University of Houston.
- Lothian, J., and M. Taylor, 1996, "Real Exchange Rate Behavior: The Recent Float from the Perspective of the Past Two Centuries." *Journal of Political Economy* 104: 488-509.
- Lothian, J., and M. Taylor, 2000, "Purchasing Power Parity Over Two Centuries: Strengthening the Case for Real Exchange Rate Stability: A Reply to Cuddington and Liang." *Journal of International Money and Finance* 19: 759-764.
- Murray, C.J., and D.H. Papell, 2002 "The Purchasing Power Parity Persistence Paradigm." *Journal of International Economics* 56: 1-19.
- Ng, S. and P. Perron, 1995, "Unit Root Test in ARMA Models with Data Dependent Methods for the Selection of the Truncation Lag," *Journal of the American Statistical Association* 90: 268-281.
- 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.
- Rogoff, K., 1996, "The Purchasing Power Parity Puzzle." *Journal of Economic Literature* 34: 647-668.

**Table 1. Median-Unbiased Half-Lives in Augmented Dickey-Fuller Regressions
U.S. Dollar/British Pound Real Exchange Rate: 1791-1990**

$$q_t = c + \alpha q_{t-1} + \sum_{i=1}^k \psi_i \Delta q_{t-i} + u_t$$

k	Rejection Level	$HL_{\alpha, MU}$	95% CI	$HL_{IRF, MU}$	95% CI
0	1%	6.58	[3.72, 34.31]	6.58	[3.72, 34.31]
1	1%	5.95	[3.49, 16.98]	6.33	[3.31, 13.48]
2	1%	5.95	[3.49, 22.76]	6.22	[3.37, 12.84]
3	5%	6.58	[3.49, 34.31]	6.72	[3.25, 14.77]
4	5%	6.58	[3.49, 34.31]	6.09	[3.08, 15.41]
5	FTR	11.20	[4.27, ∞)	4.58	[3.10, ∞)
6	FTR	16.98	[4.60, ∞)	11.62	[2.87, ∞)
7	FTR	16.98	[4.98, ∞)	11.23	[2.66, ∞)
8	FTR	13.51	[4.60, ∞)	11.98	[2.98, ∞)

Notes: Rejection level refers to the significance level at which we reject unit root null hypothesis. FTR is failure to reject the unit root null. Confidence intervals for the impulse response functions are calculated as in Andrews and Chen (1994) and Murray and Papell (2002).

**Figure 1. Median-Unbiased Impulse Response Functions and 95% Confidence Intervals
U.S. Dollar/British Pound Real Exchange Rate: 1791-1990**

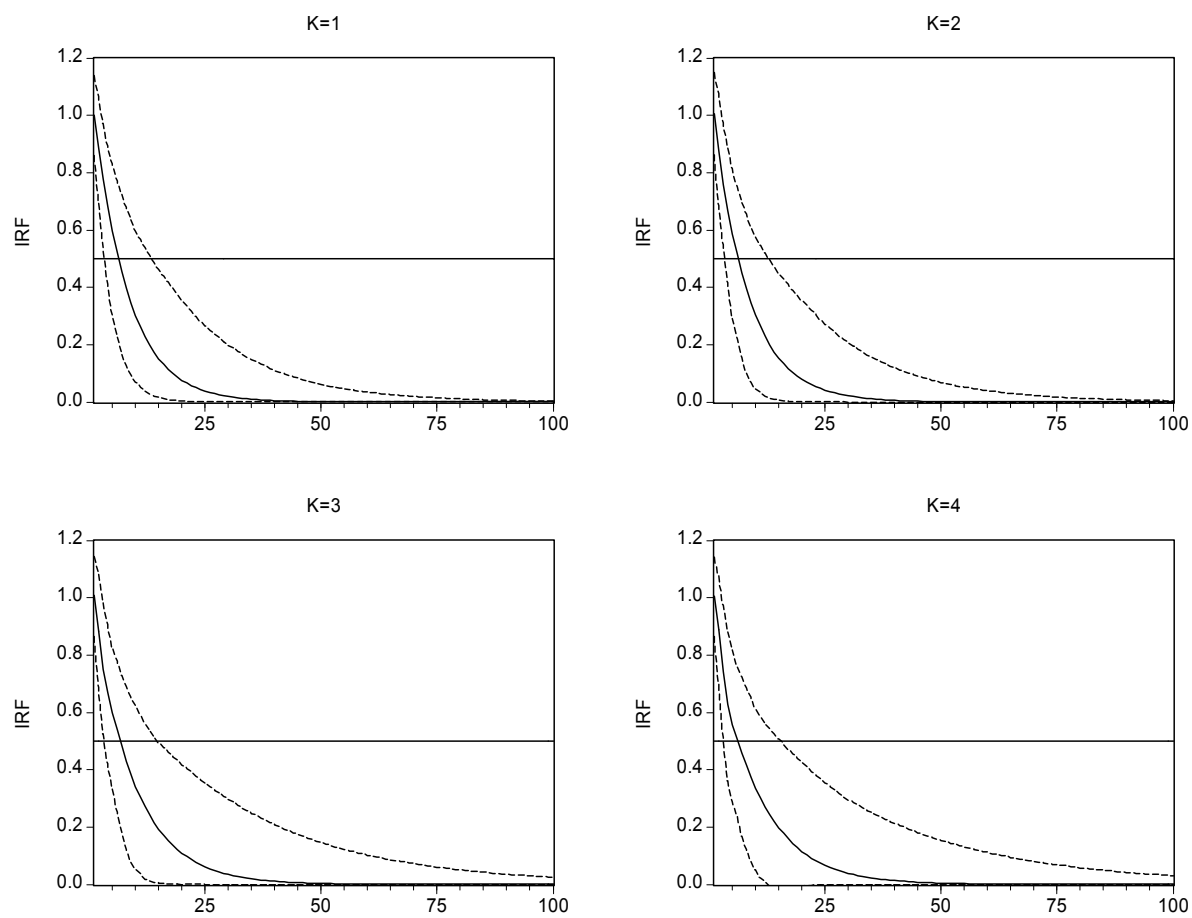


Figure 1. (Continued)

