State of the Art Unit Root Tests and the PPP Puzzle

Claude Lopez, Christian J. Murray, and David H. Papell

University of Houston

August 2003

Abstract

Using median-unbiased estimation, recent research has questioned the validity of Rogoff's "remarkable consensus" of 3-5 year half-lives of deviations from PPP. These estimates, however, are based on unit root tests with low power. We extend median-unbiased estimation to the efficient unit root test of Elliott, Rothenberg, and Stock (1996). We find that median-unbiased estimation based on the more powerful unit root test has the potential to tighten confidence intervals for half-lives. Using long horizon real exchange rate data, we find that the typical lower bound of the confidence intervals for median-unbiased half-lives is *above* 3 years. Thus, while previous confidence intervals for half-lives are consistent with virtually anything, our tighter confidence intervals now rule out economic models with nominal rigidities as candidates for explaining the observed behavior of real exchange rates. Therefore, while we obtain more information using efficient unit root tests on longer term data, this information moves us away from solving the PPP puzzle.

Correspondence:

Claude Lopez, Department of Economics, University of Cincinatti
email: claude.lopez@mail.uh.edu

Chris Murray, Department of Economics, University of Houston
tel: (713) 743 3835, fax: (713) 743-3798, email: cmurray@mail.uh.edu

David Papell, Department of Economics, University of Houston
tel: (713) 743 3807, fax: (713) 743-3798, email: dpapell@mail.uh.edu

For helpful comments and discussions we thank Lutz Kilian, Babara Rossi, Alan Taylor, seminar participants at Duke University, Texas Econometrics Camp 7, and the Southern Economic Association meetings.

1. Introduction

During the past decade, a number of studies using long-horizon data have changed the focus of research on Purchasing Power Parity (PPP) from the narrow question of whether or not the real exchange rate contains a unit root to the broader question of the persistence of deviations from PPP. Abuaf and Jorion (1990), Diebold, Husted and Rush (1991), Glen (1992), Cheung and Lai (1994), and Lothian and Taylor (1996) all reach the same conclusion: the hypothesis of a unit root in real exchange rates can be rejected and the half-life of the PPP deviations varies between 3 and 5 years.[1] In his well-known survey, Rogoff (1996) discusses the "remarkable consensus" of these half-lives and coins the phrase "purchasing power parity puzzle" to describe the difficulty in reconciling these slow speeds of adjustment with the high short-run volatility of real exchange rates. The slow speed of adjustment is problematic for models with nominal rigidities which predict faster convergence to PPP of 1 to 2 year half-lives.

Although the 3 to 5 year consensus has become the common starting point in attempts to "solve" the PPP puzzle, the consensus itself is problematic. The studies cited above generally calculate least squares point estimates of the half-lives from first order autoregressive processes. Point estimates alone do not provide a complete measure of persistence. Cheung and Lai (2000) supplement point estimates with conventional bootstrap confidence intervals in order to measure the precision of the half-life estimates. Their confidence intervals, however, are not valid under the unit root null and, even if long run PPP holds, are biased downwards in small samples.[2] In addition, the least squares estimates of half-lives are biased downward, providing an inaccurate picture of the speed of adjustment to PPP.

Two recent papers address these issues using classical estimation techniques.[3] Murray and Papell (2002) use the median-unbiased estimation methods of Andrews (1993) and the approximately median-unbiased methods of Andrews and Chen (1994) for Dickey-Fuller (DF) and Augmented Dickey-Fuller (ADF) regressions respectively. They calculate point estimates and confidence intervals for half-lives of PPP deviations for

[1] Engel (2000) raises the question of whether these rejections are caused by size distortions.

[2] See Kilian (1998), Hansen (1999), Kilian (1999), and Inoue and Kilian (2002) for further discussion of bootstrapping autoregressive processes with unit roots or near unit roots.

[3] See Kilian and Zha (2002) for a Bayesian perspective.

long-horizon annual (1990-1996) US dollar real exchange rate data for 6 countries and post-1973 quarterly US dollar real exchange rate data for 20 countries. Rossi (2002) uses the tests of Elliott and Stock (2001) and Hansen (1999) to calculate bias-corrected confidence intervals for half-lives of PPP deviations for post-1973 quarterly US dollar real exchange rate data for 17 countries. Despite the differences in methodology, the results for the post-1973 data in the two papers are nearly identical. The lower bounds of the 95% confidence intervals are mostly just above one year, while the upper bounds are generally infinite. These results, however, do not help "solve" the PPP puzzle. While the lower bounds are consistent with relatively fast convergence to PPP as predicted by models with nominal rigidities, the upper bounds are consistent with a unit root in real exchange rates and no convergence to PPP even in the very long run.

These results indicate that univariate methods are unlikely to be informative about the persistence of post-1973 real exchange rates.[4] Focusing on post-1973 rates, moreover, ignores most of the available data. While long-horizon data mixes fixed and flexible nominal exchange rate regimes and, therefore, cannot answer the question of whether PPP would hold with a century long flexible nominal exchange rate regime, it can potentially answer the question of whether PPP has held over the last century.

This potential has been greatly facilitated by the work of Taylor (2002), who develops real exchange rate data for over 100 years for 20 countries. Using the Elliott, Rothenberg and Stock (1996) generalized least squares detrended version of the Augmented Dickey-Fuller test (DF-GLS), which offers an increase in power over ADF tests, he concludes that there is strong evidence of PPP. This evidence is sufficient to lead Taylor to write: "If PPP holds in the long run, it is no longer productive to devote further attention to the stationarity question." An important contribution of Taylor's work is that, for the first time, it is possible to investigate PPP using long-horizon data with approximately the same set of advanced countries as is commonly used in studies with post-1973 data.

The purpose of this paper is to investigate the persistence of PPP deviations in long-horizon real exchange rates. We extend the methodology developed by Andrews (1993)

[4] Panel methods have been used extensively to test for unit roots in post-1973 real exchange rates. Murray and Papell (2003) examine persistence with panel methods. Elliott and Pesavento (2001) use unit root tests with stationary covariates to investigate PPP.

and Andrews and Chen (1994) for Augmented Dickey-Fuller tests to the more powerful DF-GLS test of Elliott, Rothenberg, and Stock (1996). We compute median-unbiased and approximately median-unbiased point estimates and confidence intervals for half-lives of PPP deviations for DF-GLS regressions. We use Taylor's (2002) data for 16 annual US dollar real exchange rates for developed countries with over a century of data for each country. To our knowledge, this is the first paper which corrects for median-bias in DF-GLS regressions.

While the focus of the paper is on persistence of PPP deviations rather than rejections or lack thereof of the unit root null, we need to address the latter subject first. Taylor (2002) reports much stronger rejections with DF-GLS tests than with ADF tests. He uses, however, a Lagrange Multiplier (LM) lag selection criterion that produces very short lag lengths. This causes two problems. First, the standard method to choose lag lengths in ADF tests is the general-to-specific method proposed by Hall (1994). As shown by Ng and Perron (1995), methods that choose too few lags lack power and can produce inappropriate non-rejections. Second, Ng and Perron (2001) show that the modified Akaike information criterion (MAIC) produces the best combination of size and power for DF-GLS tests. In this case, methods that choose too few lags are badly sized and can produce inappropriate rejections.

We start by reporting ADF and DF-GLS test results for the 16 real exchange rates, using the LM, GS, and MAIC criteria in each case. As would be expected from the research reported above, using GS lag selection for ADF tests raises the number of rejections and using MAIC lag selection for DF-GLS tests lowers the number of rejections. While the exact comparison depends on the choice of significance level, there is no overall qualitative difference between the number of rejections obtained with the two tests. Without including a time trend in the regression, the unit root null is rejected at the 5% level in only 8 (ADF) or 9 (DF-GLS) of 16 cases, which does not (to us) constitute strong evidence of PPP.

We next turn our attention to correcting for median-bias. Andrews (1993) shows how to calculate median-unbiased point estimates and confidence intervals for half-lives in DF regressions, and tabulates the bias for a range of parameter values and sample sizes. We conduct a similar tabulation for DF-GLS regressions. We find that, while the estimates from DF-GLS regressions are biased downwards, the extent of the bias is much less than in Dickey-Fuller regressions. In addition, the confidence intervals for median-unbiased estimators are tighter for DF-GLS regressions than for ADF regressions. This demonstrates the potential for sharper inference on the persistence of shocks to the real exchange rate than has been previously available.

We proceed to calculate median-unbiased point estimates and confidence intervals for half-lives of PPP deviations from the 16 long-horizon real exchange rates, using MAIC lag selection for DF-GLS regressions and GS lag selection for the ADF regressions. The point estimates of the half-lives are considerably larger than would be expected based on Rogoff's 3-5 year "consensus". The median value (among the 16 rates) is 11.34 years for the DF-GLS regressions and 7.55 years for the ADF regressions. The upper bounds of the 95% confidence intervals are infinite in a number of cases. The median lower bounds of the 95% confidence intervals are 3.02 years for the DF-GLS regressions and 3.25 years for the ADF regressions.

The major result in both Murray and Papell (2002) and Rossi (2002) is that, for both annual long-horizon and quarterly post 1973 real exchange rates, the confidence intervals of the half lives were so wide as to be consistent with virtually anything. We find a very different result here. With both DF-GLS and ADF regressions, the median lower bounds of the 95% confidence intervals are *over* three years. Since the half-lives that would be predicted from models with nominal rigidities are generally 1 to 2 years, our results are clearly inconsistent with the predictions from such models. Therefore, while we obtain greater information about the persistence of shocks to the real exchange rate, the PPP puzzle becomes even more problematic.

2. The Data and Unit Root Tests

Taylor (2002) collects nominal exchange rate and price level data through 1996 for 20 countries, each for over 100 years, yielding 19 US dollar denominated real exchange rates. The price levels are consumer price deflators or, if not available, GDP deflators. We extend Taylor's data through 1998, and omit Argentina, Brazil, and Mexico, in order to focus solely on developed countries. This leaves us with 16 dollar denominated real exchange rates: Australia, Belgium, Canada, Denmark, Finland, France, Germany, Italy, Japan, the Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, and the United Kingdom.

Before discussing the persistence of shocks to the real exchange rate, we first reconsider the results of Taylor's (2002) unit root tests. He reports results from ADF and DF-GLS tests, with and without deterministic time trends, and concludes that "… PPP has held in the long run over the twentieth century for my sample of twenty countries."

The ADF test, without a deterministic time trend, runs the following regression:

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

where q_t is the natural logarithm of the real exchange rate. This regression includes k lagged first differences to account for serial correlation.

The more powerful DF-GLS test runs the following auxiliary regression:

$$q_t^{\mu} = \alpha q_{t-1}^{\mu} + \sum_{i=1}^{k} \psi_i \Delta q_{t-i}^{\mu} + u_t \tag{2}$$

where q_t^{μ} is the GLS demeaned real exchange rate. That is, $q_t^{\mu} = q_t - \tilde{\beta} z_t$, where $z_t = 1$, $\tilde{\beta} = \left(\sum \tilde{z}_t^2\right)^{-1} \sum \tilde{z}_t \tilde{q}_t$, $\tilde{q}_t = (q_1, (q_2 - \alpha q_1), \ldots, (q_T - \alpha q_{T-1}))'$, $\tilde{z}_t = (1, (1-\alpha), \ldots, (1-\alpha))'$, $\alpha = 1 + c/T$, and $c = -7$.

We question Taylor's conclusion that long run PPP holds for two reasons. First, we find the conclusion too strong given his test results. For the real exchange rates that we consider in this paper, Taylor finds that with the ADF test, 6 of 16 unit root tests are rejected at the 5% level, while with the DF-GLS test, 10 of the 16 unit root tests are rejected at the 5% level. The increased rejection rate with the DF-GLS test is attributed to the higher power of the test. While the rejection rate for the DF-GLS test is over 50%, it is certainly not strong enough to support conclusion that long run PPP holds for all 16 countries, or that "it is no longer productive to devote further attention to the stationarity question."

Second, Taylor uses a Lagrange Multiplier (LM) criterion to choose the number of lags in the unit root regressions, and in most cases the selected lag is zero. Much research has been devoted to the topic of lag selection in unit root tests, and to our knowledge, the LM criterion has not been studied in this context.[5] Also, there are well documented problems with unit root tests when the chosen lag is too small. Specifically, as shown by Hall (1994) and Ng and Perron (1995), the ADF test suffers from low power when the lag length is too small. Also, Ng and Perron (2001) demonstrate that the DF-GLS test suffers from size distortions when the lag is too small. Since the former problems leads to too few rejections, and the latter problems leads to too many, lag selection alone may be responsible for the difference in rejections rate that Taylor finds between ADF and DF-GLS tests.

There exist lag selection procedures whose properties are well understood, and which offer a better combination of size and power. Specifically, the existing literature suggests using the general-to-specific (GS) procedure of Hall (1994) for ADF tests and the Modified Akaike information criterion (MAIC) of Ng and Perron (2001) for DF-GLS tests. Accordingly, we report ADF and DF-GLS tests in Table 1 and Table 2 respectively. Every series ends in 1998 but not all have the same starting date. The sample period for each series is indicated in Tables 1 and 2. For completeness, we consider LM, GS, and MAIC lag selection for each test, although we will focus on GS for ADF test, and MAIC for DF-GLS test.

With GS lag selection, we find 8 of 16 rejections at the 5% level using ADF tests; 2 more than Taylor finds for the same 16 countries with data ending in 1996. With MAIC lag selection, we find 9 of 16 rejections at the 5% level with DF-GLS test, 1 less than Taylor with the same data ending in 1996. Thus, using proper lag selection procedures, the difference in rejection rates between the ADF and DF-GLS test has almost been eliminated.[6]

Taylor (2002) also reports rejections of the unit root null based on ADF and DF-GLS tests with the inclusion of a deterministic time trend. While rejection of the unit root null in this case is not evidence of PPP in the strict sense of a mean reverting real exchange rate, one can ascribe a Balassa-Samuelson interpretation to a stationary real exchange rate

[5] See Hall (1994), Ng and Perron (1995), and Ng and Perron (2001).

[6] The 8 rejections in Table 1 are not a proper subset of the 9 rejections in Table 3. In total, there are 11 countries that reject the unit root null in Tables 1 and/or 3. We do not report 11 rejections since this testing strategy would our test to be oversized.

around a time trend.[7] A sensible strategy when testing for unit roots in persistent data, when some of the series are clearly trending, is to employ a two-step procedure. First, test the unit root null against the alternative of level stationarity. If the null is rejected, conclude that the series is level stationary, or in our case, that long run PPP holds. If the null cannot be rejected against level stationarity, test the unit root null against the alternative of trend stationarity. If the null is rejected, conclude that the series is trend stationary. If the null cannot be rejected, conclude that the series is I(1), and that long run PPP fails to hold. This testing strategy, which Taylor essentially follows, has intuitive appeal when testing for unit roots in real exchange rate data, since some series are clearly trending. Suppose that the true data generating process is level stationarity. In this case the unit root test described by equation (1) or (2) is appropriate. A time trend is an extraneous regressor, and its inclusion reduces power. However, if the true data generating process were trend stationarity, failing to include a time trend also results in a reduction in power of the test. In addition, this loss or power from excluding a time trend when it should be present is more severe than the reduction in power associated with including a time trend when it is extraneous; see West (1987). Finally, if the series is I(1), then both steps of this procedure should lead to a failure to reject the unit root null.

Using this two-step procedure, Taylor finds 3 additional 5% rejections for the ADF test, and 1 additional 5% rejection for the DF-GLS test. He cites these extra rejections from including a time trend as an important ingredient to his conclusion that PPP holds for over a century. However, it appears that it is the use of sub-optimal lag selection which leads Taylor to find so many rejections in the presence of linear time trend, not the (trend) stationarity of the data. When one uses proper lag selection methods, there is only one additional unit root rejection for the ADF test and none for the DF-GLS test.

We report the results of both tests, again using all 3 methods of lag section for completeness. The ADF unit root test with a time trend runs the following regression:

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

the results of which are reported in Table 3. For the ADF test with GS lag selection the 10% rejection for the U.S./Australia exchange rate is strengthened to a 5% rejection.

[7] See Papell and Prodan (2003) for further discussion.

Table 4 reports the results of the DF-GLS test, also allowing for a time trend under the alternative hypothesis, which runs the following auxiliary regression:

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

where q_t^τ is the GLS detrended real exchange rate. For the DF-GLS test with MAIC lag selection, there are no additional unit root rejections.

Thus, using state of the art unit root tests and lag selection, 9 of 16 countries are consistent with PPP, while the remaining 7 are consistent with the failure of PPP to hold in the long run. We maintain that this evidence is *not* sufficient to conclude that long run PPP has held over the last century for *all* of the countries in our sample.

3. Median-Unbiased Estimation in DF-GLS Regressions.

Murray and Papell (2002) use the median-unbiased techniques of Andrews (1993) and Andrews and Chen (1994) to compute point estimates and confidence intervals for PPP half-lives. Since these estimates are based on ADF regressions, they do not optimally exploit the sample information in terms of power. We propose an extension of the Andrews (1993) and Andrews and Chen (1994) methodology to the DF-GLS test. The objective here is to obtain tighter confidence intervals than those of Murray and Papell (2002) to potentially shed more light on the PPP puzzle.

The extension of median-unbiased estimation to DF-GLS regressions is straightforward. Instead of working with the data in levels as in the ADF regression (1), we simply work with the GLS demeaned (or detrended) data in the auxiliary DF-GLS regression (2). Since deterministic terms have been removed by GLS demeaning (or detrending), none are present in the above regression.[8] When k = 0, as in Andrews (1993), the median-unbiased estimator is exact, and when $k > 0$, as in Andrews and Chen (1994), the median-unbiased estimator is approximate.[9]

3.1 Exactly Median-Unbiased Estimation

We compute our exactly median-unbiased estimator for equation (2) for the sample sizes considered by Andrews (1993). We also report 90% confidence intervals. Our

[8] Again, since we are interested in the strict interpretation of PPP, we do not allow for deterministic time trends, although doing so is straightforward.

[9] See Andrews and Chen (1994) and Murray and Papell (2002) for further details concerning the computation of approximately median-unbiased estimators.

estimator is reported in the first row of Table 5, and Andrews' estimator, based on equation (1), is reported in the second row of Table 5.[10] To find the median-unbiased estimator, we find the value of α such that the median of the least squares estimator is equal to the least squares estimate. For example, if the least squares estimate of α is 0.915, and $T+1=125$, the median-unbiased estimate of α based on the DF-GLS regression is 0.930. A similar exercise leads to the construction of confidence intervals.

Two features of Table 5 are important to highlight here. First, while median-bias is present in the least squares estimator of α in DF-GLS regressions, it is not as severe as the bias in ADF regressions. This accords with intuition since bias worsens as the number of deterministic regressors increases. The auxiliary DF-GLS regression (2) contains no deterministic terms, while the ADF regression (1) contains a constant. Second, the confidence intervals from the DF-GLS regressions are tighter than from the ADF regressions. Uniformly, the lower bounds of the confidence intervals for the median-unbiased estimator of α are higher in the DF-GLS regressions than in the ADF regressions. Similarly, with only a few exceptions when $T+1=40$, the upper bounds from the DF-GLS regressions are higher than from the ADF regressions. Even though both the upper are lower bounds are higher, the confidence intervals are uniformly tighter in the DF-GLS case. This derives from the greater power of the DF-GLS test, and demonstrates the potential to extract more information on the persistence of shocks to real exchange rates than has been previously available.

3.2 Approximately Median-Unbiased Estimation

When $k > 0$, median-unbiased estimation is no longer exact, but approximate. In addition, the half life, which is based on the impulse response function, is a nonlinear transformation of an approximately median unbiased estimate, and is therefore biased. In this subsection, we conduct a simulation study of the half-life estimate to determine how our proposed half-life estimator performs relative to that of Andrews and Chen (1994), in terms of bias and precision.

We consider four values of α: 0.85, 0.90, 0.95, and 1. For each value of α, we

[10] While our subsequent empirical application reports 95% confidence intervals, we report 90% confidence intervals in Table 5 in order to directly compare our estimator to Andrews' estimator, for which he does not report 95% confidence intervals.

generate multiple parameterizations, either 2nd or 3rd order autoregressions. The true half-lives of all the parameterizations we consider range from 3.3 years to infinity. For each process, we compute the approximately median-unbiased estimate of the half-life, as well as the 95% confidence interval, using our proposed methodology, as well as than of Andrews and Chen (1994). The results are reported in Table 6.

There are two main features of Table 6 worth noting. First, although not severe, our estimate of the half-life is downward biased for every data generating process we consider. In addition, the bias we find in our estimator is greater than the bias of the Andrews and Chen (1994) estimator. While our estimator is arguably outperformed by the Andrews and Chen estimator in terms of point estimates of the half-life, it paints a more precise picture of the persistence of shocks to the real exchange rate. Our 95% confidence intervals are tighter in every case. Our lower bounds of the half-life are always higher, and except for the case where the true half-life is infinity, our upper bounds are always lower. The confidence interval of the half-life is arguably more important than the point estimate when one is trying to compare the persistence of shocks to the exchange rate with the predictions from economic models.[11] Our proposed methodology leads to notably tighter confidence intervals than those computed from the Andrews and Chen (1994) methodology, and demonstrates the ability to gain more information regarding the PPP puzzle when the median-unbiased estimator is only approximate.

4. Empirical Results: The Persistence of Shocks to the Real Exchange Rate

Turning now to the data, we compute median-unbiased estimates of half-lives, and 95% confidence intervals, for our 16 dollar denominated real exchange rates. The half-life is defined as the number of years required for a unit shock to dissipate by one-half, and is based directly on the impulse response function for each real exchange rate. In Table 7, we report half-life estimates from DF-GLS regressions where the lag length has been chosen by the MAIC. The median-unbiased estimates are exact when $k = 0$, and are approximate when $k > 0$.

[11] Using a different methodology, Rossi (2002) only reports confidence intervals for half-lives.

The point estimates of the half-lives in Table 7 are larger than what has been previously reported in the literature. The median point estimate is 11.34 years, with 12 of the 16 half-lives lying outside Rogoff's (1996) 3-5 year interval. This strengthens Murray and Papell's (2002) conclusion that the literature surveyed by Rogoff (1996) does not accurately represent the behavior of real exchange rates. Furthermore, the 95% confidence intervals paint a much different picture of the persistence of deviations from PPP, vis-à-vis models with nominal rigidities. The median confidence interval for half-lives of PPP deviations is $[3.02, \infty)$ years and, with the exception of the US/Finland real exchange rate, every lower bound is greater than 2 years.[12]

We would like to know whether the larger point estimates and lower bounds of the confidence intervals that we report (compared with previous work) are solely caused by differences in techniques, or if differences in the data also play a role. The half-lives for long-horizon data in Murray and Papell (2002) are computed from ADF regressions using GS lag selection, using data for six countries from 1900-1996. To remove one of these differences, we have also computed median-unbiased half-lives from Taylor's (2002) long term data set, based on ADF regressions with GS lag selection. These are reported in Table 8

The point estimates of the half-lives from ADF regressions in Table 8 are also larger than what has been previously reported in the literature. The median point estimate is 7.55 years, with 11 of the 16 half-lives lying outside Rogoff's (1996) 3-5 year interval. As with the DF-GLS regressions in Table 7, the 95% confidence intervals paint a much different picture of the persistence of deviations from PPP than models with nominal rigidities. The median confidence interval for half-lives of PPP deviations is [3.25, 30.21] years and, again with the exception of the US/Finland real exchange rate, every lower bound is greater than 2 years.

The differences in the point estimates and confidence intervals between Table 8 and Murray and Papell (2002) are not caused by the inclusion of additional countries. For the six countries: Canada, France, Italy, Japan, the Netherlands, and the United Kingdom,

[12] Since we have an even (16) number of real exchange rates, we report the median as the average of the eighth and ninth largest values. For the upper bound of the 95% confidence intervals, this is the average of 72.86 and infinity, which we report as infinity.

that are included in both studies, the median point estimate is 7.36 years and the median confidence interval is [3.74, 37.68] years, compared to 3.98 years and [1.69, 12.75] years in Murray and Papell (2002). The two most important findings from Tables 7 and 8 are that the median point estimate is well above Rogoff's 3-5 year consensus and the median lower bound of the 95% confidence interval is well above what would be consistent with models based on nominal rigidities. These hold equally well for the six-country subset.

Another potential explanation for the differences with previous work is the time span of the data. Taylor's (2002) begins in 1870 or 1880 for most countries, and ends in 1996. We extend the end to 1998. In Murray and Papell (2002), we use data from Lee (1976) that begins in 1900 and ends (with our augmentation) in 1996. In order to see whether the larger half-lives are caused by the longer span of data, we computed (but do not report) median-unbiased point estimates and confidence intervals, using ADF regressions, for the six-country subset using Taylor's data from 1900 - 1996. The results are quite similar to the results for the same six-country subset with the full time span of data.

Having ruled out differences in technique and differences in time span as explanations for the differences between these and previous results, we are left with differences in data. Most of the existent work on persistence of PPP deviations using long-horizon data, including Abuaf and Jorion (1990), Glen (1992), Lothian and Taylor (1996), and Murray and Papell (2002), uses real exchange rates computed from nominal exchange rates and wholesale price indexes. Taylor (2002) computes real exchange rates using consumer price indexes. Since the consumer price index contains a larger component of non-traded goods than the wholesale price index, it is not surprising that the persistence of PPP deviations is larger. Since most studies of PPP using post-1973 floating exchange rates use consumer price indexes, this is another dimension where Taylor's data provides greater comparability between long-horizon and post-1973 data than was previously available.

Given that half-life confidence intervals based on the DF-GLS regression are shown in Table 6 to be uniformly narrower than those based on the ADF regression, it may seem puzzling that the confidence intervals in Table 7 are not all tighter than those in Table 8. However, this is caused entirely by MAIC and GS choosing different lag lengths. In the five cases where MAIC and GS choose the same lag length, the half-life confidence intervals based on our estimator are tighter than those based on Andrews and Chen's (1994) estimator.

We have focused on the differences in the confidence intervals between the median-unbiased ADF and DF-GLS regressions. Both sets of confidence intervals, however, are narrower than what currently exists in the literature, and the message from Tables 7 and 8 is the same. Using the largest available dataset, we are unable to reconcile the predictions of exchange rate models with nominal rigidities with the behavior of real exchange rates. Therefore, while tighter confidence intervals translate to more information about the persistence of deviations from PPP, this increase in information moves us away from solving the PPP puzzle.

5. Conclusion

Rogoff's (1996) "remarkable consensus" of 3-5 year half-lives of PPP deviations was based on studies using biased estimates that underestimate the magnitude of the PPP puzzle. Subsequent work using data for industrialized countries from the post-1973 flexible exchange rate period has obtained ambivalent conclusions. In Murray and Papell (2002) and Rossi (2002), the confidence intervals for half-lives are so wide that they are consistent with virtually anything. They range from a speed of reversion to PPP that is predicted by models with nominal rigidities (half-lives between 1 and 2 years) to the failure of PPP to hold in the long run (infinite half-lives). Murray and Papell (2002) also examine long-horizon data for six countries, and report similar results.

In this paper, we investigate the purchasing power parity puzzle for long-horizon data using "state of the art" techniques. We extend the median-unbiased estimation methodology developed by Andrews (1993) and Andrews and Chen (1994) to the efficient DF-GLS test of Elliott, Rothenberg, and Stock (1996), and report both point estimates and confidence intervals. We show that the more powerful DF-GLS test can produce tighter confidence intervals than the more widely used ADF test.

Taylor (2002) constructs long-horizon US dollar denominated real exchange rate data for 16 developed and 3 developing countries from nominal exchange rates and consumer price deflators. This data allows researchers, for the first time, to study long-horizon real exchange rates for developed countries with comparable coverage and construction to the data that are commonly used to investigate post-1973 real exchange rates. He concludes that, using more powerful DF-GLS tests rather than ADF tests, long-run PPP has held over the twentieth century. We question this conclusion for two reasons. First, using arguably superior lag selection methods, the number of rejections rises with the ADF tests and falls with the DF-GLS tests. The difference between the two tests is almost eliminated, and if one allows deterministic time trends then the two tests produce the same result. The unit root null can be rejected at the 5% level for 9 of 16 real exchange rates. Second, solely reporting unit root rejections paints a very incomplete picture of the PPP puzzle.

Rogoff (1996) argues that the combination of high short-run real exchange rate volatility and "glacial" speeds of mean reversion produce the PPP puzzle. Using the best available data and estimation techniques, we find half-lives of PPP deviations to be much larger than his 3-5 year consensus. Another contribution of our work is to augment the information conveyed by point estimates with confidence intervals. In our earlier work using long-horizon data, as well as in work using post-1973 data, median-unbiased confidence intervals for PPP deviations were too wide to be informative. In this paper we see something much different. Similar to previous work, the upper bounds of the confidence intervals are so high that we cannot rule out the failure of PPP to hold in the long run. In contrast to previous work, however, the lower bounds are also so high that we can rule out consistency with models based on nominal rigidities. While our quantitative results are very different from those reported by Rogoff, our conclusions are in some respects very similar. Using more powerful and more complete techniques with better data moves us further away from solving the PPP puzzle.

References

Abuaf, N. and P. Jorion, 1990, "Purchasing power parity in the long run," *Journal of Finance* 45, 157-174.

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, 1994, "Mean Reversion in Real Exchange Rates," *Economics Letters* 46, 251-56.

Cheung, Y.-W. and K.S. Lai, 2000, "On the Purchasing Power Parity Puzzle," *Journal of International Economics* 52: 321-330.

Diebold, F., Husted, S., and Rush, M. 1991, "Real Exchange Rates Under the Gold Standard," *Journal of Political Economy* 99, 1252-1271.

Elliott G, and Stock, J.H., "Confidence Intervals for Autoregressive Coefficients Near One," *Journal of Econometrics* 103: 155-181.

Elliott, G. and Pesavento, E., 2001, "Higher Power Tests for the Bilateral Failure of PPP After 1973," working paper, Emory University.

Elliott, G., Rothenberg, T., and Stock, J.H., 1996, "Efficient Tests for an Autoregressive Unit Root," *Econometrica* 64:813-836.

Engel, C.M., 2000, "Long-Run PPP May Not Hold After All," *Journal of International Economics* 51: 243-273.

Frankel, J., 1986, "International Capital Mobility and Crowding Out in the U.S. Economy: Imperfect Integration of Financial Markets or of Goods Markets?" in R. Hafer, ed., *How Open is the U.S. Economy?*, Lexington Books.

Glen, J.H., 1992, "Real Exchange Rates in the Short, Medium, and Long Run," *Journal of International Economics* 33, 147-66.

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.

Hansen, B., 1999, "The Grid Bootstrap and the Autoregressive Model," *The Review of Economics and Statistics* 81, 594-607.

Inoue, A, and L. Kilian, 2002, “Bootstrapping Autoregressive Processes with Possible Unit Roots,” *Econometrica* 70: 377-391.

Kilian, L., 1998, “Small-Sample Confidence Intervals for Impulse Response Functions,” *The Review of Economics and Statistics* 80, 218-230.

Kilian, L., 1999, “Finite-Sample Properties of Percentile and Percentile-*t* Bootstrap Confidence Intervals for Impulse Responses,” *The Review of Economics and Statistics* 81, 652-660.

Kilian, L., and T. Zha, 2002, “Quantifying the Uncertainty about the Half-Life of Deviations from PPP,” *Journal of Applied Econometrics* 17:107-125.

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.

Murray, C.J., and D.H. Papell, 2002, “The Purchasing Power Parity Persistence Paradigm,” *Journal of International Economics* 56: 1-19.

Murray, C.J., and D.H. Papell, 2003, “Do Panels Help Solve the PPP Puzzle?” working paper, University of Houston

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.

Papell, D.H., and R. Prodan, 2003, “Long Run Purchasing Power Parity: Cassel or Balassa-Samuelson?” working paper, University of Houston.

Rogoff, K., 1996, “The Purchasing Power Parity Puzzle,” *Journal of Economic Literature* 34: 647-668.

Rossi, B., 2002, “Confidence Intervals for Half-Life Deviations from Purchasing Power Parity,” working paper, Duke University.

Taylor, A.M., 2002, “A Century of Purchasing Power Parity,” *The Review of Economics and Statistics* 84: 139-150.

West, K.D., 1987, “A Note on the Power of Least Squares Tests for a Unit Root,” *Economics Letters* 24: 249-252.

Table 1. ADF Unit Root Tests Without Time Trends

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

Country	Sample	ADF_{LM}	k_{LM}	ADF_{GS}	k_{GS}	ADF_{MAIC}	k_{MAIC}
Australia	1870-1998	-2.26	0	-2.62*	1	-2.26	0
Belgium	1880-1998	-3.32**	0	-4.15***	1	-2.92**	3
Canada	1870-1998	-1.62	0	-1.62	0	-1.62	0
Denmark	1880-1998	-2.27	0	-1.24	6	-1.24	6
Finland	1881-1998	-4.58***	0	-6.02***	1	-4.58***	0
France	1880-1998	-3.55***	1	-3.55***	1	-2.53	6
Germany	1880-1998	-2.95**	1	-2.95**	1	-2.39	2
Italy	1880-1998	-3.33**	0	-4.28***	2	-3.33**	0
Japan	1885-1998	-0.37	0	-1.02	1	-0.74	2
Netherlands	1870-1998	-2.14	0	-2.79*	1	-2.46	2
Norway	1870-1998	-2.58*	0	-3.67***	1	-2.23	5
Portugal	1890-1998	-2.69*	0	-2.25	5	-1.99	4
Spain	1880-1998	-2.43	0	-3.24**	1	-2.43	3
Sweden	1880-1998	-2.95**	0	-3.72***	1	-3.09**	2
Switzerland	1892-1998	-2.18	1	-1.50	2	-1.50	2
UK	1870-1998	-3.20**	0	-2.61*	4	-3.20**	0

*, **, and *** denote significance at the 10%, 5%, and 1% level respectively.

Table 2. ADF Unit Root Tests With Time Trends

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

Country	Sample	ADF_{LM}	k_{LM}	ADF_{GS}	k_{GS}	ADF_{MAIC}	k_{MAIC}
Australia	1870-1998	-3.22*	0	-3.66**	1	-3.22*	0
Belgium	1880-1998	-3.90**	0	-5.11***	1	-3.90**	0
Canada	1870-1998	-2.98	0	-2.98	0	-2.98	0
Denmark	1880-1998	-2.90	0	-1.97	6	-1.97	6
Finland	1881-1998	-4.69***	0	-6.22***	1	-4.69***	0
France	1880-1998	-4.16***	1	-4.16***	1	-1.67	8
Germany	1880-1998	-3.32*	1	-3.32*	1	-2.75	2
Italy	1880-1998	-3.33*	0	-4.27***	2	-3.33*	0
Japan	1885-1998	-2.07	0	-1.98	7	-2.32	2
Netherlands	1870-1998	-2.40	0	-3.19*	1	-2.40	0
Norway	1870-1998	-2.71	0	-3.95**	1	-2.64	5
Portugal	1890-1998	-2.60	0	-2.15	5	-1.72	4
Spain	1880-1998	-2.37	0	-3.23*	1	-2.34	3
Sweden	1880-1998	-3.40*	0	-4.52***	1	-3.40*	0
Switzerland	1892-1998	-3.63**	1	-2.78	2	-2.78	2
UK	1870-1998	-3.38*	0	-2.74	4	-3.38*	0

*, **, and *** denote significance at the 10%, 5%, and 1% level respectively.

Table 3. DF-GLS Unit Root Tests without Time Trends

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

Country	Sample	$DF\text{-}GLS_{LM}$	k_{LM}	$DF\text{-}GLS_{GS}$	k_{GS}	$DF\text{-}GLS_{MAIC}$	k_{MAIC}
Australia	1870-1998	-2.29**	0	-2.65***	1	-2.29**	0
Belgium	1880-1998	-2.84***	0	-3.57***	1	-2.38**	3
Canada	1870-1998	-1.40	1	-1.29	0	-1.29	0
Denmark	1880-1998	-2.24**	0	-1.20	6	-1.20	6
Finland	1881-1998	-4.49***	0	-5.85***	1	-4.49***	0
France	1880-1998	-2.24**	1	-1.34	4	-1.07	6
Germany	1880-1998	-2.52**	1	-2.52**	1	-1.97**	2
Italy	1880-1998	-3.35***	0	-4.29***	2	-3.35***	0
Japan	1885-1998	0.25	0	-0.08	1	-0.08	1
Netherlands	1870-1998	-1.84*	0	-2.51**	1	-2.20**	2
Norway	1870-1998	-1.61*	0	-2.49**	1	-1.31	5
Portugal	1890-1998	-1.88*	0	-1.52	5	-1.29	6
Spain	1880-1998	-2.13**	0	-2.08**	3	-2.08**	3
Sweden	1880-1998	-2.35**	0	-2.36**	2	-2.36**	2
Switzerland	1892-1998	-1.47	1	-0.76	2	-0.76	2
UK	1870-1998	-2.81***	0	-2.26**	4	-2.26**	4

*, **, and *** denote significance at the 10%, 5%, and 1% level respectively.

Table 4. DF-GLS Unit Root Tests with Time Trends

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

Country	Sample	$DF\text{-}GLS_{LM}$	k_{LM}	$DF\text{-}GLS_{GS}$	k_{GS}	$DF\text{-}GLS_{MAIC}$	k_{MAIC}
Australia	1870-1998	-2.77*	0	-3.17**	1	-2.77*	0
Belgium	1880-1998	-3.93***	0	-5.14***	1	-3.93***	0
Canada	1870-1998	-1.48	0	-1.48	0	-1.48	0
Denmark	1880-1998	-2.79*	0	-1.86	6	-1.86	6
Finland	1881-1998	-4.72***	0	-6.27***	1	-4.72***	0
France	1880-1998	-4.04***	1	-4.04***	1	-1.60	8
Germany	1880-1998	-3.35**	1	-3.35**	1	-2.78*	2
Italy	1880-1998	-3.35**	0	-4.30***	2	-3.35**	0
Japan	1885-1998	-1.92	0	-1.98	7	-2.28	2
Netherlands	1870-1998	-2.43	0	-3.23**	1	-2.43	0
Norway	1870-1998	-2.58	0	-3.81***	1	-2.60	5
Portugal	1890-1998	-2.49	0	-2.15	5	-1.71	4
Spain	1880-1998	-2.37	0	-3.22**	1	-2.36	3
Sweden	1880-1998	-3.44**	0	-4.55***	1	-3.44**	0
Switzerland	1892-1998	-3.64***	1	-2.78*	2	-2.78*	2
UK	1870-1998	-2.99*	0	-2.45	4	-2.99*	0

*, **, and *** denote significance at the 10%, 5%, and 1% level respectively

Table 5. Exactly Median-Unbiased Estimators

DF-GLS Exactly Median-Unbiased Estimator

T+1=40

α/Quantile	0.05	0.50	0.95
1	0.737	0.930	1.000
0.99	0.727	0.918	1.000
0.97	0.696	0.895	0.982
0.93	0.646	0.856	0.956
0.9	0.612	0.827	0.938
0.85	0.553	0.782	0.908
0.8	0.501	0.737	0.877

T+1=50

α/Quantile	0.05	0.50	0.95
1	0.797	0.950	1.000
0.99	0.781	0.938	1.000
0.97	0.755	0.915	0.984
0.93	0.702	0.876	0.959
0.9	0.665	0.847	0.941
0.85	0.607	0.799	0.909
0.8	0.551	0.752	0.876

T+1=60

α/Quantile	0.05	0.50	0.95
1	0.834	0.962	1.000
0.99	0.820	0.949	1.000
0.97	0.793	0.927	0.984
0.93	0.738	0.887	0.959
0.9	0.702	0.858	0.941
0.85	0.643	0.812	0.909
0.8	0.583	0.762	0.874

T+1=70

α/Quantile	0.05	0.50	0.95
1	0.862	0.970	1.000
0.99	0.845	0.957	1.000
0.97	0.817	0.936	0.985
0.93	0.766	0.896	0.960
0.9	0.727	0.866	0.941
0.85	0.665	0.818	0.908
0.8	0.610	0.771	0.872

Andrews (1993) OLS Exactly Median-Unbiased Estimator

T+1=40

α/Quantile	0.05	0.50	0.95
1	0.674	0.893	0.999
0.99	0.666	0.886	0.994
0.97	0.649	0.872	0.983
0.93	0.612	0.841	0.958
0.9	0.582	0.816	0.939
0.85	0.532	0.772	0.908
0.8	0.480	0.727	0.875

T+1=50

α/Quantile	0.05	0.50	0.95
1	0.735	0.914	0.999
0.99	0.727	0.907	0.994
0.97	0.706	0.893	0.982
0.93	0.669	0.860	0.957
0.9	0.638	0.834	0.938
0.85	0.585	0.789	0.906
0.8	0.532	0.743	0.872

T+1=60

α/Quantile	0.05	0.50	0.95
1	0.777	0.928	0.999
0.99	0.769	0.921	0.994
0.97	0.750	0.906	0.982
0.93	0.709	0.973	0.957
0.9	0.676	0.846	0.938
0.85	0.622	0.800	0.905
0.8	0.567	0.753	0.870

T+1=70

α/Quantile	0.05	0.50	0.95
1	0.807	0.938	0.999
0.99	0.799	0.931	0.994
0.97	0.780	0.916	0.982
0.93	0.737	0.882	0.957
0.9	0.704	0.854	0.937
0.85	0.648	0.807	0.903
0.8	0.593	0.760	0.867

Table 5. Exactly Median-Unbiased Estimators, Continued

DF-GLS Exactly Median-unbiased Estimator

T+1=80

α/Quantile	0.05	0.50	0.95
1	0.879	0.975	1.000
0.99	0.864	0.963	1.000
0.97	0.837	0.942	0.985
0.93	0.785	0.902	0.960
0.9	0.747	0.872	0.941
0.85	0.684	0.824	0.907
0.8	0.627	0.774	0.870

T+1=90

α/Quantile	0.05	0.50	0.95
1	0.894	0.979	1.000
0.990	0.879	0.967	1.000
0.970	0.852	0.946	0.985
0.930	0.799	0.906	0.960
0.900	0.761	0.877	0.940
0.850	0.699	0.827	0.905
0.800	0.641	0.779	0.868

T+1=100

α/Quantile	0.05	0.50	0.95
1	0.906	0.982	1.000
0.990	0.891	0.970	1.000
0.970	0.862	0.949	0.985
0.930	0.809	0.909	0.960
0.900	0.773	0.879	0.939
0.850	0.713	0.831	0.905
0.800	0.654	0.782	0.867

T+1=125

α/Quantile	0.05	0.50	0.95
1	0.927	0.987	1.000
0.990	0.912	0.976	1.000
0.970	0.884	0.955	0.985
0.930	0.832	0.915	0.960
0.900	0.793	0.885	0.938
0.850	0.735	0.835	0.901
0.800	0.678	0.787	0.863

Andrews (1993) OLS Exactly Median-Unbiased Estimator

T+1=80

α/Quantile	0.05	0.50	0.95
1	0.831	0.946	0.999
0.99	0.822	0.939	0.994
0.97	0.802	0.923	0.982
0.93	0.758	0.888	0.956
0.9	0.724	0.861	0.937
0.85	0.668	0.813	0.902
0.8	0.613	0.765	0.865

T+1=90

α/Quantile	0.05	0.50	0.95
1	0.849	0.952	0.999
0.99	0.840	0.945	0.994
0.97	0.820	0.929	0.981
0.93	0.775	0.893	0.956
0.9	0.741	0.865	0.936
0.85	0.684	0.818	0.901
0.8	0.628	0.769	0.863

T+1=100

α/Quantile	0.05	0.50	0.95
1	0.863	0.957	0.999
0.99	0.854	0.950	0.994
0.97	0.834	0.933	0.981
0.93	0.788	0.897	0.956
0.9	0.754	0.869	0.936
0.85	0.697	0.821	0.900
0.8	0.641	0.773	0.862

T+1=125

α/Quantile	0.05	0.50	0.95
1	0.890	0.965	0.999
0.99	0.881	0.958	0.994
0.97	0.859	0.941	0.981
0.93	0.813	0.904	0.955
0.9	0.778	0.876	0.934
0.85	0.721	0.827	0.897
0.8	0.664	0.778	0.858

Table 5. Exactly Median-Unbiased Estimators, Continued

DF-GLS Exactly Median-Unbiased Estimator

T+1=150

α/Quantile	0.05	0.50	0.95
1	0.941	0.990	1.000
0.99	0.924	0.979	0.999
0.97	0.896	0.958	0.985
0.93	0.845	0.918	0.959
0.9	0.808	0.888	0.937
0.85	0.749	0.839	0.899
0.8	0.692	0.789	0.860

T+1=200

α/Quantile	0.05	0.50	0.95
1	0.956	0.993	1.000
0.99	0.941	0.982	0.998
0.97	0.913	0.962	0.985
0.93	0.861	0.922	0.957
0.9	0.825	0.892	0.935
0.85	0.767	0.843	0.895
0.8	0.710	0.793	0.854

Andrews (1993) OLS Exactly Median-Unbiased Estimator

T+1=150

α/Quantile	0.05	0.50	0.95
1	0.908	0.971	0.999
0.99	0.898	0.964	0.994
0.97	0.876	0.947	0.981
0.93	0.829	0.909	0.955
0.9	0.794	0.880	0.933
0.85	0.737	0.831	0.895
0.8	0.681	0.782	0.855

T+1=200

α/Quantile	0.05	0.50	0.95
1	0.931	0.978	0.999
0.99	0.921	0.971	0.994
0.97	0.898	0.953	0.981
0.93	0.850	0.915	0.953
0.9	0.815	0.885	0.931
0.85	0.758	0.836	0.891
0.8	0.702	0.787	0.850

Table 6. Relative Performance of Approximately Median-Unbiased Half-Life Estimates Based on DF-GLS and ADF Regressions

	True Half-Life	Median Unbiased Estimates			
		DF-GLS	95% CI	ADF	95% CI
$\alpha = 1$					
$\phi_1 = 1.25$, $\phi_2 = -0.25$	∞	73.45	[12.10, ∞)	∞	[9.11, ∞)
$\phi_1 = 1.50$, $\phi_2 = -0.50$	∞	∞	[15.23, ∞)	∞	[11.64, ∞)
$\phi_1 = 0.80$, $\phi_2 = 0.20$	∞	62.89	[5.32, ∞)	∞	[3.97, ∞)
$\phi_1 = 0.60$, $\phi_2 = 0.40$	∞	52.91	[4.03, ∞)	∞	[2.74, ∞)
$\alpha = 0.95$					
$\phi_1 = 1.25$, $\phi_2 = -0.30$	14.63	12.21	[5.40, 36.64]	14.64	[5.28, ∞)
$\phi_1 = 1.50$, $\phi_2 = -0.55$	13.66	12.77	[6.90, 34.44]	13.49	[6.87, 68.34]
$\phi_1 = 0.80$, $\phi_2 = 0.15$	12.26	10.32	[3.22, 60.02]	12.05	[2.95, ∞)
$\phi_1 = 0.60$, $\phi_2 = 0.35$	9.99	8.22	[2.49, 51.97]	10.51	[2.24, ∞)
$\alpha = 0.90$					
$\phi_1 = 1.25$, $\phi_2 = -0.35$	7.41	6.79	[4.12, 14.50]	7.37	[4.01, 18.28]
$\phi_1 = 1.55$, $\phi_2 = -0.85$, $\phi_3 = 0.20$	6.62	6.29	[3.83, 11.82]	6.70	[3.74, 17.61]
$\phi_1 = 0.60$, $\phi_2 = 0.30$	4.98	4.49	[2.06, 11.77]	4.94	[2.01, 58.18]
$\alpha = 0.85$					
$\phi_1 = 1.25$, $\phi_2 = -0.40$	5.17	5.02	[3.34, 8.24]	6.11	[3.35, 9.14]
$\phi_1 = 1.55$, $\phi_2 = -0.85$, $\phi_3 = 0.15$	4.92	4.86	[3.61, 7.35]	4.93	[3.64, 8.49]
$\phi_1 = 0.60$, $\phi_2 = 0.25$	3.30	3.02	[0.86, 7.50]	3.14	[0.85, 10.96]

Table 7. Median-Unbiased Half-Lives in DF-GLS Regressions

Country	Sample	k_{MAIC}	α_{LS}	α_{MU}	95% CI	HL_{MU}	95% CI
Australia	1870-1998	0	0.913	0.92	[0.85, 1.0]	8.31	[4.27, ∞)
Belgium	1880-1998	3	0.872	0.88	[0.77, 0.98]	3.73	[2.37, 12.67]
Canada	1870-1998	0	0.967	0.98	[0.93, 1.0]	34.31	[9.55, ∞)
Denmark	1880-1998	6	0.937	0.95	[0.85, 1.0]	12.42	[3.01, ∞)
Finland	1881-1998	0	0.704	0.71	[0.58, 0.84]	2.02	[1.27, 3.98]
France	1880-1998	6	0.965	0.98	[0.89, 1.0]	12.89	[2.56, ∞)
Germany	1880-1998	2	0.943	0.95	[0.89, 1.0]	14.17	[5.14, 72.86]
Italy	1880-1998	0	0.825	0.83	[0.73, 0.94]	3.72	[2.20, 11.20]
Japan	1885-1998	1	0.999	1.0	[0.99, 1.0]	∞	[10.50, ∞)
Netherlands	1870-1998	2	0.927	0.93	[0.87, 1.0]	10.26	[4.14, 34.29]
Norway	1870-1998	5	0.961	0.97	[0.91, 1.0]	22.21	[4.22, ∞)
Portugal	1890-1998	6	0.954	0.97	[0.88, 1.0]	17.68	[2.87, ∞)
Spain	1880-1998	3	0.924	0.93	[0.85, 1.0]	9.36	[3.03, 35.07]
Sweden	1880-1998	2	0.905	0.91	[0.83, 0.99]	7.46	[2.76, 21.24]
Switzerland	1892-1998	2	0.981	0.99	[0.95, 1.0]	72.11	[7.37, ∞)
UK	1870-1998	4	0.886	0.89	[0.78, 0.99]	3.91	[2.86, 12.55]

Table 8. Median-Unbiased Half-Lives in ADF Regressions

Country	Sample	k_{GS}	α_{LS}	α_{MU}	95% CI	HL_{MU}	95% CI
Australia	1870-1998	1	0.897	0.92	[0.84, 1.0]	8.81	[3.86, 36.00]
Belgium	1880-1998	1	0.780	0.80	[0.70, 0.90]	3.78	[2.44, 6.67]
Canada	1870-1998	0	0.930	0.96	[0.88, 1.0]	16.98	[5.42, ∞)
Denmark	1880-1998	6	0.935	0.99	[0.88, 1.0]	24.96	[2.92, ∞)
Finland	1881-1998	1	0.584	0.60	[0.46, 0.74]	2.11	[1.57, 2.87]
France	1880-1998	1	0.844	0.87	[0.77, 0.96]	5.60	[2.96, 13.89]
Germany	1880-1998	1	0.910	0.93	[0.86, 1.0]	10.44	[5.00, 72.00]
Italy	1880-1998	2	0.753	0.77	[0.67, 0.86]	3.76	[2.42, 5.69]
Japan	1885-1998	1	0.984	1.0	[0.97, 1.0]	∞	[9.20, ∞)
Netherlands	1870-1998	1	0.904	0.92	[0.85, 1.0]	9.13	[4.52, 24.42]
Norway	1870-1998	1	0.871	0.89	[0.81, 0.97]	6.78	[3.88, 14.62]
Portugal	1890-1998	5	0.883	0.91	[0.81, 1.0]	8.32	[2.76, ∞)
Spain	1880-1998	1	0.875	0.89	[0.82, 0.98]	6.70	[3.54, 18.22]
Sweden	1880-1998	1	0.829	0.85	[0.75, 0.95]	4.95	[2.89, 10.42]
Switzerland	1892-1998	2	0.957	0.99	[0.92, 1.0]	71.64	[4.94, ∞)
UK	1870-1998	4	0.852	0.89	[0.76, 1.0]	4.02	[2.92, 50.94]