POTENTIAL PITFALLS FOR THE PURCHASING-POWER-PARITY
PUZZLE? SAMPLING AND SPECIFICATION BIASES IN
MEAN-REVERSION TESTS OF THE LAW OF ONE PRICE

Alan M. Taylor

Potential Pitfalls for the Purchasing-Power-Parity Puzzle?
Sampling and Specification Biases in Mean-Reversion Tests
of the Law of One Price
Alan M. Taylor
NBER Working Paper No. 7577
March 2000
JEL No. C22, C51, F10, F30, F40

ABSTRACT

The PPP puzzle is based on empirical evidence that international price differences for individual goods (LOOP) or baskets of goods (PPP) appear highly persistent or even non-stationary. The present consensus is these price differences have a half-life that is of the order of five years at best, and infinity at worst. This seems unreasonable in a world where transportation and transaction costs appear so low as to encourage arbitrage and the convergence of price gaps over much shorter horizons, typically days or weeks. However, current empirics rely on a particular choice of methodology, involving (i) relatively low-frequency monthly, quarterly, or annual data, and (ii) a linear model specification. In fact, these methodological choices are not innocent, and they can be shown to bias analysis to-wards findings of slow convergence and a random walk. Intuitively, if we suspect that the actual adjustment horizon is of the order of days then monthly and annual data cannot be expected to reveal it. If we suspect arbitrage costs are high enough to produce a substantial "band of inaction" then a linear model will fail to support convergence if the process spends considerable time random-walking in that band. Thus, when testing for PPP or LOOP, model specification and data sampling should not proceed without consideration of the actual institutional context and logistical framework of markets.

Alan M. Taylor
Department of Economics
University of California
One Shields Avenue
Davis, CA 95616-8578
and NBER
amtaylor@ucdavis.edu
http://www.econ.ucdavis.edu/faculty/amtaylor/

1   Introduction

As recent survey articles show, there has been a remarkable surge lately in research on purchasing power parity (PPP) and the law of one price (LOOP). Once dismissed as dull topics, these subjects have rightly taken a place back on the main stage of a field where issues of market integration and the extent of the market have been central ever since the very foundation of the discipline. The study of international and regional markets, the choice of open- versus closed- economy models, and many other important issues for theory, empirics, and policy rest on what economists can say about the existence of one market or many.[1]

Though practitioners of PPP- and LOOP-testing can agree on the importance of the questions they study, there is far less unanimity concerning the answers to those questions. We may speak of two teams on the playing field, or two contrary trends in the literature (for players may switch teams). Fortunately, the field is fairly level and the goalposts do not get moved around too much. Most players have converged on a methodology, the basic model of price convergence being a linear AR(1) form

$$x_t = \rho x_{t-1} + e_t, \tag{1}$$

where $x_t = p_t^2 - p_t^1$ is the price gap between two markets measured in a common currency, e_t is $N(0, \sigma^2)$, $0 < \rho < 1$, $\lambda = \rho - 1 < 0$ is the *convergence speed*, and $H = \ln \frac{1}{2} / \ln \rho$ is the *half-life* of deviations.

Differences between the two teams center on whether this model holds, or in what form, for the contemporary period. One team may be termed the "whittling down half-lives" team; for them, half-lives of deviations are small and reasonable, and price gaps are stationary; but even here, half-lives tend to be measured in years. On the other side is the "whittling up half-lives" team; for them, half-lives are much longer than seems reasonable, and could even be infinite; price gaps might follow a random walk. Considerable ink has been spilled on both sides of this debate, and still no broad agreement seems at hand. The sense of conundrum is only furthered by the realization that, for the most part, studies have been run on essentially the same models and data, deriving different conclusions from slight changes in samples, pooling and panel techniques, stationarity tests, and the like.[2]

---

[1] For surveys of the literature see Froot and Rogoff (1995); Rogoff (1996); and M. P. Taylor (1995).

[2] For example, to survey just a few articles from the downward whittlers, a recent wave of studies (for example, Wei and Parsley 1995; Chinn and Johnston 1996; Frankel and Rose 1996; Jorion and Sweeney 1996; Oh 1996; Wu 1996) finds considerable evidence against unit roots in real exchange rates, but still suggests long half-lives for deviations from PPP, with consensus estimates clustering in the range of four to five years. But the pendulum then seemed to be swinging back toward the upward whittlers, and new work cast doubt on the long-run stationarity of international relative prices, even using wide panels and even when attention was restricted to supposedly tradable goods (for example, O'Connell 1998; Engel 2000; Canzoneri, Cumby, and Diba 1999; Papell 1997).

The present paper—and let the cards go on the table right away—is unabashedly in the whittling-down camp. However, the purpose is not to settle the debate using the same model and the same data as the bulk of the literature. Rather it is to point up a pair of problems for PPP puzzlers involving model specification and sampling, which, taken separately, could imply very large biases in the estimation procedure, and taken together could explain away much of the puzzle.

The first problem is one of temporal aggregation. Typically, PPP and LOOP have been tested with aggregate data at annual, quarterly, and sometimes monthly frequencies. When the half-lives turn out to be low, surprise is expressed that we do not find markets adjusting in days and weeks, but rather in four to five years. I will show that this result is not a surprise. Sampling the data at low frequencies will never allow one to identify a high-frequency adjustment process. Instead, a large bias could be introduced towards the finding of a long half-life, and the bias grows larger the greater the degree of temporal aggregation. In both empirical simulations and in theory, it can be shown exactly how large these biases are. This observation makes no stunning contribution to the literature on temporal aggregation problems; the novelty here is to highlight what this means for the PPP puzzle, an implication that most of the literature has thus far ignored.

The second problem is linear specification. The standard model assumes that reversion to zero place at the same rate regardless of how far the process is from zero. But recent advances in theory and empirics cast doubt on this idea. It is well known that fixed and variable trading costs, or risk aversion, can under certain scenarios lead to a "band of inaction" in which no arbitrage occurs despite a nonzero price gap. Only when prices move apart sufficiently will arbitrage occur and reversion begin. Again, there is nothing new in the idea, which dates back to Eli Heckscher's (1916) original exposition, a forerunner of Gustav Cassel's (1922) idea of PPP. Heckscher called these boundaries the "commodity points"—by analogy with the gold points of the gold-standard adjustment mechanism. The originality lies in the estimation of the potential impact of this effect on the standard techniques used to infer convergence speeds.[3] Recent research has demonstrated the relevance of this kind of adjustment process in many market situations, requiring the use of nonlinear dynamic models, and the present paper extends these findings with a detailed simulation approach.[4]

The paper is structured as follows. In Section 2 we will examine temporal-aggregation biases. Delays between sampling introduce error, but time-averaging introduces serious biases, as demonstrated by simulation. A theoretical result places

---

[3]See also Yeager (1976).

[4]See, for example, Williams and Wright (1991); Dumas (1992); Coleman (1995); Sercu, Uppal, and Van Hulle (1995); Balke and Fomby (1997); O'Connell (1996); Prakash (1996); Prakash and A. M. Taylor (1997); Obstfeld and A. M. Taylor (1997); M. P. Taylor and Peel (1999).

bounds on the size of these biases. Section 3 examines biases arising from an improper use of linear models. If the true model is nonlinear, then applying the linear model results in the unwarranted pooling of stationary-reverting data outside the band, and nonstationary-nonreverting data within the band. Once again, simulation analysis demonstrates the potential size of these biases, which are nontrivial. Section 4 considers the consequences of having *both* biases present at the same time. The biases do add up and could make the estimation even more inaccurate. If such biases afflicted typical tests of PPP then half-lives would be very exaggerated and the PPP puzzle would not be a puzzle at all. Section 5 confirms an expected corollary that raises doubts about the tests of random walks used to evaluate PPP or LOOP. When either of our two kinds of biases are present, not only do they bias the convergence speed estimate, they can also render conventional unit root tests even more powerless than they already are, offering an explanation for why it might be so hard to reject a random walk hypothesis in so many cases. Section 7 considers how such biases could be a problem in practice. A favored contemporary data source is certainly afflicted with many sampling problems, but on the question of temporal aggregation the source is very clearly tainted. The conclusion notes that, given these lessons from theory and history, contemporary tests of PPP and LOOP should perhaps give more attention to sampling issues and nonlinear dynamics, since the potential problems of bias and power are quite serious.

2   Pitfall One: Temporal Aggregation

The problem of temporal aggregation has been appreciated in the econometric literature ever since the seminal contribution of Holbrook Working (1960). Coincidentally, he too was concerned with autoregressions of price differences and biases introduced by time averaging, especially in the estimate of serial correlation parameters. In this section I will develop and generalize Working's approach, but I will apply it not only through a few specific examples, as he did, but also via a broader study of all possible temporal aggregations. I will show how to develop a general formula for the bias in the standard AR(1) model of price adjustment.

It is curious that Working's approach has left virtually no imprint on the PPP and LOOP literature, especially given his original focus on price adjustment and the data problems endemic to this field. In the last decade, when literally hundreds of papers have appeared on PPP and LOOP, only a handful cite Working and the temporal aggregation problem, and just one of those appeared in the last five years. Glen (1992) noted the need for the correction when using Lee's (1978) annual average historical data for the twentieth century. Apte, Kane, and Sercu (1994) noted the Working correction in passing but temporal aggregation was not central

to their study. The most recent paper, by Maeso-Fernandez (1998), is an important contribution that, like the present paper, notes how temporal aggregation can affect unit root tests—in this case the variance ratio test of Cochrane (1988) as extended and applied by Lo and MacKinlay (1988) and Grill and Kaminsky (1991), though the implications for half-life bias are not discussed.

The missing connection is also strange given that many economists are aware of these issues, and cite Working and the temporal aggregation problem in other kinds of economic studies. An extensive literature notes the role of temporal aggregation as it affects models of consumption, the permanent-income hypothesis, credit constraints as they relate to volatility, the dynamics of real wages, and various other business-cycle phenomena.[5] The same aggregation issues have arisen in a number of other fields ranging from stock-market and exchange-rate volatility to the asset-pricing of trees.[6] Finally, since so much of the testing of PPP and LOOP has focused on stationarity testing, we should also note that the unit-root literature has also paid some attention to the Working effect, and has concluded that problems of temporal aggregation can certainly affect the way we make inferences from standard unit root tests by biasing coefficients and lower power (Schwert 1987; Ng 1995). Still, this finding has not yet made an impact on the mainstream of the PPP and LOOP literature, an oversight I hope to address in this paper. First, in this section, I will show how temporal aggregation promotes upwardly biased half-lives; later I will show how it saps the power from unit root tests.

In order to consider the temporal aggregation problem as it affects the testing of PPP and LOOP, we must first see when the problem might arise. It is obvious that certain kinds of temporal aggregation do not introduce bias into the standard model (1). For example, if we had a daily price process but only observed it at weekly intervals, then, assuming it was still an AR1 process, estimation in the standard way would not introduce bias. The errors would compound over time, but the sampled data would still follow an AR1 process.[7] However, another kind of

---

[5]See, for example, Hall (1988); Ermini (1988, 1993); Ramey (1989); Campbell and Mankiw (1990); Christiano, Eichenbaum, and Marshall (1991); Rossana and Seater 1992; Heaton (1993); Bayoumi (1993); Blundell, Wignall, Browne, and Tarditi (1995); Berloffa (1997); Bacchetta and Gerlach (1997); and Sarno and M. P. Taylor (1998).

[6]On stock markets see, for example, Poterba and Summers (1988); Lam (1990); and Snowden (1990); on exchange rates, Baillie and Bollerslev (1991); on stumpage, Washburn and Binkley (1990a; 1990b).

[7]Formally, the process with weekly sampling would be of the form

$$x_t = \rho' x_{t-7} + e'_t,$$

where $\rho' = \rho^7$ and $e'_t = e_t + \rho e_{t-1} + \rho^2 e_{t-2} + \cdots + \rho^6 e_{t-6}$. The only problem here would be the loss of data and information: for a given span of time, the estimate using low frequency data would be much less precise than that using high frequency data.

temporal aggregation, time-averaging, is not so innocent, and can produce severe biases, as we now show.

Suppose that prices are contracted daily in the relevant markets, and that arbitrage happens each day after prices are set, but that we only observe data that is a weekly average of the market prices. In that case, the true process for the price gap is the daily x_t, where the index $t = 0, \ldots, T - 1$ runs over each successive day. Instead, the limitations of the data force us to use a moving average of the data, in the form $x_s^* = \frac{1}{P}(x_{Ps} + \ldots + x_{Ps+P-1})$, where $s = 0, \ldots, S - 1$. Here, P denotes the period over which averaging takes places. In this example with a daily process and weekly sampling $P = 7$. In another example, with a weekly process and annual sampling we would have $P = 52$.

The dynamics of the sampled data x_s^* will not correspond to those of the process x_t, even adjusting for a rescaling in the time dimension. In particular, we can show that estimates of the convergence speed of the process x_s^* will always be slower than the process x_t, and that this bias increases as the degree of temporal aggregation P increases.

This is not to say, of course, that a correct ARMA representation could not, in principle, be constructed to match the true form of sample averaging in the data. With such a model in hand, one could indeed, with appropriate coefficient restrictions, recover the true parameters of the unaggregated AR(1) process from the aggregated data. However, the point here is merely to note that even a thorough search of the PPP and LOOP literature would lead one to conclude that the universal empirical practice is *not* to implement a wide search over such ARMA specifications. Rather, the quintessential methodology has been the simple, yet far from ideal, AR(1) type of model; so the purpose of this exercise is to see what biases such a narrow specification might impose under certain plausible aggregations.

In practice, suppose one were to estimate the basic model on temporally-aggregated data, employing a regression

$$x_s^* = \rho^* x_{s-1}^* + e_s^*. \tag{2}$$

Since the process of adjustment has been reduced to one parameter, the convergence speed (or the half-life, equivalently), we can focus on the behavior of the estimate of this parameter as a function of the temporal aggregation P. In the model (2), the autoregressive coefficient has an expected value of

$$\rho^*(P; \rho) = \mathrm{E}(\rho^*) = \frac{\mathrm{Cov}(x_s^*, x_{s-1}^*)}{\mathrm{Var}(x_{s-1}^*)}, \tag{3}$$

and this would yield a half-life estimate of

$$H^* = H^*(P; \rho) = \frac{P \ln \frac{1}{2}}{\ln \rho^*} = \frac{\ln \frac{1}{2}}{\ln \rho^{*\frac{1}{P}}}. \tag{4}$$

5

Note that the factor P enters because, with the temporal aggregation, time is being measured in P-unit intervals in the model (2). Thus, for comparison with normal time, the derived half-life has to be scaled by P.

A series of results now establishes that this half-life estimate $H^*(P)$ is an upwardly biased estimate of the true half-life H for $P > 1$, that the *bias factor* $H^*(P)/H > 1$ increases as the degree of temporal aggregation P increases, and that in the limit as P grows very large we obtain infinite half-life estimates $H^*(P)$ whatever the true (finite) half-life H. However, one can also show that for given P, the size of the bias decreases as the true half-life H increases. Moreover, at long horizons, so long as the sampling frequency accords with the true frequency of the data, so that P is of order H, the size of the bias $H^*(P)/H$ is bounded by a finite limit.

In order to proceed we need a formula for ρ^*. From (3) we see that this amounts to a tedious calculation involving the covariance of moving averages of an autoregressive process. Details can be found in the appendix but the key result is as follows.

Proposition 1 *In the case of P-period temporal aggregation, the autoregressive coefficient has an expected value*

$$\rho^* = \frac{\rho(1 - \rho^P)^2}{P(1 - \rho^2) - 2\rho(1 - \rho^P)}. \tag{5}$$

Proof See appendix. □

Using the above result, we can apply (4) to obtain an expression for the predicted half-life H^* as a function of the true ρ and the temporal aggregation P,

$$H^*(P; \rho) = \frac{P \ln(2)}{\ln\left(\frac{P(1-\rho^2)-2\rho(1-\rho^P)}{\rho(1-\rho^P)^2}\right)}. \tag{6}$$

We can find H^* as a function of P and H (the true half-life), once we make the substitution $\rho = 2^{-1/H}$, whereby

$$H^*(P; H) = \frac{P \ln(2)}{\ln\left(\frac{P(1-2^{-2/H})-2^{-1/H}(1-2^{-P/H})}{2^{-1/H}(1-2^{-P/H})^2}\right)}. \tag{7}$$

It is instructive to examine the implications of temporal aggregation using some numerical examples by way of illustration. Table 1 shows the values of the estimated

6

Table 1: Temporal-Aggregation Bias for Half-Life Estimates

temporal aggregation (P)	true half-life (H, days)								
	1	2	7	14	31	90	180	365	730
	estimated half-life (H*, days)								
1	1.0	2.0	7.0	14.0	30.0	90.0	180.0	365.0	730.0
2	1.4	2.7	9.4	18.7	40.1	120.1	240.1	486.7	973.4
7	2.2	3.7	11.0	21.4	45.1	134.2	267.8	542.5	1,084.5
14	3.3	4.7	11.8	22.2	46.1	135.9	270.5	547.3	1,093.4
30	5.5	7.0	13.8	24.0	47.8	137.6	272.5	549.8	1,097.0
90	12.8	15.1	22.6	31.8	54.4	143.4	278.1	555.4	1,102.8
180	22.3	25.8	35.5	45.2	66.4	152.8	286.7	563.6	1,110.9
365	40.1	45.6	59.5	71.7	93.7	175.2	306.2	581.4	1,127.8
730	72.3	81.1	102.0	119.0	146.3	226.7	350.5	619.8	1,162.8
	bias factor (H*/H)								
1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
2	1.41	1.37	1.34	1.34	1.34	1.33	1.33	1.33	1.33
7	2.25	1.83	1.57	1.53	1.50	1.49	1.49	1.49	1.49
14	3.30	2.33	1.69	1.59	1.54	1.51	1.50	1.50	1.50
30	5.53	3.52	1.97	1.71	1.59	1.53	1.51	1.51	1.50
90	12.76	7.57	3.22	2.27	1.81	1.59	1.54	1.52	1.51
180	22.32	12.91	5.07	3.23	2.21	1.70	1.59	1.54	1.52
365	40.15	22.81	8.50	5.12	3.12	1.95	1.70	1.59	1.54
730	72.32	40.53	14.57	8.50	4.88	2.52	1.95	1.70	1.59

Notes: See text. Given P and H, H^* is calculated from (7).

half-life H^* using the above formula, and the bias factor H^*/H, for various values of the true half-life H and the temporal-aggregation parameter P. We show here just examples of daily data, with true half-lives and sampling frequencies varying between one day, two days, one week, two weeks, one month, one quarter, six months, one year, and two years. The results show that temporal aggregation biases are nontrivial. For example, look at column three: a process with a true half-life of 7 days (one week) in daily data, would be estimated to have a half-life of about 14.0 days (two weeks) with monthly data, 22.6 days (over three weeks) with quarterly data, and 59.5 days (over eight weeks) with annual data. Or look at a process of very slow convergence, in the penultimate column: a process with a half-life of one year in daily data would appear to have a half-life of almost two years in annual data.

In fact, aside from noting the size of these biases, some more systematic variation is apparent. First, we note that the bias is always upwards: the estimated half-life always exceeds the true half-life, except in the trivial case when there is no

aggregation (that is, $H^* > H$ except that $H^* = H$ when $P = 1$, in the first row). Second, reading across the table, we note that the estimated half-life appears to be an increasing function of the true half-life, as expected (that is, for a given P, H^* is an increasing function of H, or, equivalently, a decreasing function of ρ). Third, reading down the table, it appears that the size of the bias increases as the degree of temporal aggregation increases (that is, for given H, H^* is an increasing function of P). Fourth, reading across the lower table, it seems that for given P, the bias factor is a decreasing function of H, and converges monotonically to a limit for large H. Fifth, it appears that the extent of the bias potentially increases without limit, as might be guessed from reading down the first column (that is, it could be that H^*/H tends to infinity as P tends to infinity). Sixth, we may note that above and to the right of the diagonal in the lower panel, the bias factor is bounded; it reaches its highest value on the diagonal; and as we move down the diagonal, this bias factor appears to reach an upper limit of at most 59%. The last observation is encouraging, since it suggests that we might be able to limit temporal aggregation biases by appropriate choices of sampling frequency.

The above observations are driven by the numerical examples of Table 1, but to deliver more formal conclusions I now show that all are true in general via the following set of results.

Proposition 2 *Let H be the true half-life, P the degree of temporal aggregation, and $H^*(P; H)$ the estimated half-life, as defined at (7). Then:*

(a) $H^*/H = 1$ when $P = 1$, and $H^*/H > 1$ when $P > 1$;

(b) H^*/H and H^* are strictly increasing in P;

(c) H^* is strictly increasing, H^*/H strictly decreasing, in H when $P > 1$;

(d) As P tends to infinity with H fixed, H^*/H tends to infinity;

(e) As H tends to infinity with P fixed, H^*/H tends to a finite limit

$$3P^2/(2P^2 + 1);$$

(f) If P and H tend to infinity together with $P = cH$ for some constant c, then H^*/H is strictly increasing and tends to a finite limit

$$\frac{c \ln 2}{\ln \left(\frac{2c \ln 2 - 2(1 - 2^{-c})}{(1 - 2^{-c})^2} \right)};$$

(g) If $2 \leq P \leq H$ then H^* exceeds H by at least 33% and at most 59% (approximately, rounding to 2 significant figures).

Proof See appendix. □

8

With these results we can now characterize the impact of temporal aggregation on half-life estimates in mean-reversion tests of PPP and LOOP.

Consider first the upper right portion of Table 1, on or above the diagonal. The results here are not too disturbing. The minimal bias is H^* 33% above the true H, which obtains for $P = 2$ as H grows large. Closer to the diagonal the bias can be larger, but moving right-to-left, or top-to-bottom, it increases monotonically and reaches it maximum on the diagonal. On the diagonal itself, the bias increases as we move toward the bottom-right corner, with $P = H$ tending to infinity; here the upper bound on the bias is 59%. Thus, when the extent of aggregation P is small relative to the half-life H, the bias is quite narrowly bounded between plus 33% and plus 59%. A crude correction, say of dividing the estimate of H by 1.45, would yield a pretty good estimate within about 10% of the true H.

The results below the diagonal are, in contrast, very disturbing. These results, where $P > H$, indicate that, when temporal aggregation is large compared to the half-life, the bias can grow very large and is unbounded. The bias is increasing in P for fixed H, and decreasing in H for fixed P. In this region, the half-life bias can not untypically amount to a several-fold multiplication of the true half-life.

The lessons are clear: in order to get fairly accurate estimates of half lives we need to ensure that the temporal aggregation in the data is of an order of magnitude no greater than the size of the half-life itself. This will in turn require some prior about what the half-life might be, so as to design a decent experiment. On the other hand, if we sample at points separated by much more than one half-life, we should expect severe biases. This is of great import in the LOOP and PPP literature.

3   Pitfall Two: Linear Specification

The second pitfall involves the problems caused by a failure to recognize nonlinear adjustment dynamics and the use of an inappropriate linear AR(1) specification. Nonlinear adjustment might be caused by transaction costs or other features that could lead to a "band of inaction" for price adjustment, as noted. This idea has recently been developed and applied to the question of PPP and LOOP testing in an ever-expanding set of studies. Many papers use the simplest kind of nonlinear model, the three-regime threshold autoregression (TAR) structure I employ here for illustration (for example, Prakash 1996; Prakash and A. M. Taylor 1997; and Obstfeld and A. M. Taylor 1997). Alternative nonlinear models have also been developed involving more regimes and thresholds or involving smooth transitions with more complex nonlinearities (for example, O'Connell 1998; Michael, Nobay, and Peel 1997; and M. P. Taylor and Peel 1999).

This paper makes a new contribution, and a stronger case for being aware of

nonlinearities, in two ways. First, unlike the large family of empirical studies, I do not examine another specific dataset. Instead, I step back to assess the range of possible biases using a simulation framework under a wide range of plausible parameter values. No overarching simulation study of this kind has been conducted. It is a worthwhile enterprise in that it usefully complements the results derived from particular case studies by giving us an uncluttered view of the essential nature of the problem. Second, I will also examine the question of how the biases due to nonlinearity may interact with those due to temporal aggregation. It is of interest to know whether these two kinds of bias can add up, interact, or cancel out. The question can be probed using simulations that are again tailored to a range of reasonable "real world" parameter values.

The simplest econometric structure which implements the notion of commodity points, a "band of inaction," and surrounding areas of mean reversion, is a symmetric two-regime threshold autoregression. Specifically, suppose that the price gap x_t follows the process

$$
x_t = \begin{cases} +c + \rho(x_{t-1} - c) + e_t & \text{if } x_{t-1} > c; \\ x_{t-1} + e_t & \text{if } c \geq x_{t-1} \geq -c; \\ -c + \rho(x_{t-1} + c) + e_t & \text{if } -c > x_{t-1}; \end{cases} \quad (8)
$$

where $e_t = N(0, \sigma^2)$. This TAR is parameterized by ρ, the autoregressive coefficient for deviations from the band edge, and c, the amplitude of the band edge.

Inside the band the process is a random walk with a unit autoregressive coefficient; outside the band the autoregressive coefficient is ρ. Thus, a naïve application of the standard model (1) would be inappropriate since it would amount to a pooling of observations from two different regimes of behavior. Specifically, in the TAR model (8) it is correct to speak of x_t being at equilibrium in the entire interval $[-c, +c]$ of the band. Outside the band, the process converges to equilibrium by converging to the band edge. It still makes sense, then, to speak of convergence to equilibrium at a speed $\lambda = \rho - 1$, where this is interpreted as convergence relative to the band.

However, if the standard model (1) is applied, it is clear that the estimates of ρ and λ will be biased, and will not converge to the true parameters in the model (8). Depending on the amounts of time the process spends inside and outside the band, the standard model will produce an estimate of the TAR value of ρ biased toward one, and an estimate of the the TAR convergence speed λ biased toward zero. Half-life estimates will, under these circumstances, be biased upwards.

To get a measure of this bias, we note that applying the standard model (1) to the TAR process (8) yields an estimated coefficient $\hat{\rho}$ with

$$
E(\hat{\rho}) = \frac{\text{Cov}(x_t, x_{t-1})}{\text{Var}(x_{t-1})} = \frac{E(x_t x_{t-1})}{E(x_{t-1}^2)} = \frac{E(x_t x_{t-1} I_{in}(x_{t-1}))}{E(x_{t-1}^2)} + \frac{E(x_t x_{t-1} I_{out}(x_{t-1}))}{E(x_{t-1}^2)}
$$

10

$$= \theta_{in} \left(\frac{\text{Cov}(x_t, x_{t-1}|I_{in}(x_{t-1}))}{\text{Var}(x_{t-1}|I_{in}(x_{t-1}))} \right) + \theta_{out} \left(\frac{\text{Cov}(x_t, x_{t-1}|I_{out}(x_{t-1}))}{\text{Var}(x_{t-1}|I_{out}(x_{t-1}))} \right)$$

$$= \theta_{in} \left(\frac{\text{Cov}(x_t, x_{t-1}|I_{in}(x_{t-1}))}{\text{Var}(x_{t-1}|I_{in}(x_{t-1}))} \right) + \theta_{out} \left(\frac{\text{Cov}(x_t, x_{t-1}|I_{out}(x_{t-1}))}{\text{Var}(x_{t-1}|I_{out}(x_{t-1}))} \right)$$

$$= \theta_{in} + \theta_{out}\rho = \rho + \theta_{in}(1 - \rho),$$

where $I_{in}(x_{t-1})$ and $I_{out}(x_{t-1})$ are indicator functions, for inside and outside the band (respectively), and where

$$\theta_{in} = E(I_{in}(x_{t-1})) \left(\frac{\text{Var}(x_{t-1}|I_{in}(x_{t-1}))}{\text{Var}(x_{t-1})} \right);$$

$$\theta_{out} = E(I_{out}(x_{t-1})) \left(\frac{\text{Var}(x_{t-1}|I_{out}(x_{t-1}))}{\text{Var}(x_{t-1})} \right);$$

$$\theta_{out} + \theta_{in} = 1.$$

Thus, roughly speaking, the bias is more severe the more time is spent inside the band: intuitively, a more accurate estimate of ρ can only arise when the process spends more time in the regime (outside the band) where the true parameter is ρ. Finally, with this use of the standard model, we note that using the expected value of ρ we obtain an estimated half-life, which we again write H^*, with

$$H^* = \frac{\ln 0.5}{\ln E(\hat{\rho})} = H \left(\frac{\ln(\rho)}{\ln(\rho + \theta_{in}(1 - \rho))} \right), \tag{9}$$

and clearly $H^* > H$, which shows that the half-life is biased upwards.

How serious are such linear specification biases? It is difficult to obtain a closed-form solution for this bias. Instead we proceed using simulations to give an illustration of how severe these biases might be for selected half-lives, under various forms of the TAR model structure (8). It is clear from inspection that the system (8) is invariant to a rescaling of the variables such that x_t is replaced by x/σ, c is replaced by c/σ, and the error term $e_t = N(0, \sigma)$ is replaced by a new error term $e_t = N(0, 1)$. Thus, for simplicity, we investigate this latter system with a variety of "normalized" thresholds defined by $k = c/\sigma$, and for a variety of outside-the-band half-lives $H = \ln(0.5)/\ln(\rho)$. We generate T=10,000 observations of each TAR process and then estimate the standard AR(1) model. The results appear in Table 2.

These biases turn out to be potentially quite large. Obviously, when $k = 0$ there is no bias; and when k is small, and the process spends little time inside the band, then the bias is also correspondingly small. The bias decreases as the half-life outside the band gets large, as this too means that the process spends relatively little time in the band, ceteris paribus. But for small half-lives, say less than 30 days, and

11

Table 2: Linear-Specification Bias for Half-Life Estimates

threshold parameter (k)	true half-life (H, days)								
	1	2	7	14	30	90	180	365	730
	fraction of observations outside the band								
0.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.10	0.93	0.94	0.97	0.98	0.98	0.99	0.99	1.00	1.00
0.50	0.71	0.77	0.85	0.89	0.92	0.96	0.97	0.97	0.98
1.00	0.54	0.61	0.73	0.80	0.83	0.92	0.92	0.96	0.99
2.00	0.35	0.42	0.60	0.67	0.72	0.82	0.90	0.89	0.90
5.00	0.18	0.23	0.38	0.47	0.59	0.73	0.73	0.79	0.83
	estimated half-life (H *, days)								
0.00	1.0	2.0	7.0	14.0	30.0	90.0	180.0	365.0	730.0
0.10	1.1	2.1	7.2	14.3	30.6	90.6	181.3	366.8	733.3
0.50	1.6	2.7	8.3	15.8	32.6	94.0	185.9	376.6	744.6
1.00	2.2	3.5	9.7	17.6	36.3	97.9	196.0	381.4	739.5
2.00	3.6	5.3	11.9	21.2	41.7	109.4	199.6	411.5	813.3
5.00	7.2	10.0	19.1	30.1	51.4	123.9	245.4	459.4	880.5
	bias factor (H */H)								
0.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.10	1.11	1.07	1.03	1.02	1.02	1.01	1.01	1.00	1.00
0.50	1.57	1.36	1.19	1.13	1.09	1.04	1.03	1.03	1.02
1.00	2.19	1.75	1.38	1.25	1.21	1.09	1.09	1.04	1.01
2.00	3.56	2.65	1.70	1.51	1.39	1.22	1.11	1.13	1.11
5.00	7.20	4.99	2.73	2.15	1.71	1.38	1.36	1.26	1.21

Notes: We generated 10,000 observations of the TAR process (8) for given $\rho = 2^{-1/H}$, $\sigma = 1$, and $k = c$. We then estimated the standard model (1) on this generated data. We derived H^* from the estimated ρ and compared it with the true H.

for large thresholds with a k of at least 0.5, the biases can be considerable, that is, a value of H^*/H of at least 1.5, and often greater than 2.

Thus, we must note that use of a linear specification might also lead to half-life overestimation in conventional approaches to testing PPP and LOOP. If, as seems plausible in many instances, there is a "band of inaction" corresponding to the limits of price arbitrage, then such nonlinearities need to be properly modeled.[8]

[8]One example of such an approach was given by Obstfeld and A. M. Taylor (1997), who showed that PPP adjustment dynamics might more closely fir TAR rather than AR dynamics. In their econometric work, they found half-lives were overestimated by at least a factor of two using monthly data for about thirty world locations. Half-lives measured at 2 to 3 years using AR methods were revised down to about 12 months using the TAR method. Thus, the corrections we have identified as having theoretical importance are potentially important in practice.

4 Combined Biases

It is of interest to briefly ask: what if the two biases we have already identified are present simultaneously? If these biases operate in any kind of additive fashion, then, for any H, nonlinear dynamics (measured by thresholds k) and temporal aggregation (measured by period averaging over P periods) could together yield even bigger biases than we have seen so far with each source of bias operating independently.

Following the method of the last section, it is simplest again to proceed via simulation. We simulate the TAR model (8) with a variety of "normalized" thresholds defined by $k = c/\sigma$, and for a variety of outside-the-band half-lives $H = \ln(0.5)/\ln(\rho)$. Having generated single-period data in this fashion, we then generate temporally aggregated data by taking P-period averages of these generated series, to generate a new low-frequency series. We proceed in this fashion to generate $T =$10,000 observations of each low-frequency series, for each H, k, and P. The results are shown in Table 3. The table shows results for values of P from 2 to 90, for compactness. Nonetheless, as compared with Tables 1 and 2, here we can clearly see that the two sources of bias we have examined, powerful as they are individually, can be even more serious when combined. For example, suppose there is a threshold $k = 2$, a period-averaging over $P = 7$ periods, and a true half-life of $H = 14$ periods. The earlier tables would have suggested an overestimate by a factor of only 1.53 due to temporal aggregation bias, and 1.51 due to linear specification bias. But here, with both biases present, we would overestimate the half-life of this process by a factor $H^*/H = 2.39$. Or take a more extreme example, with $k = 2$, $P = 30$, and $H = 7$, where a weekly half-life nonlinear TAR is being estimated on monthly averaged data. The earlier tables would have suggested an overestimate by a factor of 1.97 due to temporal aggregation bias, and 1.70 due to linear specification bias. But here, with both biases present, we would overestimate the half-life of this process by a factor $H^*/H = 3.47$. In both these examples, and most others reported here, the effect is approximately multiplicative. In general the effects seen in each case individually still hold in combination: the extent of the bias is usually increasing in P and k and decreasing in H, ceteris paribus.

The main conclusion to be drawn here is that both sources of bias could be severe enough on their own to warrant concern. If both biases are operating together, as may be plausible in any characterization of real world data, it might be necessary to re-examine the nature of the PPP and LOOP puzzles that currently abound in the literature. A half-life of price deviations of four or five years might indeed seem like a puzzle, if one is expecting half-life of a year or less. But if averaging in the data and arbitrage costs are not being properly accounted for in the econometric methodology, it could easily be the case that a four or five fold bias in the half-life has been introduced by sampling and specification problems.

Table 3: Combined Bias for Half-Life Estimates

temporal aggregation (P)	threshold parameter (k)	true half-life (H, days)								
		1	2	7	14	30	90	180	365	730
		bias factor (H */H)								
2	0.00	1.39	1.37	1.32	1.35	1.81	1.16	1.05	0.93	2.23
	0.10	1.57	1.45	1.33	1.31	1.38	1.53	1.33	1.56	0.89
	0.50	2.27	1.92	1.58	1.47	1.48	1.58	1.47	1.26	0.95
	1.00	3.52	2.51	1.86	1.82	1.41	1.52	1.41	2.06	1.38
	2.00	6.67	4.59	2.61	2.22	1.79	1.65	1.56	1.42	1.34
	5.00	23.67	13.79	5.94	4.49	2.93	2.49	1.55	1.98	1.40
7	0.00	2.24	1.75	1.58	1.57	1.63	1.48	1.61	1.48	1.36
	0.10	2.38	1.94	1.64	1.55	1.54	1.48	1.82	1.68	1.69
	0.50	3.15	2.36	1.83	1.67	1.59	1.48	1.51	1.68	1.30
	1.00	4.47	3.27	2.25	1.94	1.80	1.62	1.55	1.63	1.38
	2.00	8.29	5.37	3.16	2.39	2.13	1.58	1.51	1.67	1.48
	5.00	31.38	16.99	6.94	4.92	3.24	2.33	1.73	2.09	1.45
14	0.00	3.40	2.31	1.72	1.59	1.55	1.47	1.35	1.38	1.64
	0.10	3.30	2.63	1.73	1.62	1.53	1.66	1.45	1.65	1.45
	0.50	4.35	2.90	1.99	1.83	1.59	1.55	1.45	1.54	1.48
	1.00	5.32	3.50	2.29	2.07	1.82	1.77	1.75	1.45	1.74
	2.00	9.33	6.11	3.19	2.66	2.25	1.82	1.58	1.61	1.96
	5.00	31.08	18.17	7.85	4.80	3.47	2.47	2.04	2.04	1.96
30	0.00	5.25	3.38	2.03	1.70	1.58	1.56	1.47	1.60	1.40
	0.10	4.84	3.90	1.94	1.70	1.62	1.54	1.43	1.53	1.48
	0.50	6.79	4.52	2.19	1.93	1.70	1.66	1.48	1.64	1.63
	1.00	7.09	5.00	2.77	2.12	1.89	1.73	1.77	1.46	1.76
	2.00	11.44	7.30	3.47	2.77	2.19	1.82	1.72	1.86	1.60
	5.00	34.37	18.89	7.84	5.08	3.66	2.55	2.28	1.85	1.84
90	0.00	7.14	7.23	2.65	2.45	1.83	1.51	1.54	1.50	1.69
	0.10	10.60	8.55	2.99	2.34	1.81	1.62	1.60	1.51	1.55
	0.50	16.32	8.57	3.36	2.52	1.96	1.67	1.65	1.45	1.50
	1.00	17.72	9.12	3.78	2.83	2.12	1.74	1.67	1.69	1.54
	2.00	21.66	12.20	4.52	3.32	2.44	1.98	1.73	1.65	1.50
	5.00	43.39	23.14	8.73	5.79	3.94	2.53	2.14	1.86	1.67

Notes: See text and Table 2. Note that, in principle, the results for $k = 0$ should correspond exactly to those in Table 2. However, these results are generated by simulation in finite samples and not from an exact formula. Thus, there are small discrepancies between the two tables, especially for larger P, where aggregation means a shorter span in the finite-sample averaged data.

5 Power of Unit Root Tests

Thus far I have considered only the question of bias as it applies to the estimates of convergence speed and half-life in standard AR(1) tests. That is, I have only considered point estimates that relate to the issue of quantitative significance, the problem of how large we think the half-life is. Obviously, having shown how biased the AR(1) estimate can be under certain conditions, it behooves me to spell out briefly what the implications might be for the main test of statistical significance invoked in the PPP and LOOP literature, the unit root test.

Intuitively, the impact is easy to guess: power will be reduced by these types of biases. Consider for example, the standard Dickey-Fuller test. It is the t-test for the restriction of the unit root null, $\rho = 1$, in the AR(1) regression we have been studying. But if the two pitfalls we have considered tend to bias ρ upwards toward 1, then, clearly, this will bias the t-statistic downward, making it harder to reject the unit root null.[9] For illustration, I consider the simplest Dickey-Fuller t-test with no lags. For each set of these true parameters, I generated 10,000 simulations of the series, each with T observations, and I then applied the Dickey-Fuller test. The results report the power of the test for a $p = .05$ significance level under this scenario—that is, the fraction of times the test rejected a unit root.

There is one delicate problem, however, that still needs to be addressed: what span of data T should be chosen? It is well known in the unit root literature that the main determinant of the power of a conventional unit root test is neither the length nor the frequency of a dataset, but its span (Shiller and Perron 1985). This point was noted in the context of the PPP debate by Jeffrey Frankel (1991).

Frankel showed that, for the consensus half-life estimate of, say, 4–5 years, one would need typically about 100 years of data to have an adequately powerful test. He did so by invoking the asymptotic standard error (ASE) formula for the AR(1) coefficient ρ, where $\text{ASE}^2 = (1 - \rho^2)/T$. Suppose that we have annual data and the true half life is 5 years, so that $\rho = .87$. Consider the Dickey Fuller t-test of $\rho = 1$; to get below the asymptotic $p = .05$ critical value of $t = -3$ we would need $T = 3^2(1 - .87^2)/(1 - .87)^2 = 128$ observations. With a half life of 4 years this would fall to 104 observations.

Clearly, we cannot go any further without appreciating the relevant scaling issues in the time dimension: we might reasonably surmise that doubling the half-life and doubling the span should essentially leave the power problem unaffected. This is certainly true as an approximation for large H, or equivalently, for small $\lambda = \rho - 1 < 0$. Using approximations for small λ in the ASE formula, it is easy

---

[9]For a discussion of stationarity testing with thresholds see Balke and Fomby (1997), who develop a cointegration test for use with TAR models.

Table 4: Power of Unit Root Tests

temporal aggregation (P)	threshold parameter (k)	25H half-life (H, days) 1	25H 7	25H 30	50H half-life (H, days) 1	50H 7	50H 30	100H half-life (H, days) 1	100H 7	100H 30
1	0.00	0.52	0.63	0.65	0.97	1.00	1.00	1.00	1.00	1.00
	0.50	0.26	0.47	0.56	0.77	0.98	1.00	1.00	1.00	1.00
	2.00	0.11	0.17	0.33	0.16	0.64	0.95	0.44	1.00	1.00
7	0.00	0.52	0.22	0.23	0.97	0.76	0.88	1.00	1.00	1.00
	0.50	0.26	0.15	0.17	0.77	0.60	0.81	1.00	1.00	1.00
	2.00	0.11	0.06	0.08	0.16	0.18	0.51	0.44	0.85	1.00
30	0.00	0.52	0.22	0.21	0.97	0.76	0.76	1.00	1.00	1.00
	0.50	0.26	0.15	0.17	0.77	0.60	0.69	1.00	1.00	1.00
	2.00	0.11	0.06	0.10	0.16	0.18	0.42	0.44	0.85	0.99

Notes: See text.

to show that $\text{ASE}^2 = 2|\lambda|/T + O(\lambda^2)$, and hence $t^2 \approx |\lambda|T/2$. Next, using the approximation $H \approx \ln 2/|\lambda|$, we obtain $t^2 \approx \frac{1}{2}(\ln 2)(T/H)$. This last formula shows that the test is approximately invariant to time rescaling in T and H, and the test is likely to be more powerful the higher is the span T relative to the half-life H.

With this in mind, let us now focus on the ratio T/H, the span as it relates to the half-life. If we aim to be below $t = -3$, the $p = .05$ critical level, then we must aim a span such that $T/H > 2t^2/\ln 2 \approx 25$. The above suggests that, for the consensus half-life in the PPP literature, one needs a span equal to about 25 times the half-life as a minimum, that is, about 100 to 125 years of data. More powerful tests would result if the span were larger, say, 50 or 100 times the half-life. Reassuringly, we do know that long-run tests run over a century or more of data can reject a unit root (Lothian 1996; Lothian and M. P. Taylor 1996, 1997; A. M. Taylor 1996). Given this intuition, let us now play with some simulated data to see how the power of the test is also affected by a variety of spans.

The results are shown in Table 4. To save space, I will consider a limited range of half-lives ($H = 1, 7, 30$), temporal aggregations ($P = 1, 7, 30$), and thresholds ($k = 0, 0.5, 2$). The data spans employed are $T = 25H, 50H, 100H$, as indicated above. The implications are quite striking. Consider first the situation with no aggregation and no threshold ($P = 1, k = 0, T = 25H, 50H, 100H$). This corresponds to what we know already in the standard situation: power is about 0.5–0.6 with $T = 25H$ but rises close to 1.0 at $T = 50H$, and equals 1.0 at $T = 100H$. Next, allow in some temporal aggregation with a short span ($P = 1, 7, 30; k = 0; T = 25H$). We see that power is not affected at all for the

case $H = 1$. But when $H > 1$ power falls dramatically with aggregation, from around 0.6 to around 0.2. At least if we throw enough data at the test this problem goes away ($P = 1, 7, 30; k = 0; T = 50H, 100H$). Finally, toss in a threshold when there is a short span ($P = 1; k = 0, 0.5, 2.0; T = 25H$), and we again put a severe dent in the power of the test, lowering it from 0.5 to around 0.1. This problem is there for all values of the half life H, and temporal aggregation P. Doubling the span ($T = 50H$) alleviates the problem somewhat, though more in the case of the longer half life; a very long span ($T = 100H$) again goes to show that if you can muster enough data your inferences should be sound.[10]

Three main results appeared in this section. First, and most importantly, the power of unit root tests can be spectacularly weakened by temporal aggregation and nonlinearity problems. Second, if you have enough data to throw at the problem, it will probably go away. Third, when nonlinearities are present the power of the test is no longer just a function of the span of the data in relation to the half life; it also depends on the span of the data too. The first two results directly bear on PPP and LOOP testing, and might explain why AR(1) stationarity tests rarely reject a unit root: weak as those tests are, they may be even weaker than we thought once we allow for the two pitfalls. The third results shows another interesting deviation from the Shiller and Perron (1985) finding that power depends only on span for linear AR tests; the result seems not to hold once nonlinearities enter the picture, just as it fails when moving average terms are present (Ng 1995).

6   Implications

The conclusions of the empirical and theoretical work are clear enough, but would be of no consequence could we not demonstrate the importance of these potential biases in real applications. In this concluding section I claim that key contemporary sources used for tests of PPP and LOOP suffer from various deficiencies in empirical design and sampling, including the temporal aggregation biases noted above.

Probably the most widely used international macroeconomic dataset today is the International Monetary Fund's regularly updated series *International Financial Statistics* (IMF 1998 and preceding issues). Published in monthly and annual volumes, and now available on CD-ROM, this data set has been the basis for countless empirical articles in macroeconomics and international finance. Inevitably, the data on exchange rates and prices has been used in many econometric studies of PPP.

---

[10]Ng (1995) showed that with certain kinds of ARMA forms the power of the test may not be monotonically increasing in the span, as in the standard AR case studied by Shiller and Perron (1985). For reasons of space, we only explore a small range of spans here, so we cannot say whether the same non-monotonicity affects our power-versus-span relationship.

However, although some of these studies have risen to remarkable levels of econometric sophistication, similar care over the sampling and construction of the underlying real exchange-rate data cannot be assumed. Recall that for clean use in, say, a monthly data set, we would be hoping for the primary source to contain temporally unaggregated (say, end-of-month) series for exchange rates and price indices. The former is not in too much doubt, as the foreign exchange market has always been easy to observe and record at high frequency. The latter is a vain hope.

Of course, the scarce documentation supplied in the print copy of *International Financial Statistics* (and even less on the CD-ROM) does not permit the scholar full information on the actual techniques used. For this, a deeper archival dig is required to unearth one of the obscure supplements to the *International Financial Statistics* series, number 12, published in 1986, which is concerned with the methodology used to compute the price statistics. Aside from its turgid style and far-from-scintillating subject matter, this supplement makes for discouraging reading even for one concerned with the matters at hand. PPP studies for the contemporary cross-section of nations presume that the observed data corresponds to the true data in many ways. All prices are assumed to be accurate. All prices are supposed to be for comparable goods at each location. All prices should be market prices, whether for commodities, manufactured goods or services, and whether the good is freely traded or subject to some market controls. Prices should be collected at representative locations in each economy. The supplement bravely spells out the extensive desiderata, only to crush the hopes of the wishful scholar by confessing all of the manifold ways in which these criteria cannot be met by price-collecting statisticians in over one hundred countries, some developed and some very underdeveloped.

The news so far is bad enough. But as regards our concerns over temporal aggregation there is yet worse news to come. To meet the desired standard we would be hoping that hundreds of price inspectors would leave a hundred or more capital cities on the final day of each month, scour every market in all representative locations, for all goods, and come back at the end of a long day, with a synchronized set of observations from Seoul to Santiago, from Vancouver to Vanuatu. We cannot pretend that this happens. Moreover, the IMF doesn't pretend this happens. They have given up, and one does not need the supplement to know this. In every month's issue of the *International Financial Statistics*, in stark italics above each time series of price indices, one finds the brief notation: "current period averages."[11]

That is, each monthly, quarterly, or annual observation on prices in the IMF

---

[11]Although the IMF labels this column a monthly average in every country table this does *not* imply that every country's data is, in fact, always a monthly average. The IMF documentation is vague, but for many OECD countries, it appears that, based on what we know of national sources, the price data is sampled on a particular day of the month, notwithstanding the IMF labels. I thank Charles Engel for alerting me to this problem.

data has precisely the form of temporal aggregation we have been studying: not merely infrequent sampling, but time averaging. At least this is so in principle. In practice, even the supplement cannot go into the details needed to tell us when the price of bread was calculated in Bangkok versus when the price of milk was figured out in Bangui. And in reality some of the data may be polluted in different ways in different places. There may be more or less averaging in some countries than in others. All of these imperfections make the empirical question far less clean than the standard thought-experiments considered in this paper and in the rest of the literature.

This example picks on the IMF, but similar concerns affect many price collection exercises. Admittedly, the month-to-month (and, to a lesser extent, year-to-year) variations in prices are generally dwarfed by the variations in exchange rates, and these are, in the standard sources from the IMF and elsewhere, usually collected more scrupulously on a regular end-of-period basis. To that extent, the problem may be mitigated, though that is almost certainly less true for most historical studies where lower frequency (quarterly and annual) time series for exchange rates are often a melange of period averages, end-of-period, and other approximations. One could assess where and how averaging versus end-of-period data affects the results by comparing the tests on those samples where we can use both kinds of data, for example in certain OECD countries, though such an exercise is beyond the scope of this paper.

Different sources may suffer greater or lesser degrees of time averaging in their construction. But averaging is often an issue. It is out there polluting more sources than we know, perhaps in more ways than we know. Its implications for calculations of convergence speeds in price processes need to be considered, and the present paper takes one step in that direction. These are important concerns that need to be addressed before the extent of the PPP puzzle can be accurately gauged.

7 Conclusion

The pollution of data by temporal aggregation and the existence of linear estimation biases create serious pitfalls on the way to an understanding of the PPP puzzle. If measurements of half-lives can be out by, say, a factor of ten, then we need to— first, put the puzzle idea on hold for a while until we know more; second, search for higher-frequency data with which to pursue testing, looking for frequencies that match the underlying arbitrage process; third, consider the implications of nonlinear models that might deliver more rapid adjustments outside any "bands of inaction." Furthermore, we are dogged by the same problems if we want to execute reliable unit root tests.

Researchers, it would seem, are between a rock an a hard place. In an ideal world high-frequency data could solve our problems. But we have been forced to use low-frequency data for most PPP and LOOP testing in contemporary and historical data, because that's all we have had available. With rare exceptions, high-frequency data has only existed for short spans. Long-span low-frequency data has been used to support PPP and LOOP using more than a century of data with linear tests, but as our unit root tests indicate, even this modest glimmer of hope may be illusory: in many circumstances, nonlinearities and temporal aggregation can offset the power-boosting effects of a lengthened span. Finally, even in the linear studies that do reject a unit root the estimated half-lives might still be way off target. If a complete reworking of the research agenda were contemplated, a case can be made that we should: (i) adopt a new empirical strategy of developing shorter spans of high-frequency data to get more powerful tests of stationarity and more reliable estimates of half lives; and (ii) consider replacing the unit root null with a stationary null, given that PPP and LOOP are so firmly entrenched in the profession's priors.[12]

It looks like the two teams in the PPP debate are heading into overtime. They'll be out on the playing field for a good while yet, so the referee can postpone the final whistle, the fans watching on the sidelines will almost certainly grow restless if they haven't gone home already, the players should stay warmed-up, and of course there is still plenty of time for everyone to change teams a few more times.

---

[12]I thank Mark Taylor for comments on these issues.

Appendix

Proof of Proposition 1

From (3) we obtain

$$\mathrm{E}(\rho^*) = \frac{\mathrm{Cov}(x_{Ps} + \ldots + x_{Ps+P-1}, x_{Ps-P} + \ldots + x_{Ps-1})}{\mathrm{Var}(x_{Ps-P} + \ldots + x_{Ps-1})} = \frac{n(P)}{d(P)},$$

where $n(P)$ and $d(P)$ refer to the numerator and denominator. By consideration of the true process (1), we know that, for any k, the kth autocorrelation coefficient is given by $\mathrm{Cov}(x_t, x_{t-k}) = \rho^k \mathrm{Var}(x)$, where $\mathrm{Var}(x)$ is the unconditional variance of x. Thus, we can obtain the following expressions using the numerator $n(P)$ and denominator $d(P)$,

$$N(P) = n(P)/\mathrm{Var}(x) = \begin{array}{llll} \rho & +\rho^2 & +\ldots & +\rho^P \\ +\rho^2 & +\rho^3 & +\ldots & +\rho^{P+1} \\ +\ldots & +\ldots & +\ldots & +\ldots \\ +\rho^P & +\rho^{P+1} & +\ldots & +\rho^{2P-1} \end{array}$$

$$= \rho(1 + \rho + \rho^2 + \ldots + \rho^{P-1})^2 = \rho(1 - \rho^P)^2/(1-\rho)^2;$$

$$D(P) = d(P)/\mathrm{Var}(x) = \begin{array}{llll} 1 & +\rho & +\ldots & +\rho^{P-1} \\ +\rho & +1 & +\ldots & +\rho^{P-2} \\ +\ldots & +\ldots & +\ldots & +\ldots \\ +\rho^{P-1} & +\rho^{P-2} & +\ldots & +1 \end{array}$$

$$= P(1 - \rho^2) - 2\rho(1 - \rho^P)/(1 - \rho)^2;$$

the final formula for $D(P)$ can be obtained by induction; note that $D(1) = 1$, so that

$$D(P) = D(P-1) + 1 + 2\rho + \ldots + 2\rho^{P-1} = D(P-1) + \frac{2(1 - \rho^P)}{(1 - \rho)} - 1;$$

and it is straightforward to solve this difference equation for $D(P)$.

Dividing the formula for $N(P)$ by that for $D(P)$ yields the desired expression for ρ^*.

$\square$

Proof of Proposition 2

(a) The first property is trivial. The second follows from (b).

(b) Consider the gradient of the function $h = H^*/H$ with respect to the arguments H and P. We can write

$$\partial h/\partial P = (\partial h/\partial P, \partial h/\partial H)(1, 1)^T + (-\partial h/\partial H) > 0$$

where, by the monotonocity properties (c) and (f), we know that both terms in the middle expression are positive. Hence H^*/H, and by extension H^* are strictly increasing in P.

(c) Let $E = \partial \ln \rho^{*\frac{1}{P}}/\partial \ln \rho$. I claim it suffices to show that $0 < E < 1$.

First, if $0 < E$ then the mapping $\ln \rho^{*\frac{1}{P}} \to \ln \rho$ is strictly increasing. But the mappings $\ln \rho^{*\frac{1}{P}} \to H^*$ and $\ln \rho \to H$ are strictly increasing, so immediately it follows that the map $H \to H^*$ is strictly increasing too.

Second, if $E < 1$ then $\frac{\partial(\ln \rho^{*\frac{1}{P}} - \ln \rho)}{\partial \ln \rho} < 0$. But note that

$$H^*/H = \ln \rho / \ln \rho^{*\frac{1}{P}} = (1 - x)^{-1}$$

where $x = (\ln \rho^{*\frac{1}{P}} - \ln \rho)/\ln \rho > 0$.

Now clearly H^*/H is increasing in x. As for x itself, its denominator is negative and strictly increasing in $\ln \rho$, and since $E < 1$ its numerator is negative and strictly decreasing in $\ln \rho$. Therefore the mapping $x \to \ln \rho$ is strictly decreasing. But we know already that the mappings $\ln \rho \to H$ and $x \to H^*/H$ are strictly increasing. Hence, we can infer that the mapping $H \to H^*/H$ is strictly decreasing.

Thus, our proof is complete if we can show $0 < E < 1$. We compute the elasticity next using perturbation analysis and show that the inequalities are satisfied. Let

$$y = \ln \rho^{*\frac{1}{P}} = \frac{N}{D} = \frac{(\rho(1 - \rho^P)^2)^{\frac{1}{P}}}{(P(1 - \rho^2) - 2\rho(1 - \rho^P))^{\frac{1}{P}}}$$

where N and D refer to the numerator and denominator respectively.

Consider a perturbation where ρ changes to $\rho' = \rho(1 + \epsilon)$. The perturbation yields a new value of y given by $y' = N'/D'$ where

$$N' = \left\{ \rho(1 + \epsilon)(1 - \rho^P(1 + \epsilon)^P)^2 \right\}^{\frac{1}{P}} + O(\epsilon^2)$$

$$= N \left\{ 1 + \epsilon \left(\frac{1}{P} - \frac{2\rho^P}{(1 - \rho^P)^2} \right) \right\} + O(\epsilon^2),$$

$$D' = \left\{ P(1 - \rho^2 - 2\epsilon\rho^2) - 2\rho(1 + \epsilon)\left(1 - \rho^P(1 + P\epsilon) \right) \right\}^{\frac{1}{P}} + O(\epsilon^2)$$

$$= D \left\{ 1 + \epsilon \left(\frac{1}{P} - \frac{1 - 2\rho^{P+1} + \rho^2}{P(1 - \rho^2) - 2\rho(1 - \rho^P)} \right) \right\} + O(\epsilon^2),$$

and we have omitted some tedious intermediate steps. Thus, we find that

$$y' = \frac{N'}{D'} = \frac{N}{D}(1 + \epsilon E) + O(\epsilon^2),$$

where

$$E = -\frac{2\rho^P}{(1 - \rho^P)^2} + \frac{1 - 2\rho^{P+1} + \rho^2}{P(1 - \rho^2) - 2\rho(1 - \rho^P)}.$$

It is trivial to show that $E = 1$ when $P = 1$.

The first property we show is that $E > 0$. Rearranging the above formula, this property holds if and only if

$$L_P \equiv (1 - \rho^P)^2(1 - 2\rho^{P+1} + \rho^2) > 2\rho^P P(1 - \rho^2) - 2\rho(1 - \rho^P) \equiv R_P.$$

Consider a fixed ρ. Trivially the property holds for $P = 1$ when $E = 1 > 0$. It is also easily shown to hold for $P = 2$. We show it holds for all $P > 2$ by showing that L_P is

22

increasing in P and R_P decreasing in P. The former is clear by inspection. For the latter it can be shown that,

$$R_{P+1} - R_P = 2\rho^P (1 - \rho) \left\{ (1 - P) + 3\rho(1 - \rho^P) \right\} < 0,$$

where the inequality holds because the term in braces is negative for $P > 2$.

Now we conclude by showing that $E < 1$. Rearranging terms, this property holds if and only if

$$\frac{1 - 2\rho^{P+1} + \rho^2}{P(1 - \rho^2) - 2\rho(1 - \rho^P)} < \frac{2\rho^P}{(1 - \rho^P)^2} + 1 = \frac{1 + \rho^{2P}}{(1 - \rho^P)^2};$$

and rearranging again this holds if and only if

$$0 > (1 - 2\rho^{P+1} + \rho^2)(1 - \rho^P)^2 - (1 + \rho^{2P})(P(1 - \rho^2) - 2\rho(1 - \rho^P)).$$

Collecting terms on the right, this may be written

$$0 > (1 - P)(1 + \rho^2 + \rho^{2P} + \rho^{2P+2}) - r\rho - 2(\rho^P - \rho^{P+2}) - 2\rho^{2P+1},$$

and clearly all the terms shown here are negative for $P > 1$.

(d) For fixed H, and, thus, fixed ρ, we use (6) and the fact that $\rho = 2^{-1/H}$ to show that

$$\frac{H^*}{H} = \frac{P \ln(\rho^{-1})}{\ln\left(\frac{P(1 - \rho^2) - 2\rho(1 - \rho^P)}{\rho(1 - \rho^P)^2} \right)}.$$

By considering the limiting properties of this expression as $P \to \infty$, given that $0 < \rho < 1$, we can see that

$$\frac{H^*}{H} \sim \frac{P \ln(\rho^{-1})}{\ln(P(1 - \rho^2)/\rho)} \sim \frac{P \ln(\rho^{-1})}{\ln P},$$

and clearly this last term tends to infinity.

(e) Let $x = \rho^{-1} > 1$. Then, from (6) we can obtain

$$\frac{H^*}{H} = \frac{P \ln x}{\ln\left(\frac{P(1 - x^{-2}) - 2x^{-1}(1 - x^{-P})}{x^{-1}(1 - x^{-P})^2} \right)}.$$

The limit $H \to \infty$ corresponds to the limit $\rho \to 1$ from below, and thus to $x \to 1$ from above. Therefore, let $x = 1 + z$, with $z > 0$. Then the above numerator can be written as

$$zP + O(z^2).$$

The denominator, with much tedious manipulation of Taylor expansions, can be written as

$$\ln\left(1 + \frac{P(1 - x^{-2}) - 2x^{-1}(1 - x^{-P}) - x^{-1}(1 - x^{-P})^2}{x^{-1}(1 - x^{-P})^2} \right)$$

23

$$= \ln\left(1 + \frac{z^3(\frac{2}{3}P^3 + \frac{1}{3}P) + O(z^4)}{z^2 P^2 + O(z^3)}\right) = \frac{z(2P^2 + 1)}{3P} + O(z^2),$$

where the intermediate step uses the substitutions

$$x^{-1} = 1 - z + z^2 - z^3 + O(z^4); \quad x^{-2} = 1 - 2z + 3z^2 - 4z^3 + O(z^4);$$

$$x^{-P} = 1 - Pz + \frac{P(P+1)}{2}z^2 - \frac{P(P+1)(P+2)}{6}z^3 + O(z^4).$$

Dividing numerator by denominator yields the desired result as $z \to 0$.

(f) Let $x = \rho^{-1} > 1$, and set $P = cH$. Then $P = c\ln 2/\ln x$, and $x^{-P} = \rho^P = \rho^{cH} = 2^{-c}$, and from (6) we can obtain

$$\frac{H^*}{H} = \frac{c\ln 2}{\ln\left(\frac{(c\ln 2/\ln x)(1-x^{-2})-2x^{-1}(1-2^{-c})}{x^{-1}(1-2^{-c})^2}\right)} = \frac{c\ln 2}{\ln\left(\frac{c\ln 2}{(1-2^{-c})^2}\frac{(x-x^{-1})}{\ln x} - 2(1-2^{-c})^{-1}\right)}$$

Thus,

$$\lim_{H\to\infty}\frac{H^*}{H} = \frac{c\ln 2}{\ln\left(\frac{c\ln 2}{(1-2^{-c})^2}\lim_{x\to 1}\left\{\frac{(x-x^{-1})}{\ln x}\right\} - 2(1-2^{-c})^{-1}\right)}$$

$$= \frac{c\ln 2}{\ln\left(2c\ln 2(1-2^{-c})^{-2} - 2(1-2^{-c})^{-1}\right)},$$

where one can use L'Hôpital's rule to show that $\lim_{x\to 1}(x - x^{-1})/\ln x = 2$.

To show monotonicity, it suffices to show that $\partial(H^*/H)/\partial x < 0$. Now, by the above formulae, we see that H^*/H is a strictly decreasing function of $(x - x^{-1})/\ln x$. So it suffices to show that the latter is strictly increasing in x. The result follows since

$$\frac{\partial}{\partial x}(\frac{x - x^{-1}}{\ln x}) = \frac{(1 + x^{-2})\ln x - (x - x^{-1})x^{-1}}{(\ln x)^2} = \frac{x^{-2}(x^2 + 1)}{(\ln x)^2}\left\{\ln x - \frac{x^2 - 1}{x^2 + 1}\right\} > 0,$$

and this property holds because the term in braces is positive for $x > 1$.

(g) By the monotonicity properties in (b) and (c), we see that, for $P \le H$,

$$\lim_{H\to\infty}\frac{H^*(2;H)}{H} \le \frac{H^*(2;H)}{H} \le \frac{H^*(P;H)}{H} \le \frac{H^*(H;H)}{H} \le \lim_{H\to\infty}\frac{H^*(H;H)}{H}.$$

Hence, by the limit formulae in (e) and (f), with $c = 1$,

$$\frac{4}{3} = \frac{3.(2^2)}{2.(2^2) + 1} \le \frac{H^*}{H} \le \frac{\ln 2}{\ln\left(2\ln 2(1 - 2^{-1})^{-2} - 2(1 - 2^{-1})^{-1}\right)} = \frac{\ln 2}{8\ln 2 - 4}.$$

The lower bound on the left is approximately 1.33 and the upper bound on the right is approximately 1.59; hence the result. □

References

Apte, P., M. Kane, and P. Sercu. 1994. Relative PPP in the Medium Run. *Journal of International Money and Finance* 13 (October): 602–22.

Bacchetta, P., and S. Gerlach. 1997. Consumption and Credit Constraints: International Evidence. *Journal of Monetary Economics* 40 (October): 207–38.

Baillie, R. T., and T. Bollerslev. 1991. Intraday and Inter-Market Volatility in Foreign Exchange Rates. *Review of Economic Studies* 58 (May): 565–585.

Balke, N. S., and T. B. Fomby. 1997. Threshold Cointegration. *International Economic Review* 38 (August): 627–646.

Bayoumi, T. 1993. Financial Deregulation and Consumption in the United Kingdom. *Review of Economics and Statistics* 75 (August): 536–39.

Berloffa, G. 1997. Temporary and Permanent Changes in Consumption Growth. *Economic Journal* 107 (March): 345–58.

Blundell, R. et al. 1995. Financial Liberalization and the Permanent Income Hypothesis. *Manchester School of Economic and Social Studies* 63 (June): 125–44.

Campbell, J. Y., and N. G. Mankiw. 1990. Permanent Income, Current Income, and Consumption. *Journal of Business, Economics, and Statistics* 8 (July): 265–279.

Canzoneri, M. B., R. E. Cumby, and B. Diba. 1999. Relative Labor Productivity and the Real Exchange Rate in the Long Run: Evidence for a Panel of OECD Countries. *Journal of International Economics* 47 (April): 245–66.

Cassel, G. 1922. *Money and Foreign Exchange After 1914*. New York: Macmillan.

Chinn, M., and L. Johnston. 1996. Real Exchange Rate Levels, Productivity and Demand Shocks: Evidence from a Panel of 14 Countries. Working Paper Series no. 5709, National Bureau of Economic Research (August).

Christiano, L. J., M. Eichenbaum, and D. Marshall. 1991. The Permanent Income Hypothesis Revisited. *Econometrica* 59 (March): 397–423.

Cochrane, J. H. 1988. How Big is the Random Walk in GNP? *Journal of Political Economy* 96 (October): 893–920.

Coleman, A. M. G. 1995. Arbitrage, Storage, and the 'Law of One Price': New Theory for the Time Series Analysis of an Old Problem. Princeton University (September). Photocopy.

Dumas, B. 1992. Dynamic Equilibrium and the Real Exchange Rate in a Spatially Separated World. *Review of Financial Studies* 5: 153–180.

Engel, C. 2000. Long-Run PPP May Not Hold After All. *Journal of International Economics*. Forthcoming.

Ermini, L. 1988. Temporal Aggregation and Hall's Model of Consumption Behavior. *Applied Economics* 20 (October): 1317–20.

Ermini, L. 1993. Effects of Transitory Consumption and Temporal Aggregation on the Permanent Income Hypothesis. *Review of Economics and Statistics* 75 (November): 736–40.

Frankel, J. A. 1991. Quantifying International Capital Mobility in the 1980s. In *National Saving and Economic Performance*, edited by B. D. Bernheim and J. B. Shoven. Chicago: University of Chicago Press.

Frankel, J. A., and A. K. Rose. 1996. A Panel Project on Purchasing Power Parity: Mean Reversion Within and Between Countries. *Journal of International Economics* 40 (February): 209–224.

Froot, K. A., and K. Rogoff. 1995. Perspectives on PPP and Long-Run Real Exchange Rates. In *Handbook of International Economics*, vol. 3, edited by G. Grossman and K. Rogoff. Amsterdam: Elsevier.

Glen, J. D. 1992. Real Exchange Rates in the Short, Medium, and Long Run. *Journal of International Economics* 33 (August): 147–166.

Grilli, V., and G. Kaminsky. 1991. Nominal Exchange Rate Regimes and the Real Exchange Rate: Evidence from the United States and Great Britain, 1885–1986. *Journal of Monetary Economics* 27: 191–212.

Hall, R. E. 1988. Intertemporal Substitution in Consumption. *Journal of Political Economy* 96 (April): 339–357.

Heaton, J. 1993. The Interaction Between Time-Nonseparable Preferences and Time Aggregation. *Econometrica* 61 (March): 353–385.

Heckscher, E. F. 1916. Växelkursens Grundval vid Pappersmyntfot. *Ekonomisk Tidskrift* 18 (October): 309–312.

International Monetary Fund. 1986. *International Financial Statistics: Supplement on Price Statistics*. International Financial Statistics. Supplement Series no. 12. Washington, D.C.: International Monetary Fund.

International Monetary Fund. 1998. *International Financial Statistics*. Washington, D.C.: International Monetary Fund.

Jorion, P., and R. J. Sweeney. 1996. Mean Reversion in Real Exchange Rates: Evidence and Implications for Forecasting. *Journal of International Money and Finance* 15 (August): 535–550.

Lam, K. 1990. Working's Effect Revisited: Fitting Univariate Time Series to Stock Price Data. *Omega-International Journal of Management Science* 18 : 337–338.

Lee, M. H. 1978. *Purchasing Power Parity*. New York: Marcel Dekker.

Lo, A. W., and A. G. MacKinlay. 1988. Stock Market Prices Do Not Follow Random Walks: Evidence from a Simple Specification Test. *Review Of Financial Studies* 1: 41–66.

Lothian, J. R. 1997. Multi-Country Evidence on the Behavior of Purchasing Power Parity under the Current Float. *Journal of International Money and Finance* 16 (February): 19–35.

Lothian, J. R., and M. P. Taylor. 1996. Real Exchange Rate Behavior: The Recent Float From the Perspective of the Past Two Centuries. *Journal of Political Economy* 104 (June): 488–509.

Lothian, J. R., and M. P. Taylor. 1997. Real Exchange Rate Behavior. *Journal of International Money and Finance* 16 (December): 945–54.

Maeso-Fernandez, F. 1998. Econometric Methods and Purchasing Power Parity: Short- and Long-Run PPP. *Applied Economics* 30 (November): 1443–57.

Michael, P., A. R. Nobay, and D. A. Peel. 1997. Transaction Costs and Nonlinear Adjustment in Real Exchange Rates: An Empirical Investigation. *Journal of Political Economy* 105 (August): 862–79.

Ng, S. 1995. Testing for Unit Roots in Flow Data Sampled at Different Frequencies. *Economics Letters* 47 (March): 237–42.

O'Connell, P. G. J. 1996. Market Frictions and Relative Traded Goods Prices. *Journal of International Money and Finance* 17 (February): 71–95.

O'Connell, P. G. J. 1998. The Overvaluation of Purchasing Power Parity. *Journal of International Economics* 44 (February): 1–19.

Obstfeld, M., and A. M. Taylor. 1997. Nonlinear Aspects of Goods-Market Arbitrage and Adjustment: Heckscher's Commodity Points Revisited. *Journal of the Japanese and International Economies* 11 (December): 441–79.

Oh, K.-Y. 1996. Purchasing Power Parity and Unit Root Tests Using Panel Data. *Journal of International Money and Finance* 15 (June): 405–418.

Papell, D. H. 1997. Searching for Stationarity: Purchasing Power Parity Under the Current Float. *Journal of International Economics* 43 : 313–32.

Poterba, J., and L. H. Summers. 1988. Mean Reversion in Stock Prices: Evidence and Implications. *Journal of Financial Economics* 22 (October): 27–59.

Prakash, G. 1996. Pace of Market Integration. Northwestern University (September). Photocopy.

Prakash, G., and A. M. Taylor. 1997. Measuring Market Integration: A Model of Arbitrage with an Econometric Application to the Gold Standard, 1880–1913. Working Paper Series no. 6073, National Bureau of Economic Research (June).

Ramey, V. A. 1989. Inventories as Factors of Production and Economic Fluctuations. *American Economic Review* 79 (June): 338–54.

Rogoff, K. 1996. The Purchasing Power Parity Puzzle. *Journal of Economic Literature* 34 (June): 647–668.

Rossana, R. J., and J. J. Seater. 1992. Aggregation, Unit Roots, and the Time-Series Structure of Manufacturing Real Wages. *International Economic Review* 33 (Feburary): 159–179.

Sarno, L., and M. P. Taylor. 1998. Real Interest Rates, Liquidity Constraints and Financial Deregulation: Private Consumption Behavior in the U.K. *Journal of Macroeconomics* 20 (Spring): 221–42.

Sercu, P., R. Uppal, and C. Van Hulle. 1995. The Exchange Rate in the Presence of Transaction Costs: Implications for Tests of Purchasing Power Parity. *Journal of Finance* 50 (September): 1309–1319.

Schwert, G. W. 1987. Effects of Model Specification on Tests for Unit Roots in Macroeconomic Data. *Journal of Monetary Economics* 20 (July): 73–103.

Shiller, R. J., and P. Perron. 1985. Testing the Random Walk Hypothesis: Power Versus Frequency of Observation. *Economics Letters* 18 : 381–86.

Snowden, K. A. 1990. Historical Returns and Security Market Development, 1872–1925. *Explorations in Economic History* 27 (October): 381–420.

Taylor, A. M. 1996. International Capital Mobility in History: Purchasing Power Parity in the Long-Run. Working Paper Series no. 5742, National Bureau of Economic Research (September).

Taylor, M. P. 1995. The Economics of Exchange Rates. *Journal of Economic Literature* 33 (March): 13–47.

Taylor, M. P., and D. A. Peel. 1999. Nonlinear Mean-Reversion in Real Exchange Rates: Towards a Solution to the Purchasing Power Parity Puzzle. University College, Oxford. Photocopy.

Washburn, C. L., and C. S. Binkley. 1990a. Informational Efficiency of Markets for Stumpage. *American Journal of Agricultural Economics* 72 (May): 394–405.

Washburn, C. L., and C. S. Binkley. 1990b. On the Use of Period-Average Stumpage Prices to Estimate Forest Asset Pricing Models. *Land Economy* 66 (November): 379–393.

Wei, S.-J., and D. C. Parsley. 1995. Purchasing Power *Dis*-Parity during the Floating Rate Period. Harvard University and Vanderbilt University (May). Photocopy.

Working, H. 1960. Note on the Correlation of First Differences of Averages in a Random Chain. *Econometrica* 28 (October): 916–918.

Wu, Y. 1996. Are Real Exchange Rates Nonstationary? Evidence from a Panel-Data Test. *Journal of Money, Credit, and Banking* 28 (February): 54–63.

Yeager, L. B. 1976. *International Monetary Relations: Theory, History, and Policy*. 2d ed. New York: Harper and Row.