

**RE-EXAMINING THE PURCHASING POWER PARITY HYPOTHESIS  
OVER TWO CENTURIES**

**John T. Cuddington and Hong Liang\***

December 31, 1997

Department of Economics  
Georgetown University  
580 Intercultural Center  
Washington, DC 20057-1036  
Ph: (202) 687-6260  
Fx: (202) 687-6102

*<http://www.georgetown.edu/deparments/economics>*

---

\* The authors thank Jim Lothian for providing the data set used in this study.

# Re-examining the Purchasing Parity Hypothesis Over Two Centuries

## Abstract

This paper reexamines the stationarity of the dollar-sterling real exchange rate using the two centuries of data analyzed in Lothian and Taylor (LT) (1996). We find univariate time series models for the real exchange rate that dominate both the stationary AR(1) specification chosen by LT and the random walk alternative they consider, both in terms of in-sample fit and out-of-sample forecasting ability. Our results suggest, contrary to LT, that long-run PPP does not hold for the dollar-sterling exchange rate. The differences in our conclusions arise from two primary sources: (1) choice of lag lengths in the augmented Dickey-Fuller regressions on which the unit root tests are based, and (2) realization that the absence of a unit root is necessary but not sufficient for stationarity. Deterministic time trends (and structural breaks) can also give rise to nonstationary RER series.

LT (1996) also compare the out-of-sample forecasting results of their stationary AR(1) model with the random walk model over the 1973-90 period. They find that the AR(1) forecasts outperform the random walk model. Moreover, the relative superiority of the AR(1) model increases monotonically as the forecast horizon extends from one year to five years. In light of the unit root test results, we find that the dollar-sterling real exchange rate is better modeled as either trend stationary or difference stationary with an MA(5) error process. Compared with these two non-stationary specifications, the AR(1) model no longer dominates in out-of-sample forecasts on all time horizons. When the out-of-sample forecasting exercise extends from the post-1973 years to the entire post-World War II period, the non-stationary models strictly dominate the stationary AR(1) specification. In the meantime, the forecast exercise also indicates the stability of the non-stationary models over the pre-float and recent floating rate period.

## **Introduction**

In the last ten to fifteen years, a large literature has emerged on testing the long-run validity of PPP, or equivalently the stationarity of the real exchange rate (RER), using modern time-series econometrics techniques. (See Rogoff (1996) for recent references.) Nonstationarity in the RER may take the form of a unit root process (with or without drift), a deterministic trend, and/or structural breaks. An important contribution by Lothian and Taylor (LT) (1996) emphasizes the potential role that low power in standard unit root tests has played in leading some authors to conclude that the RER follows a random walk and is therefore nonstationary. They present new unit root tests for the pound-sterling and dollar-sterling real exchange rates using annual time series covering two hundred years. With the increased test power obtained by this large data sample, they are able to reject the unit root hypothesis using both ADF and Phillips-Perron tests. They therefore conclude that PPP is valid in the long run for the two bilateral RERs considered.

The present study reexamines the dollar-sterling RER and concludes, contrary to LT, that it is not stationary.<sup>1</sup> We find univariate time series models for the real exchange rate that dominate both the stationary AR(1) specification chosen by LT and the random walk alternative they consider, both in terms of in-sample fit and out-of sample forecasting ability. Our results suggest, contrary to LT, that long-run PPP does not hold for the dollar-sterling exchange rate. The differences in our conclusions arise from two primary sources: (1) choice of lag lengths in the augmented Dickey-Fuller regressions on which the unit root tests are based, and (2) realization that the absence of a unit root is necessary but not sufficient for stationarity.

---

<sup>1</sup> We also examined the franc-sterling RER and agree with LT that it is stationary. These results are available

Deterministic time trends (and structural breaks) can also give rise to nonstationary RER series.

The remainder of the paper is organized as follows. Section II discusses the main issues relating to tests for unit roots and stationarity. Section III presents our unit root tests and conclusions of stationarity. In section IV, the relative superiority of different models in out-of-sample forecast is re-evaluated. Section V concludes.

## II. Unit Root Tests and Stationarity

Unit root testing is hazardous terrain. It is now well-known that unit root tests have low power, and that the form of the test (i.e. whether an intercept and time trend are included in the regression used to obtain the ADF or Phillips-Perron statistics is critical in interpreting the results). In general, the appropriate procedure is to use the general-to-specific (GTS) methodology by first estimating the regression:

$$(1) \Delta q_t = \alpha_0 + \alpha_1 t + \gamma q_{t-1} + \sum_{i=1}^p \beta_i \Delta q_{t-i} + \varepsilon_t$$

where  $q_t \equiv s_t + p_t - p_t^*$ .  $s_t$  is the logarithm of the nominal exchange rate and  $p_t$  and  $p_t^*$  are the domestic and foreign price level, respectively. Note that (1) includes both the intercept and time trend. The arguments in favor of this approach are the usual ones involving omitted variable bias versus loss of efficiency caused by redundant regressors. Moreover, a time trend must be included initially to allow for the possibility of a deterministic trend in the alternative hypothesis when the null hypothesis of a unit root is tested. (On this, see the excellent discussion in Hamilton (1995,...). If the null hypothesis of a unit root is not rejected in the most general

version of (1) including the intercept and trend, the significance of the trend and intercept can then be tested in turn to see if they can be omitted from (1), thereby increasing the power of the unit root tests<sup>2</sup>.

The GTS methodology is also useful in choosing the optimal lag length ( $p$ ) on the polynomial involving lagged values of the dependent variable in (1). (Recall that too many lags reduces efficiency, while too few implies serial correlation in the residuals thereby invalidating standard significance tests.) In fact, recent Monte Carlo evidence strongly recommends its use of other lag selection methods based on the Akaike or Schwartz criteria. On the latter, see Hall (1994) and Ng and Perron (1995). The GTS suggests starting with a "large" number of lags, with the square root of sample size being a good rule of thumb. Examine the  $t$ -statistic on the last lag (which is asymptotically normal). If it is insignificant, drop the last lag and rerun the test equation (1). Continue dropping lags until a significant lag (at, say, the 95% level) is found. Stop at that point, leaving all shorter lags in the regression.

Nonstationarity can take the form of either unit root, structural break, and/or deterministic trend. Therefore when testing the validity of long-run PPP it is important to allow for the possibility of a deterministic trend in the RER. If PPP holds, the trend should not be significant *and* the unit root hypothesis should be rejected. Lothian and Taylor (1996) focus on nonstationarity resulting from the presence of unit roots, but not deterministic trends. Their out-of-sample forecasting exercises examine the possibility of structural breaks.

---

<sup>2</sup> See Enders (1995, Table 4.1) for a detailed discussion of the GTS methodology and the appropriate critical values for testing the significance of the trend and intercept terms in the various specifications.

### **III. Empirical Results**

As in Lothian and Taylor (1996), this study considers both the ADF and the Phillips-Perron unit root tests. The latter do not require lagged values of the dependent variable in (1) to account for possible serial correlation.

#### *i The ADF Unit Root Test*

Implementing the GTS method, an initial lag length of  $p=15$  was chosen. Eliminating redundant lags one-by-one led to a chosen lag length of 14, well in excess of the 5 lag specification used in LT. With 14 lags of the dependent variable in (1), the unit root hypothesis can not be rejected. Subsequent tests on the trend and constant terms find that neither is significant. Hence, a more restricted specification without the two terms is tested against the null. The t-statistic on  $\gamma$  is  $-0.58$ , which is far below the critical value of  $-1.95$  to reject the unit root hypothesis.

Suppose that 14 lags is for some a priori reason considered to be “too many”.

For the full sample period (1791-1990) if the 14th lag is ignored by starting the GTS procedure with 13 lags, the chosen lag length would be eight. The ADF test results based on equation (1) now indicate rejection of the unit root hypothesis, which concurs with the results from the five-lag specification used by LT. **Note, however, that in the 8-lag (and the 5 lag) specification of (1), the time trend is now negative and statistically significant.** The point estimate is  $\hat{\rho}$ 0.00059, with an associated t-statistic of  $\hat{\rho}$ 3.47. Note that the t-statistics are asymptotically normal when the unit root hypothesis is rejected. Similar conclusions are found when the sample period ends in 1945. The unit root hypothesis can not be rejected, however, for the post-World War II period 1946-1990. The test results are presented in Table 1.

**Table 1. Unit Root Tests for Dollar-Sterling RER**

	<i>ADF Unit Root Test</i>			<i>Phillips Perron Unit Root Test</i>			
	$\tau$	Constant	time	$\tau$	constant	time	
1791-1990:				1791-1990:			
14 lags	-0.58			5 lags	-3.80*	4.26*	-2.62*
8 lags	-4.10*	4.08*	-3.47*	1791-1945:			
5 lags	-4.30*	4.29*	-3.53*	4 lags	-4.93*	4.52*	-2.55*
1791-1945:				1946-1990:			
14 lags	-2.80	2.82	-2.92*	4 lags	-3.15	2.87	2.93*
8 lags	-4.59*	4.61*	-3.85*				
5 lags	-4.85*	4.89*	-3.80*				
1946-1990:							
8 lags	-3.09	2.91	3.06*				
5 lags	-3.13	2.94	3.34*				

\* indicates significance at 5% level.

*ii. The Phillips-Perron Unit Root Test*

Phillips and Perron (1988) developed a generalization of the ADF test that allows for a weaker set of assumptions concerning the error process. To choose the number of periods of serial correlations to include, Newey-West (1987) suggest that it should equal to  $4(T/100)^{2/9}$ ,

which is 5 for the real exchange rate series used in this study. Results on equation (1) indicate rejection of the unit root hypothesis. Again, however, the coefficient on the time trend is statistically significant (as the t-statistics are asymptotically normal when the unit root hypothesis is rejected).

Although the Phillips-Perron test allows for weaker assumptions about the error process, Monte Carlo studies suggest that in the presence of *negative* moving average terms the test tends to reject the null of a unit root whether or not the true data generating process contains a negative unit root. (Enders, 1995; p242) LT have estimated ARIMA(1,0,1) models for the various sample sizes and concluded that the moving average terms is small and insignificant. However, the correlograms of the residuals from an AR(1) specification indicates that there is a spike at lag 5. Estimates of an ARIMA(1,0,5) model find a significant and negative moving average term at lag 5. The presence of this negative moving average term may have subjected the Phillips-Perron statistics to considerable distortions (Schwert, 1989).

### iii The Alternative Models

The findings on the unit root tests are mixed, but they all suggest the non-stationarity of the dollar-sterling real exchange rate over the past two hundred years. Thus, the AR(1) stationary model proposed by LT is misspecified. We propose two alternative non-stationary models as follows.

$$(2) q_t = a_0 + b * time + a_1 q_{t-1} + u_t + a_2 u_{t-5}$$

$$(3) \Delta q_t = \lambda_0 + e_t + \lambda_1 e_{t-5}$$

Equation (2) is a trend stationary model (TS) with AR(1) and MA(5) error process.

Equation (3) is a difference stationary model (DS) with MA(5) error process.

If  $b=a_0=0$ , the TS model collapses to the LT model. Hence it is possible to nest the LT specification in equation (2) and test the restrictions implied by their model. The likelihood ratio (LR) statistics are used to test the following hypotheses:

$$H_A: b=0; H_B: a_2=0; H_C: b=a_2=0$$

The LR statistics has an asymptotic  $\chi^2$  distribution with degrees of freedom equal to the number of restrictions. Table 2 reports the test results. It can be seen that all three hypotheses regarding the TS model can be rejected with very small  $p$ -value.

<b>Table 2. Estimated Non-stationary Models (1791-1990)</b>		
<i>The TS Model</i>		
$q_t = 1.771 - 0.002*time + e_t$		
(33.76) (-3.90)		
$e_t = 0.827e_{t-1} + u_t - 0.235 u_{t-5}$		
(18.83) (-3.11)		
$R^2 = 0.81; Q(39) = 35.06$		
$H_A: b=0$ LR=8.55 (p=0.0035)	$H_B: a_2=0$ LR=9.30 (p=0.0023)	$H_C: b=a_2=0$ LR=16.13 (p=0.0003)
<i>The DS Model</i>		
$\Delta q_t = -0.0004 + u_t - 0.252u_{t-5}$		
(-0.102) (-3.58)		
$R^2 = 0.06; Q(39) = 40.21$		
$H_O: \lambda_1=0$ LR=11.34 (p=0.0008)		

t-statistics are in the parenthesis unless otherwise stated.

In summary, the unit root test results reveal that the real exchange rate for dollar-sterling is non-stationary, rejecting the PPP hypothesis. Similar analysis for franc-sterling real exchange rate confirms the conclusions reached by LT.

#### IV. Out of Sample Forecasting

LT compare the forecasting ability of their stationary AR(1) specification to the simple random walk model preferred by earlier authors (e.g. Roll (1979), Adler and Lehamann (1983)).

Their model fitting to the pre-1973 period produces lower root mean square error (RMSE) when forecasting the real exchange rate over the floating rate period 1974-1990. In addition, they find that the relative superiority of the AR(1) model increases monotonically as the forecast horizon is extended.

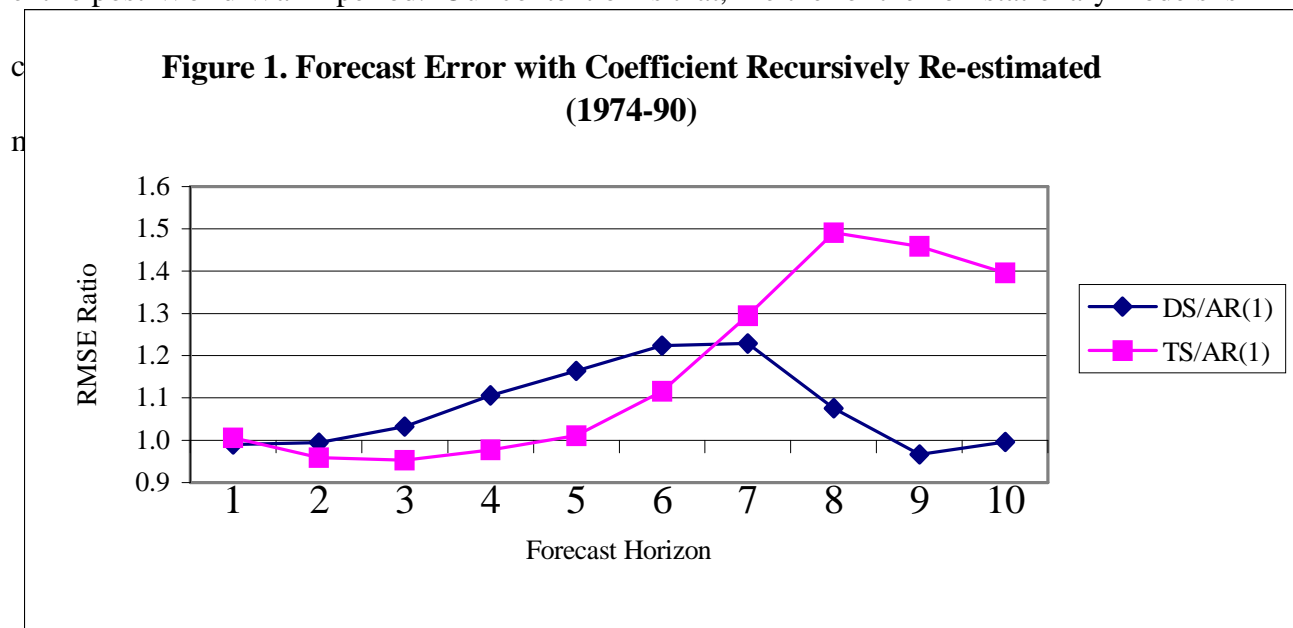
The empirical results presented so far demonstrate that the alternative non-stationary models fit the sample better than the AR(1) specification. To compare these models with the AR(1) model in out-of-sample forecasting, we first repeat the LT exercise but using a longer forecasting horizon of 10 years. We thus constructed four series of 1-to-10-year-ahead dynamic forecasts for each year after 1973. The first two series are obtained by using the TS and the DS models with the coefficients held fixed at their pre-1973 values respectively, while the other two series by using recursively re-estimated coefficients. The ratios of RMSEs were then constructed against the yardstick AR(1) model. Hence, a ratio lower than one indicates better performance of the non-stationary models over the AR(1) specification in out-of-sample forecasting<sup>1</sup> When trying to replicate the results of LT, we find that the LT=s fixed-coefficient RMSEs were obtained by holding the coefficients fixed at their whole sample period value (i.e. at the estimates using data through 1990), instead of at their pre-floating (through 1973) value as they claimed in the paper..

The results of this exercise are reported in Figure 1 and 2. The line graphs show how the relative superiority of alternative models changes as the forecasting horizon extends. The first point to notice is that the AR(1) model no longer monotonically dominates in out-of-sample forecasts. In the short run (1-to-5-year horizon), the TS model performs the best. More interestingly, the DS model bounces back and outperforms the AR(1) model when the time horizon extends to nine years. Although the AR(1) model no longer monotonically dominates, in

many cases it seems to produce better forecasts. Of course, the imposition of an incorrect restriction (i.e., stationarity by the LT model) may lead to superior forecasting performance. Next, we want to check this conjecture by looking at different sample period for forecasting.

Second, there is little difference in the fixed coefficient and recursive-estimation results. For the TS model, the fixed-coefficient RMSEs are slightly larger than the recursive-estimation RMSEs; for the DS model, the relative superiority of different forecast methods changes with forecast horizon, as well as is very marginal. The little difference in RMSEs between the two forecasting methods may be taken as an indication of the stability of these specifications over the period.

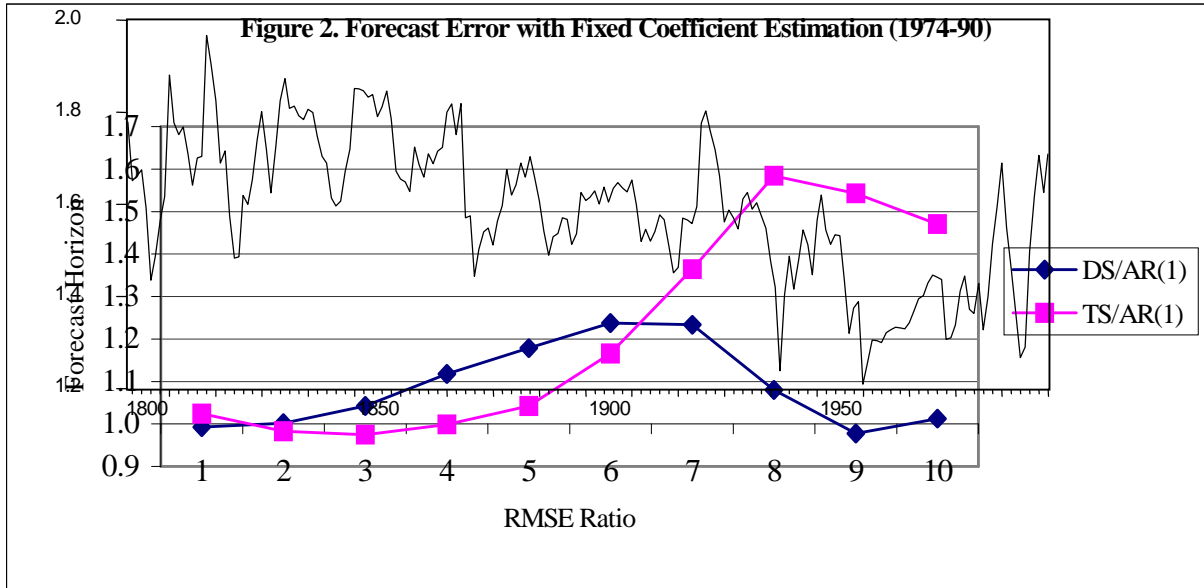
Looking at the dollar-sterling real exchange rate in Figure 3, we notice that the out-of-sample forecasting exercise by LT corresponds to the period when there appeared to be some mean-reverting behavior of the rate. There is a possibility that this period contains too few observations (and/or too many outliers) that it artificially influences the power of alternative models in out-of-sample forecasting. So, we repeat the out-of-sample forecasting tests for the entire post-World War II period. Our contention is that, if either of the non-stationary models is



RMSE Ratios of Fixed Coefficient Forecasts over Re-estimated Coefficient Forecasts

Horizon (years)	1	2	3	4	5	6	7	8	9	10
DS model	0.998	1.002	1.003	1.003	1.005	1.005	1.003	0.998	0.995	0.994
TS Model	1.014	1.017	1.013	1.010	1.018	1.034	1.056	1.083	1.076	1.061

Figure 3. The Dollar-Sterling Real Exchange Rate: 1791-1990



The results of the out-of-sample forecasting exercise for the period 1946-90 are presented in Table 3. Both non-stationary models strictly dominate the stationary AR(1) model at every forecasting horizon, although by only a small margin. We regard this as additional evidence that the non-stationary models characterize the dollar-sterling real exchange rate behavior better than

the stationary model proposed by LT.

RMSE Ratio (Coefficients Held Fixed)										
Horizon	1	2	3	4	5	6	7	8	9	10
TS/AR(1)	0.9999	0.9997	0.9996	0.9995	0.9990	0.9989	0.9990	0.9991	0.9993	0.9996
DS/AR(1)	0.9993	0.9989	0.9983	0.9981	0.9980	0.9981	0.9984	0.9988	0.9992	0.9997
RMSE Ratio (Coefficients Re-estimated)										
TS/AR(1)	0.9999	0.9997	0.9996	0.9995	0.9990	0.9989	0.9990	0.9991	0.9993	0.9996
DS/AR(1)	0.9996	0.9993	0.9988	0.9986	0.9984	0.9984	0.9986	0.9989	0.9993	0.9997

## V. Conclusions

The findings of this paper suggest that the sterling-dollar real exchange rate is non-stationary over the past two centuries. Hence, PPP does not hold even in the long run. These findings contradict the conclusions reached by Lothian and Taylor (1996). In a series of dynamic out-of-sample forecasts, we also reject the dominance of LT's stationary model.

Theory, of course, provides reasons why PPP may not hold in the long run. There are a number of factors that can cause changes in the long-run real exchange rate and hence temporary or permanent deviations from PPP, such as productivity changes, a natural resource discovery, changes in consumers' preferences and so on. It is, however, difficult to see why this would lead to a *deterministic* time trend. We prefer the *stochastic* trend model chosen by the GTS method using longer lag-lengths. The long lag length is consistent with the puzzle of very persistent time series behavior of real exchange rates discussed in Rogoff (1996).

An intriguing question from our reexamination of PPP using the LT data is why it apparently holds for the franc-sterling but not the dollar-sterling rate. There are several possible explanations. One is that the geographic distance is greater between the US and Europe than between the UK and France, but that effective distance has shrunk over time due to improvements in transportation and communications technology. Another more interesting

explanation may be that fixed nominal exchange rate prevailed longer between the UK and France than between the US and Europe. A full explanation must await future research.

## References

- Adler, M. and B. Lehmann. 1983. "Deviations from Purchasing Power Parity in the Long Run," *Journal of Finance*, 38, 1471-87.
- Enders, W. 1995. Applied Econometric Time Series, John Wiley & Sons, Inc.
- Hall, A. 1994. "Testing for a Unit Root in Time Series with Pretest Data-Based Model Selection," *Journal of Business and Economic Statistics* 12, 4, 461-470.
- Hamilton, J. 1994. Time Series.... Princeton University Press.
- 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, 3, 488-509.
- Newey, W.K. and K.D. West. 1987. "A Simple Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica* 55, 703-8.
- Ng, S. and P. Perron. 1995. "Unit Root Tests in ARMA Models with Data-Dependent Methods for the Selection of the Truncation Lag," *Journal of Business and Economic statistics* 90, 429, 268-281.
- Rogoff, K. 1996. "The Purchasing Power Parity Puzzle," *Journal of Economic Literature* 34, 2, 647-6
- Roll, R. 1979. "Violations of Purchasing Power Parity and Their Implications for Efficient International Commodity Markets," in M. Sarnat and G.P. Szego eds. *International Finance and Trade*, Vol 1. Cambridge, Mass.: Ballinger.
- Schwert, G.W. 1989. "Tests for Unit Roots: A Monte Carlo Investigation," *Journal of Business and Economic Statistics* 7, 147-59.