7/12/00: A revised, shortened version of this paper is forthcoming in the *Journal of International Money and Finance* (2000)

RE-EXAMINING THE PURCHASING POWER PARITY HYPOTHESIS OVER TWO CENTURIES

John T. Cuddington and Hong Liang*

May 6, 1998

Department of Economics
Georgetown University
580 Intercultural Center
Washington, DC 20057-1036
Ph: (202) 687-5830
Fax: (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 that the dollar-sterling real exchange rate is well modeled as either trend stationary process or as a difference stationary process with an MA(5) error. In terms of goodness of fit, these specifications dominate the stationary AR(1) specification chosen by LT as well as the random walk alternative they consider. The differences in our conclusions arise from two primary sources: (1) choice of lag length in the augmented Dickey-Fuller regressions on which the unit root tests are based, and/or (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. The out-of-sample forecasting ability of the various models are compared.
Our results suggest that long-run PPP does ***not*** hold for the dollar-sterling exchange rate.

I. 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.) 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 test results for the franc-sterling and dollar-sterling real exchange rates using annual time series spanning two centuries. 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.

Although we agree with LT that the franc-sterling exchange rate is stationary, our re-examination of the dollar-sterling RER concludes that it is not stationary. Our skepticism about their findings for the dollar-sterling rate initially arose when we found that the time trend coefficient in their Dickey-Fuller and Phillips-Perron regressions was highly significant. This is important because the absence of a unit root, is necessary but not sufficient for mean stationarity. Nonstationarities may also take the form of deterministic trends or structural breaks. Given the implausibility of deterministic as opposed to stochastic trends in RERS, and the potential difficulty of distinguishing them, this led us to re-examine the specification of LT's unit root test, in particular their choice of lag length in the ADF tests. The differences in our conclusions arise from two primary sources: (1) choice of lag length in the augmented Dickey-Fuller regressions on which the unit root tests are based, and (2) realization that time trends -- stochastic or deterministic -- imply nonstationarity, and hence rejection of the PPP hypothesis.

Analyzing the underlying causes of the often-observed non-stationarity of real exchange rates continues to be an active research area. Most explanations focus on the distinction between tradeable and nontradeable goods, differential rates of productivity growth, and imperfect competition in the world market. Our conclusion that PPP holds for the long-run franc-sterling rate but not for the dollar-sterling rate suggests the potential importance of geographic factors in influencing the adjustment of relative prices[1].

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 test results and estimates univariate models for the RER. In section IV, the relative superiority of different models in out-of-sample forecasting is 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 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 following regression which includes both an intercept and time trend:

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

where $q_t \equiv s_t + p_t^* - p_t$ is the logarithm of the RER. s_t is the logarithm of the nominal exchange rate and p_t and p_t^* are the domestic and foreign price level, respectively. The arguments in favor of beginning with the most general specification including the intercept and time trend are the usual ones involving omitted variable bias versus loss of efficiency caused by redundant regressors; 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 (1994, 501-2)). If the null hypothesis of a unit root is not rejected using (1), then the significance of the trend and intercept can then be tested in turn to see if they can be omitted from the test equation, thereby increasing the power of the unit root tests[2].

The GTS methodology can also be used in choosing the optimal lag length (p) on the lag polynomial of the dependent variable in (1)[3]. It dominates Akaike and Shwartz criteria in recent Monte Carlo studies. Ng and Perron (1995) compare two classes of lag length selection methods that are widely used in practice. The "deterministic" rule refers to any *a priori* rule that presets the value of p in equation (1). Including in this class is the Newey-West rule (used by LT), which sets p as a fixed function of T. The second class of lag selection methods is "adaptive" or "data-dependent" rules. Ng and Perron (1995) demonstrate that any deterministic rule is likely to result in size distortion and/or power loss unless p happens to be chosen appropriately. One important conclusion reached in their study is that:

> [An] overly parsimonious model can have large size distortions, but an over parameterized model may have low power. However, the size problem is more severe than power loss in the sense that discrepancies in power across selection procedures diminishes as T increases, but size

1 See Engel and Rogers (1995) for a study on the relationship between the law of one price and the roles of geography and currencies.

2 See Enders (1995, Chapter 4) 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 (in his Table 4.1).

3 Recall that too many lags reduce efficiency, while too few imply serial correlation in the residuals thereby invalidating standard significance tests.

> distortions persist even for large sample sizes for some methods of selecting [*p*]. (p.277)

Within the class of data dependent rules, Ng and Perron (1995) recommend the GTS method that sequentially tests for the significance of the coefficients on lags of the dependent variable. (See also Hall (1994)). The GTS method involves 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 re-estimate the test equation (1). Continue dropping the last lag in the lag polynomial, one-by-one, until a significant lag (at, say, the 95% level) is found. Stop at that point, leaving all shorter lags in the regression.

III. Empirical Results

This section consists two parts. Section III.1 presents the unit root test results on the dollar-sterling RER series. Based on these results, two alternative non-stationary models are proposed in Section III.2. Their appropriateness is evaluated in terms of in-sample fit, before turning to out-of-sample forecasting in Section IV.

III.1. The Unit Root Test 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 of (1) without these two terms is used in obtaining the ADF statistic. The t-statistic on γ is -0.58, which is far below the Dickey-Fuller-MacKinnon critical value of -1.95. Thus we fail to reject the unit root hypothesis. This result is shown in the first row of summary Table 1.

Table 1. Unit Root Tests for Dollar-Sterling RER

	ADF Unit Root Test			*Phillips Perron Unit Root Test*			
	τ	Constant	time		τ	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.

Suppose that 14 lags are for some *a priori* reason considered being “too many.” If the 14th lag is ignored by starting the GTS procedure with 13 lags, the chosen lag length is eight. The ADF test results based on equation (1) with eight lags 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 -0.00059, with an associated t-statistic of -3.47[4]. see column 3, row 2 (8 lags) in the Table. Similar

4 Note that the t-statistic on the time trend is asymptotically normal when the unit root hypothesis is rejected.

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.

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 maximum order of serial correlations to consider, Newey-West (1987) use $4(T/100)^{2/9}$, which is roughly 5 for the real exchange rate series used in this study. The Phillips-Perron unit root tests, reported in Table 1, indicate rejection of the unit root hypothesis as in LT. We find, however, that the coefficient on the time trend is again statistically significant, as it is in the ADF tests.   Thus, the conclusion is that the RER is not mean stationary.

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 PP test tends to reject the null of a unit root whether or not the true data generating process contains a negative unit root." (Enders (1995,242) LT estimated ARIMA(1,0,1) models for the various sample sizes and concluded that the MA(1) term was 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.2 The Alternative Models

The findings on the unit root tests are mixed, but all of the specifications in Table 1 suggest the non-stationarity of the dollar-sterling real exchange rate over the past two hundred years, either because of the presence of a unit root or a deterministic time trend. Based on these results, we propose two alternative nonstationary models:

(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 an ARMA(1,5) error process. Equation (3) is a difference stationary model (DS) with MA(5) error process.

It is possible to nest the LT specification in equation (2). It is the special case where $b=a_2=0$ and $a_1<1$. 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 χ^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*-values.

Table 2. Estimated Non-stationary Models (1791-1990)		
	The TS Model	
	$q_t = 1.771 - 0.002*time + e_t$	
	(33.76) (-3.90)	
	$e_t = 0.827 e_{t-1} + u_t \quad 0.235 u_{t-5}$	
	(18.83) (-3.11)	
	$R^2 = 0.81$; $Q(39)=35.06$	
H_A: $b=0$	H_B: $a_2=0$	H_C: $b=a_2=0$
LR=8.55 (p=0.0035)	LR=9.30 (p=0.0023)	LR=16.13 (p=0.0003)
	The DS Model	
	$\Delta q_t = -0.0004 + u_t \quad 0.252 u_{t-5}$	
	(-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 point to deterministic or stochastic trends, depending on one's judgement on lag lengths in the ADF regressions. In any event, estimated TS and DS models for the respective interpretations both imply that the real dollar-sterling exchange rate is nonstationary. The PPP hypothesis is rejected.

IV. Out of Sample Forecasting

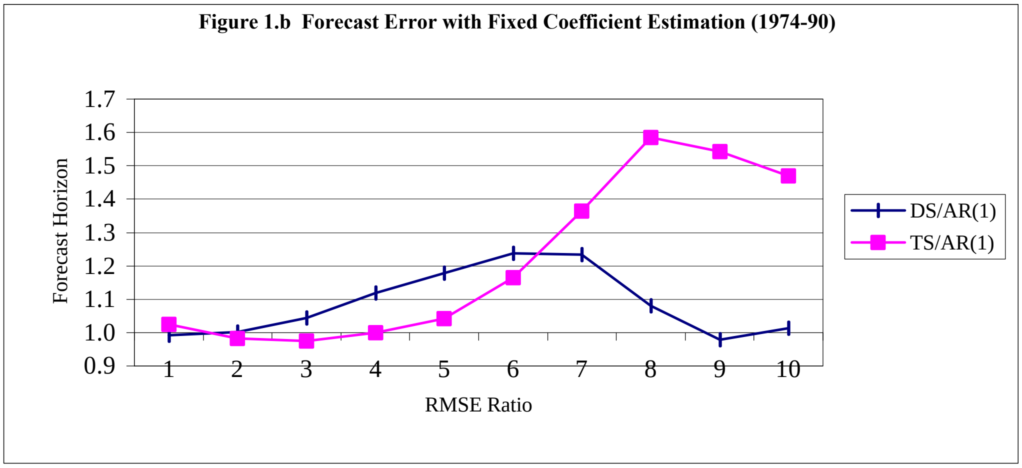
LT compare the forecasting ability of their stationary AR(1) specification to the simple random walk model preferred by earlier authors (e.g. Roll (1979), Adler and Lehamann (1983)). Their model fitted to the pre-1973 period produces lower root mean square error (RMSE) when forecasting the real exchange rate out-of-sample into 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 (i.e., 2-year ahead vs. 1-year ahead forecasts).

The empirical results presented in Section III demonstrate that the alternative nonstationary models fit the sample better than the stationary AR(1) specification. To compare these models with the AR(1) model in out-of-sample forecasting, we first repeat the LT exercise but considering forecasting horizons of up to 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 our nonstationary models over the AR(1) specification in out-of-sample forecasting[5].

The results of this exercise are reported in Figure 1 and Table 3. First, the line graphs show how the relative superiority of alternative models changes as longer forecasting horizons are considered. 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 because of the specific period selected for forecasting. Later, we want to check this conjecture by looking at different sample period for forecasting.

Second, Table 3 shows that 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, but is very marginal. The minor differences in RMSEs between the two forecasting methods may be taken as an indication of the stability of these specifications over the period.

5 When trying to replicate the results of LT, we found that their fixed-coefficient RMSEs were obtained by holding the coefficients fixed at their estimated value when the entire sample through 1990 was chosen, instead of at their pre-floating (through 1973) values, as they claimed in the paper.



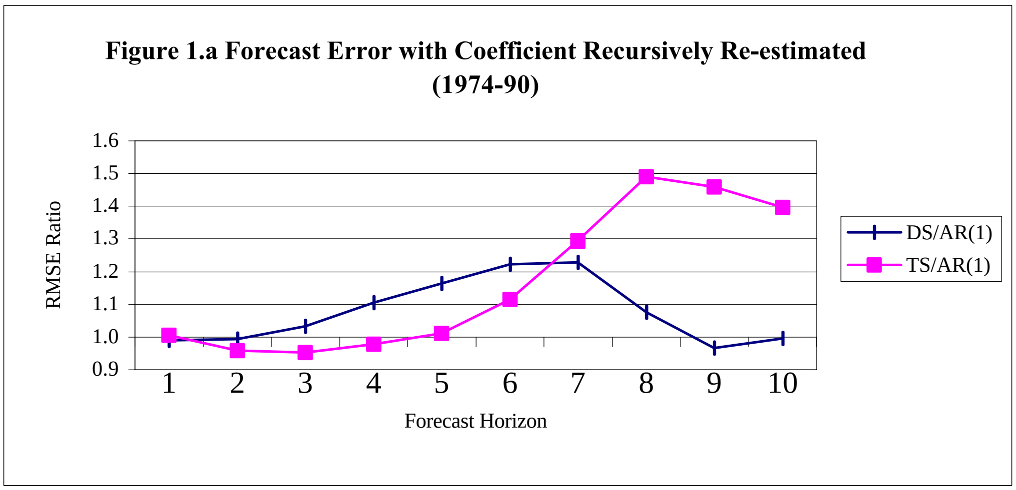
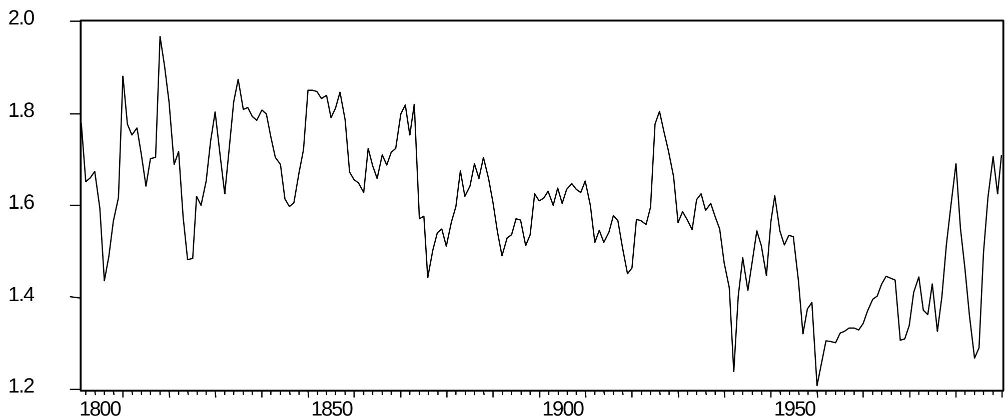


Table 3. RMSE Ratios of Fixed Coefficient Forecasts over Re-estimated Coefficient Forecasts

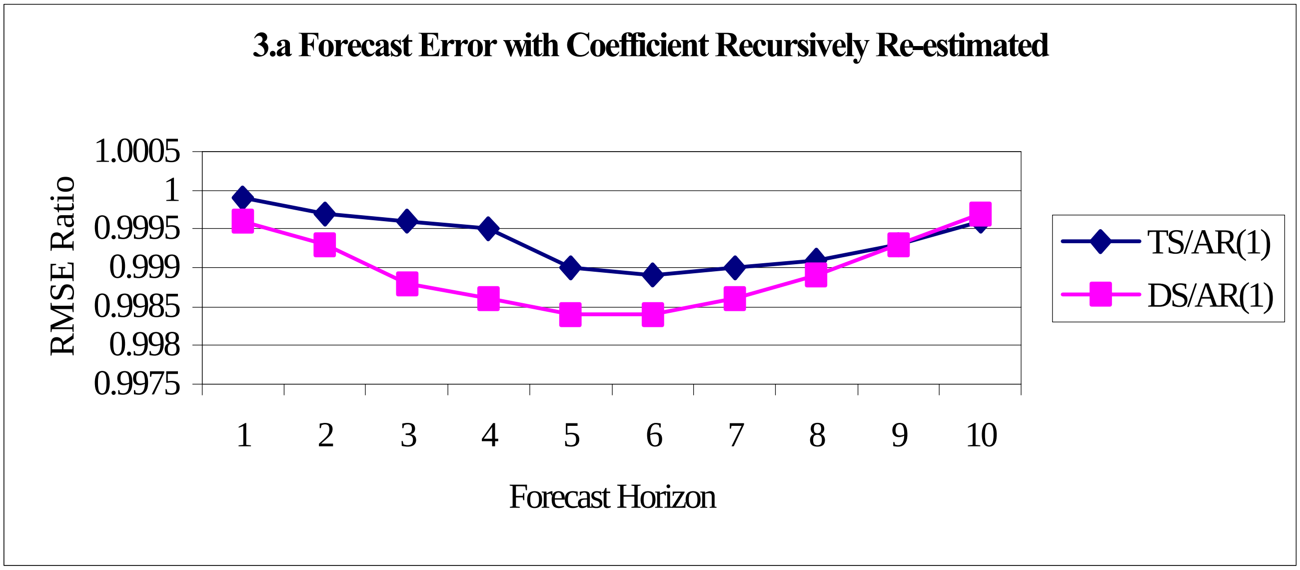
Horizon (years)	1	2	3	4	5	6	7	8	9	10
DS model	0.998	1.002	1.003	1.003	1.005	1.005	1.003	0.998	0.995	0.994
TS Model	1.014	1.017	1.013	1.010	1.018	1.034	1.056	1.083	1.076	1.061

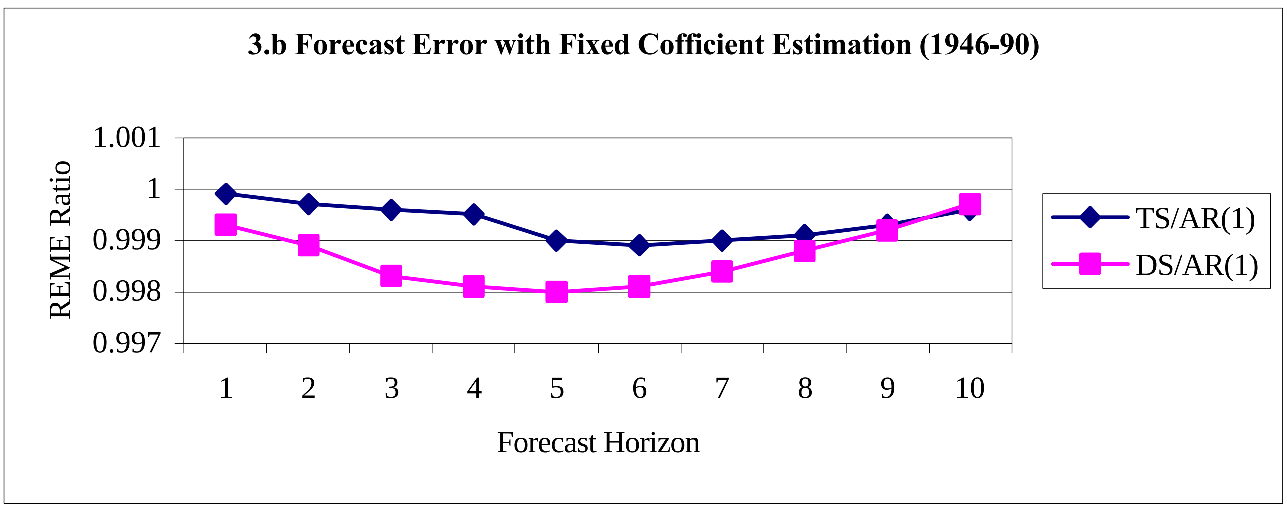
Looking at the dollar-sterling real exchange rate in Figure 2, we notice that the out-of-sample forecasting exercise by LT corresponds to the period when there appeared to be some mean-reverting behavior of the rate. There is a possibility that this period is somewhat unrepresentative. So, we repeat the out-of-sample forecasting tests for the entire post-World War II period. Our contention is that, if either of our nonstationary specifications is closer to the true data generating process, it should outperform the stationary AR(1) model when more observations are included.

Figure 2. 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 Figure 3. Both nonstationary 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 nonstationary models characterize the dollar-sterling real exchange rate behavior better than the stationary model proposed by LT.





V. <u>Conclusions</u>

The findings of this paper suggest that the sterling-dollar real exchange rate is nonstationary 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 factors, such as productivity changes, a natural resource discovery, changes in consumers' preferences and so on, that can cause changes in the long-run real exchange rate and hence temporary or permanent deviations from PPP. Of the many models that try to explain deviations from PPP, one well-known hypothesis advanced by Balassa (1964) and Samuelson (1964) emphasizes the role of non-traded goods and different productivity growth between fast growing and slowing countries. Empirical evidence, however, also suggests that there are persistent deviations from the law of one price in traded goods[6]. In addition, the short-run size and direction of these deviations appear to be closely related to nominal exchange rate movements[7]. One possible explanation points to the existence of market imperfection and segmentation. It is, however, difficult to see why these factors would lead to a *deterministic* time trend in the RER. Hence, we prefer the *stochastic* trend or DS model. This is the specification that is suggested when longer lag-lengths indicated by the GTS method are chosen. The long lag length is consistent with Rogoff's (1996) observation that there an important puzzle to resolve explaining the very persistent time series behavior of real exchange rates.

There is an intriguing question from our reexamination of PPP using the LT data. Why does PPP apparently hold for the franc-sterling but not the dollar-sterling exchange 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 possibility is that nominal exchange rate variability was lower

[6] See Knetter (1989), Feenstra and Kendall (1994), and Ghosh and Wolf (1994).
[7] See Giovannini (1988), and Engel (1993).

between the UK and France than between the US and Europe. Interestingly, Engel and Rogers (1995) find that the law of one price holds more nearly for country pairs that are within geographic regions. Theories on pricing to market and nominal price stickiness have offered many interesting ideas as to why distance and nominal exchange variability may matter for long-run RER movements[8]. A full explanation, however, still awaits further research.

[8] See Goldberg and Knetter 91997) for an excellent survey on recent literature on goods prices and the nominal exchange rates.

References

Adler, M. and B. Lehmann. 1983. "Deviations from Purchasing Power Parity in the Long Run," *Journal of Finance*, 38, 1471-87.

Balassa, B. 1964. "The Purchasing Power Parity Doctrine: A Reappraisal," *Journal of Political Economy*, 72, 584-596.

Enders, W. 1995. *Applied Econometric Time Series*, John Wiley & Sons, Inc.

Engel, C. 1993. "Is Real Exchange Rate Variability Caused by Relative Price Changes? An Empirical Investigation," *Journal of Monetary Economics,* 32, 35-50.

Engel, C. and J.H. Rogers. 1995. "Regional Patterns in the Law of One Price: the Roles of Geography vs. Currencies," *NBER Working Paper*, No. 5395.

Feenstra, R. and J. Kendall. 1994. "Pass-through of Exchange Rates and Purchasing Power Parity," *NBER Working Paper*, No. 4842.

Ghosh, A. and H. Wolf. 1994. "Pricing in International Markets: Lessons from the Economist," *NBER Working Paper*, No. 4806.

Giovannini, A. 1988. "Exchange Rates and Traded Goods Prices," *Journal of International Economics*, 24, 45-68.

Goldberg, P.K. and M.M. Knetter. 1997. "Goods Prices and Exchange Rates: What Have We Learned?" *Journal of Economic Literature,* XXXV, 1243-1272.

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 Analysis***, Princeton University Press.

Knetter, M. 1989. "Price Discrimination by U.S. and German Exporters," *American Economic Review*, 79, 198-210.

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 American Statistical Association* 90, 429, 268-281.

Rogoff, K. 1996. “The Purchasing Power Parity Puzzle,” *Journal of Economic Literature* 34, 2, 647-668.

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.

Samuelson, P.A. 1964. “Theoretical Notes on Trade Problems,” *Review of Economics and Statistics*, 46, 145-164.

Schwert, G.W. 1989. “Tests for Unit Roots: A Monte Carlo Investigation,” *Journal of Business and Economic Statistics* 7, 147-59.