THE JOURNAL OF FINANCE • VOL. XLV, NO. 1 • MARCH 1990

Purchasing Power Parity in the Long Run

NISO ABUAF and PHILIPPE JORION *

ABSTRACT

This paper re-examines the evidence on Purchasing Power Parity (PPP) in the long run. Previous studies have generally been unable to reject the hypothesis that the real exchange rate follows a random walk. If true, this implies that PPP does not hold. In contrast, this paper casts serious doubt on this random walk hypothesis. The results follow from more powerful estimation techniques, applied in a multilateral framework. Deviations from PPP, while substantial in the short run, appear to take about three years to be reduced in half.

LONG-RUN PURCHASING POWER PARITY (PPP) is a fundamental building block of most models of exchange rate determination. Dynamic exchange rate models, as in Dornbusch (1976) and Mussa (1982), usually rely on PPP as a long-run equilibrium condition for the exchange rate. Yet the PPP doctrine has not fared well in recent tests.[1] In particular, Roll (1979) and Adler and Lehmann (1983) have been unable to reject the hypothesis that the real exchange rate follows a random walk. If true, the random walk hypothesis has the disturbing implication that shocks to the real exchange rate are never reversed, which clearly implies that there is no tendency for PPP to hold in the long run. This paper presents evidence which casts doubt on the random walk hypothesis for the real exchange rate.

In our opinion, the negative results obtained in previous empirical research[2] reflect the poor power of the tests rather than evidence against PPP. In other words, the methodology employed so far will fail to reject the random walk assumption even in situations where the real exchange rate exhibits slow reversals to PPP values. This is why Hakkio (1986) concludes that, "although the hypothesis that the exchange rate follows a random walk cannot be rejected, not much weight should be put on this conclusion."

This paper shows that PPP may hold in the long run after all. The stronger conclusions of this study can be traced to the use of more powerful tests, primarily the statistics advocated by Dickey and Fuller (1979), employed in a multivariate setting. We find that using a system of univariate autoregressions constraining

* Salomon Brothers and Graduate School of Business, Columbia University, respectively. This material is related to Abuaf's Ph.D. dissertation at the University of Chicago. The suggestions of the referees led to numerous improvements in the paper. The views expressed here do not necessarily reflect those of Salomon Brothers.

[1] See for instance Frenkel (1981), Cumby and Obstfeld (1984), and Hakkio (1984).

[2] Exceptions are Cumby and Obstfeld (1984) and Cumby and Huizinga (1988), who report that expected exchange rate changes are biased predictors of relative inflation rates. This implies that real exchange rates changes are somewhat predictable. Huizinga (1987) also finds evidence of mean reversion, although not statistically significant.

the autoregressive coefficient to be the same across countries leads to more precise parameter estimates than the usual country-by-country setting. The small sample distributions of tests on the slope coefficient are derived from Monte Carlo simulations, which also demonstrate the increased power of these tests. In addition, we analyze annual data over the period 1900–1972. The annual data yield clear evidence of mean reversion, much more so than the 1973–1987 monthly data set, which spans the recent flexible exchange rate period.

Our results suggest that, though the real exchange rate is well captured by a first-order autoregressive process, the root of this process is slightly below unity. If so, then long-run PPP holds. Nonetheless, there are substantial short-term deviations from PPP, which take on average three years to be reduced in half.

This paper is organized as follows. Section I presents a first-order autoregressive model of the real exchange rate. This model is consistent with rational expectations and will be used as the alternative hypothesis against the random walk assumption. Econometric issues are reviewed in Section II, where we demonstrate the superiority of a multilateral test framework. The empirical results are presented in Section III, and the last section concludes the paper.

I. An Autoregressive Model for Real Exchange Rates

The real exchange rate is defined as the nominal exchange rate deflated by a ratio of domestic to foreign price levels. Taking logarithms,

$$e \equiv s + p^* - p = \ln(SP^*/P), \tag{1}$$

where e is the log of the real exchange rate, s is the log of the spot exchange rate defined in local currency units per foreign currency unit, and p and p^* are the log of the price levels, with asterisks denoting foreign quantities.

Under *long-run* PPP,[3] the long-run equilibrium real exchange rate is a time-invariant constant, which equals one for the absolute version of PPP. In contrast, *short-run* PPP is violated whenever the instantaneous real exchange rate does not equal its long-run equilibrium value. Short-term deviations from PPP are commonly observed, and the real question is whether these deviations tend toward zero when economic forces such as commodity arbitrage or capital movements are allowed to take full effect.

For instance, suppose that e follows a first-order autoregressive process:

$$e_{t+1} = k_0 + k_1 e_t + u_{1t+1}, \tag{2}$$

where k_0 and k_1 are constants and the error term u_1 is normally and independently distributed over time. The log of the long-term equilibrium real exchange rate, $\bar{e}$ can be defined as the unconditional expectation of the process in (2). Assuming that $|k_1| < 1$, it can be written as

$$\bar{e} = k_0/(1 - k_1). \tag{3}$$

Long-run PPP is violated if $|k_1| \geq 1$ and if k_0 and/or k_1 are not time-invariant constants. Provided that long-run PPP holds, short-run PPP is violated whenever

[3] See, for instance, Frenkel (1976), (1978), (1981), and Officer (1976).

e_t does not equal its long-run value $\bar{e}$.[4] If $k_1 < 1$, however, shocks to the system are corrected at the rate of $(1 - k_1)$ per period. For monthly data, for instance, a value of k_1 of 0.98 implies that a given deviation from equilibrium would take 34 months to be reduced in half.[5] For annual data, k_1 equal to 0.8 implies a half-life of 3.1 years.

It is important to note that the previous specification is perfectly consistent with efficiently functioning capital markets. For instance, under rational expectations and risk-averse speculation in the foreign exchange market, it can be shown that movements in the real exchange rate can be described by

$$\Delta e_{t+1} \equiv e_{t+1} - e_t = E_t[r_{t+1}] - E_t[r^*_{t+1}] - d_t + u_{2t+1}, \tag{4}$$

where $E_t[r_{t+1}]$ is the expectation of the real interest rate from time t to $t + 1$, based on information available at time t, and d_t is the risk premium in the forward rate, defined as the difference between the forward rate and the expected future spot rate. Equation (4) shows that the serial correlation properties of Δe depend on the time-series characteristics of expected real return differentials and of the risk premium. Consequently, the martingale model for Δe breaks down when real interest differentials are mean-reverting, or when risk premia embodied in forward rates are time-varying.[6]

Alternatively, Stulz (1987) derives a general equilibrium model where movements in real exchange rates depend on the time-series properties of output shocks for nontraded goods. In his model, the real exchange rate follows a random walk only in the absence of serial correlation in the output of nontraded goods. Thus, there is no a priori reason to believe that changes in the real exchange rate have to be uncorrelated over time—even in the context of efficient financial speculation.

II. The Econometric Issues

Roll (1979) and Adler and Lehmann (1983) analyze the empirical validity of long-run PPP by testing whether the real exchange rate has a unit root. Both studies include a univariate regression model based on *differences* in the real exchange rate:

$$\Delta e_{t+1} = b_0 + b_1 \Delta e_t + \cdots + b_k \Delta e_{t-k+1} + u_{t+1}, \tag{5}$$

where k is the number of lags. The random walk hypothesis implies that the b_j's

[4] Note that the intercept k_0 and thus $\bar{e}$, may be different from zero for two reasons. First, price levels are only reported as indices, as opposed to a currency price of a basket of goods. As a result of this normalization, even if absolute PPP held, this measure of the real exchange rate may be different from one, and thus $\bar{e}$ from zero. Second, a non-zero intercept may be consistent with the relative version of long-run PPP, as opposed to the absolute version. Relative PPP may be verified when the factors causing deviations from absolute PPP are stable over time.

[5] If a and b are the initial and final percentage deviations from equilibrium, respectively, the number of intervals from a to b is given by $n = (\ln b - \ln a)/\ln k_1$. Thus the half-life is $\ln (1/2)/(\ln k_1)$. In practice, since k_1 is estimated from the data, n will be measured with error.

[6] For empirical evidence on the inequality of ex ante real interest rates across countries, see Mishkin (1984). The empirical evidence on time varying risk premia is reviewed in Korajczyk (1985).

coefficients should be zero for all j. These studies are unable to reject the null hypothesis,[7] which is seen to imply that PPP will be violated in the long run. We argue that this result should not be taken at face value, because the failure to reject the random walk model could be due to the low power of the tests used in these studies.

In fact, as Dickey and Fuller (1979) have shown, regressions in first differences have little power against the alternative of a stable near random walk model. In contrast, regressions in *levels* are likely to yield more powerful tests. This paper uses a first-order autoregression in levels with $k_1 < 1$ as the alternative hypothesis. Thus we choose to employ one-sided tests only.

One drawback of this specification in levels is that the usual test statistics do not apply. Kendall (1973), for instance, shows that the OLS estimate $\hat{k}_1$, although consistent, is centered at values less than one in finite samples when $k_1 = 1$. Generally, the downward bias is of the order of $-(1 + 3k_1)/T$, where T is the sample size.[8] Fuller (1976) tabulates new tests under the null $k_0 = 0$ and $k_1 = 1$. The first statistic is defined as:

$$\hat{\rho}_\mu = T(\hat{k}_1 - 1), \tag{6}$$

where k_1 is the OLS estimate of k_1 in (2), and the subscript μ indicates that the regression contains an intercept. The second test statistic is analogous to the "regression t-statistic":

$$\hat{\tau}_\mu = (\hat{k}_1 - 1)/\sigma(\hat{k}_1), \tag{7}$$

where $\sigma(k_1)$ is the OLS standard error of k_1. Given a significance level, the (negative) critical value of $\hat{\tau}_\mu$ is lower than that of the t distribution, since k_1 is downward biased in finite samples. For instance, for a one-sided test based on 180 observations, the 5% critical value of $\hat{\tau}_\mu$ is -2.88, against -1.65 for the t distribution. Dickey and Fuller (1979), (1981) indicate that the ρ_μ test is more powerful than the τ_μ test, and also generally superior to likelihood ratio tests and Box-Pierce statistics.

While these tests improve on the traditional test, they still have low power to discriminate between k_1 equal to one and slightly less than one. As seen in the power calculations of Dickey and Fuller (1981), the probability of rejecting the null at the 5% level based on $\hat{\rho}_\mu$ and $\hat{\tau}_\mu$ is only 19% and 12%, respectively, when the true value of the autoregression coefficient is 0.95 and the sample size is 100. This is a discouraging situation, since it shows that even the Dickey and Fuller tests are not reliable if k_1 is in the economically plausible range of $0.95 - 1.00$ for monthly data. Hakkio (1986) also reports that a variety of other univariate tests have low power.

In such a situation, it is essential to devise tests with increased power. This is achieved by extending the Dickey and Fuller tests to a system of univariate autoregressions, estimated jointly by Generalized Least Squares (GLS). As Zellner (1962) points out, the seemingly unrelated regression (SURE) method is

[7] Roll (1979) reports individual t-statistics, whereas Adler and Lehmann (1983) report joint F-tests that all coefficients are zero. They are both unable to reject the hypothesis that the coefficients are different from zero.

[8] Evans and Savin (1981) have derived the exact sampling distribution of k_1.

more efficient than OLS equation by equation because it fully exploits the information in cross-equation correlations. In addition, estimating the regression country by country is clearly inefficient given that the null hypothesis is that *all* values of k_1 are unity. Clearly, this restriction should be imposed across all countries.

The contribution of this paper is to estimate the autoregression model in a multivariate setting which should allow for more powerful tests. The drawback of this method is that the distribution of these test statistics is unknown for a system of equations. Therefore, the empirical distributions had to be derived by simulation analysis, as in Fuller (1976).

The simulation is conducted as follows. The data generation process is based on the autoregressive model with $k_0 = 0$ and $k_1 = 1$. The sample size is chosen to be $T = 180$, which corresponds to the number of observations in the floating exchange rate period 1973–1987, and $N = 10$ countries. In each experiment, N error terms u_t are generated jointly T times from a multivariate normal distribution with mean zero and a given covariance matrix. Table I describes means, standard deviations and correlations of changes in the logarithm of real exchange rates over the sample period January 1973–December 1987. The covariance matrix for the simulation is taken from this data set.

This procedure allows to simulate e_t from the autoregressive model (2) and to compute a sample value of the statistics. Each experiment is then replicated 5000 times, which generates a sample distribution of the statistics for a known autoregression coefficient. The power function, which reports the probability of rejecting the null for various values of the parameter k_1, is found in a similar fashion with 1000 replications.[9]

The top half of Table II presents the 5% and 10% critical values for $\hat{\rho}_\mu$ and $\hat{\tau}_\mu$, both for the usual univariate OLS tests (from Fuller) and for the GLS restricted tests (from the simulation). Given these critical levels, the power functions are calculated by simulation under the alternatives $k_1 = 0.90, 0.95, 0.975, 0.99, 1.00$, and reported in the middle of the table. Let us focus, for instance, on the tests with size 5%. For the OLS univariate regressions, the empirical probability of rejecting the null is about 4.5% for both ρ_μ and τ_μ, which confirms the size of the Fuller tests. Unfortunately, these tests have low power: for values of k_1 such as 0.975, the probability of rejecting the null is only 16.5% and 10.5%, respectively for ρ_μ and τ_μ. But this is not all: Evans and Savin (1984) have shown that if the intercept k_0 is different from zero in the true model, the power falls even further, and the tests may be biased, i.e. reject less frequently than expected when $k_1 = 1$. Since the null hypothesis does not preclude a random walk with a trend, the power functions reported in the table should be considered as upper limits to the power of the Dickey and Fuller tests. Finally, the second line from the bottom in the table shows that the usual OLS estimator k_1 is downward biased. The asymptotic bias is at most $-(1 + 3)/180 = -0.022$, which is consistent with the results found here.

In contrast, the GLS method, which restricts the autoregressive term to be the

[9] Because a different number of replications is used to compute the critical values and the power functions, the empirical power under the null hypothesis $k_1 = 1$ may be slightly different from the test size.

Table I
Changes in Real Exchange Rates: Statistics for 10 Countries January 1973–December 1987

The statistics refer to the change in the log of the real exchange rate $\Delta e_{t+1} \equiv e_{t+1} - e_t$, where e_t is defined as $\ln(SP^*/P)$, with S measured in dollars per foreign currency unit, P^* and P the foreign and domestic price levels. The ARCH parameters are estimated from the following model for the conditional variance h_t: $h_t = \alpha_0 + \alpha_1 (\Delta e_{t-1} - \mu)^2$.

	Belg.	Can.	France	Germ.	Italy	Japan	Neth.	Norw.	Switz.	UK.
Mean	.00134	−.00074	.00145	.00149	.00127	.00436	.00184	.00183	.00359	.00166
Std. Dev.	.03597	.01341	.03410	.03595	.02971	.03460	.03577	.03076	.03984	.03286
Correlation										
Belgium	1.000									
Canada	.287	1.000								
France	.896	.253	1.000							
Germany	.959	.278	.895	1.000						
Italy	.765	.179	.806	.765	1.000					
Japan	.613	.186	.630	.614	.550	1.000				
Nether.	.955	.293	.889	.957	.772	.611	1.000			
Norway	.855	.217	.830	.865	.668	.600	.835	1.000		
Switzer.	.855	.243	.823	.878	.743	.657	.862	.774	1.000	
UK	.631	.238	.621	.604	.539	.481	.628	.619	.575	1.000
ARCH Parameters:										
α_0	.00118	.00016	.00107	.00119	.00081	.00102	.00112	.00072	.00139	.00102
α_1	.081	.086	.074	.077	.074	.156	.115	.261	.142	.060

Table II

Comparison of Univariate OLS and Restricted GLS Tests by Simulation
Null Hypothesis: $k_1 = 1$, $k_0 = 0$

The test statistics are computed from the regression $e_{t+1} = k_0 + k_1 e_t + u_{1t+1}$ as $\rho_\mu = T(k_1 - 1)$, $\tau_\mu = (k_1 - 1)/\sigma(k_1)$, using a sample size of 180 months. The "OLS univariate" regressions are estimated individually country by country, while the "GLS restricted" regressions are based on simultaneous estimation of 10 series, constraining the coefficient k_1 to be equal across series. The test size is the probability that the sample value of the test statistics ρ_μ, τ_μ will be lower than the associated critical value, thus leading to rejection of the null hypothesis. The power functions describe the probability (in percent) of rejecting the null $k_1 = 1$ for various true values of k_1. Simulations assume $k_0 = 0$ and the covariance matrix of real exchange rates changes from Table I. All simulations consist of 1000 experiments, except for the critical values, which are derived from 5000 experiments.

		Test Size									
		5%					10%				
Critical Values											
OLS Univariate Tests: (from Fuller)	ρ_μ			−13.9					−11.1		
	τ_μ			−2.88					−2.57		
GLS Restricted Tests: (from simulation)	ρ_μ			−5.44					−4.91		
	τ_μ			−5.63					−5.30		
		$k_1 =$					$k_1 =$				
		0.90	0.95	0.975	0.99	1.00	0.90	0.95	0.975	0.99	1.00
Power Functions (percent)											
OLS Univariate Tests:	ρ_μ	90.7	39.5	16.5	8.5	4.5	98.1	60.6	29.7	16.5	9.5
	τ_μ	78.3	25.5	10.5	6.3	4.6	92.3	43.8	20.3	12.3	9.5
GLS Restricted Tests:	ρ_μ	100.0	100.0	95.3	44.2	4.6	100.0	100.0	98.3	61.7	9.2
	τ_μ	100.0	100.0	71.4	21.8	4.3	100.0	100.0	86.3	37.4	9.6
Average $\hat{k}_1$											
OLS Univariate:		.877	.925	.948	.961	.973					
GLS Restricted:		.886	.934	.957	.970	.981					

same across countries, yields much more powerful tests. For instance, when $k_0 = 0$ and $k_1 = 0.975$, the probability of rejection is now 95.3% and 71.4% for ρ_μ and τ_μ, respectively. This dramatic improvement can be explained by means of Figures 1 and 2, which compare the sample distributions, based on 5000 observations, of the OLS and GLS autoregressive coefficient k_1. Although the GLS estimator is also downward biased, the GLS distribution is tighter than the OLS distribution, which appears to be highly skewed to the left. For instance, the 5% lower cutoff value of the GLS distribution is 0.970, while it is 0.923 for the OLS distribution. This explains why the critical values of the GLS ρ_μ statistic are higher than those of the traditional Dickey and Fuller tests. This effect, combined with lower standard errors, also applies to the GLS τ_μ statistic. Thus pooling across assets reduces the dispersion in the distribution of k_1 and makes it more likely to reject false hypotheses.

Since the null hypothesis does not preclude a drift, the simulations were also performed with the true k_0's equal to their sample estimates, described in Table I. Table III reports the true critical values under this new scenario, as well as the empirical probabilities of rejection using the critical values in Table II, that assume $k_0 = 0$. As in the case of the Dickey and Fuller tests, it appears that the GLS tests based on the assumption of no drift do not reject often enough if the k_0's are truly non-zero. For instance, the true 5% critical level of ρ_μ is -4.49 instead of -5.44 if a drift is allowed for, so that rejection occurs only with a 1.2%

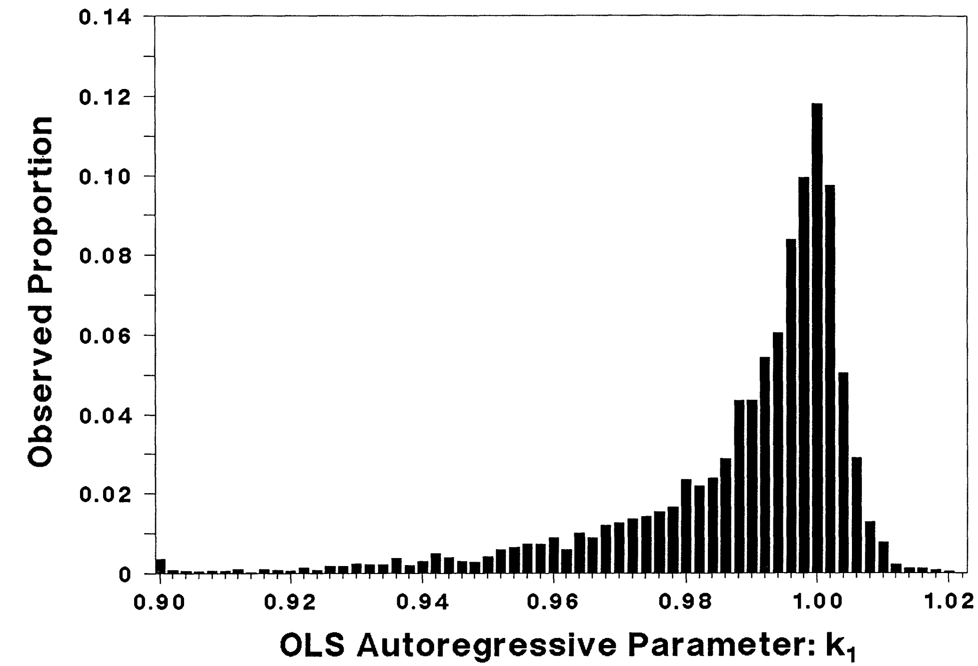

Figure 1. Histogram of univariate OLS k_1. Frequency distribution of the autoregressive parameter k_1 in univariate regressions estimated by Ordinary Least Squares (OLS). Each bar indicates the proportion of times the corresponding value of k_1 on the horizontal axis was observed in the simulaton. The average value of k_1 is 0.973.

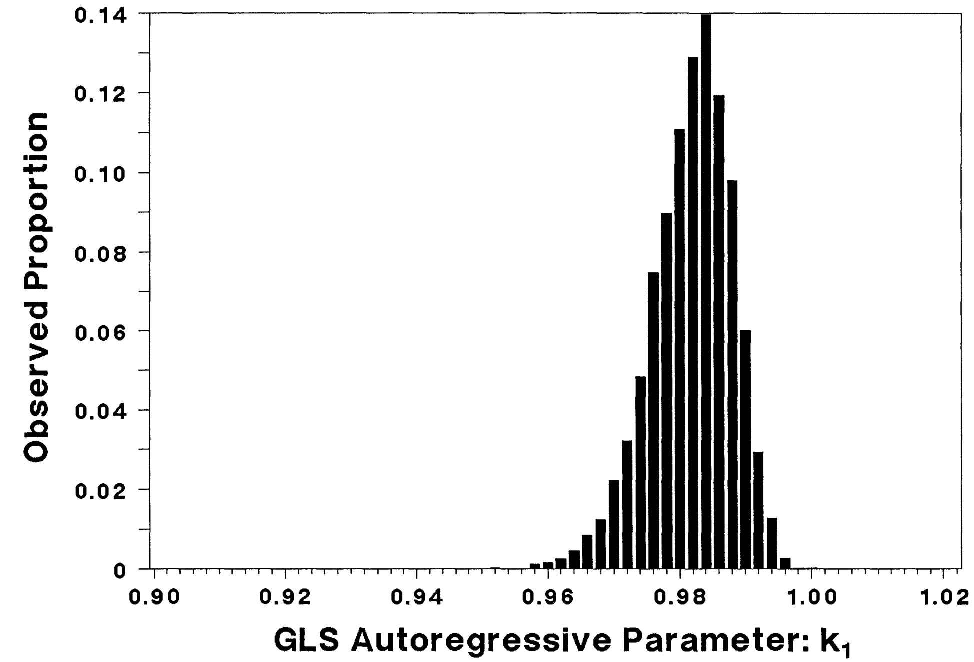

Figure 2. Histogram of restricted GLS k_1. Frequency distribution of the autoregressive parameter k_1 in a system of regressions estimated by Generalized Least Squares (GLS), where k_1 is restricted to be the same across equations. Each bar indicates the proportion of times the corresponding value of k_1 on the horizontal axis was observed in the simulation. The average value of k_1 is 0.981.

frequency under a unit root. Thus, the power functions reported in Table II, although considerably improved relative to the classical tests, will not be attained in practice if the real exchange rate displays drifts.

Next, we investigate whether the previous simulations are sensitive to misspecifications in the covariance matrix. Table IV reports the rejection frequency for the GLS tests, again using the 5% critical values from Table II. Three homoskedastic formulations are used: (i) the "Full Historical Matrix," (ii) a "Half-Correlations Matrix," where the historical correlations are replaced by half their value while keeping variances unchanged, (iii) a "Diagonal Matrix" which sets correlations to zero. The table also reports the true critical values, under each specification for the covariance matrix, as well as average autoregressive coefficients. As the table shows, these changes in the structure of correlations have little effect on the tests.

Further, to assess whether these results are sensitive to heteroskedasticity, the simulations were also run with the errors generated by an Autoregressive Conditionally Heteroskedastic (ARCH) process. Specifically, an ARCH(1) process was first fitted to the 10 real exchange rate changes over the period 1973–1987:

$$\Delta e_{t+1} \mid t \sim N(\mu,\ h_{t+1}),\ h_{t+1} = \alpha_0 + \alpha_1\ (\Delta e_t - \mu)^2, \tag{8}$$

where the conditional variance h_{h+1} is a linear function of the last squared innovation only. The parameter α_0 and α_1 are reported in the last two lines of

Table III

Simulations of GLS Tests with Non-Zero Drift Parameters

The test statistics are computed from the regression $e_{t+1} = k_0 + k_1 e_t + u_{1t+1}$ as $\rho_\mu = T(k_1 - 1)$, $\tau_\mu = (k_1 - 1)/\sigma(k_1)$, using a sample size of 180 months. The "OLS univariate" regressions are estimated individually country by country, while the "GLS restricted" regressions are based on simultaneous estimation of 10 series, constraining the coefficient k_1 to be equal across series. The difference with Table II is that the true values of k_0 are now equal to their sample counterparts, reported in Table I. The probability of rejection is computed using the critical values from Table II.

		Test Size									
		5%					10%				
True Critical Values											
GLS Restricted Tests:	ρ_μ			−4.49					−3.97		
	τ_μ			−5.24					−4.88		
				$k_1 =$					$k_1 =$		
		0.90	0.95	0.975	0.99	1.00	0.90	0.95	0.975	0.99	1.00
Probability of Rejection (percent)											
GLS Restricted Tests:	ρ_μ	100.0	100.0	94.5	33.9	1.2	100.0	100.0	98.1	51.1	3.2
	τ_μ	100.0	100.0	72.5	20.9	1.6	100.0	100.0	87.0	34.3	4.3
Average $\hat{k}_1$											
GLS Restricted Tests:		.886	.934	.958	.972	.985					

Table IV

Simulations of GLS Tests with Misspecified Covariance Matrices 5% Test Size

The test statistics are computed from the regression $e_{t+1} = k_0 + k_1 e_t + u_{1t+1}$ as $\rho_\mu = T(k_1 - 1)$, $\tau_\mu = (k_1 - 1)/\sigma(k_1)$, using a sample size of 180 months. The "OLS univariate" regressions are estimated individually country by country, while the "GLS restricted" regressions are based on simultaneous estimation of 10 series, constraining the coefficient k_1 to be equal across series. The probability of rejection is computed using the 5% critical values from Table II. The covariance matrix for the error terms is modified as follows. "Full Historical Matrix" is the historical covariance matrix. "Half Correlations Matrix" replaces the correlation coefficients by half their value while keeping variances unchanged. "Diagonal Matrix" sets correlations to zero. "ARCH Variances" uses conditional variances from a first-order Autoregressive Conditionally Heteroskedastic model, where (1) the parameters have been estimated from historical data, (2) the parameter α_1 has been set to 0.6.

		$k_1 =$					True Critical Values
		0.90	0.95	0.975	0.99	1.00	
Probability of Rejection (percent)							
Full Historical Matrix	ρ_μ	100.0	100.0	95.3	44.2	4.6	−5.44
	τ_μ	100.0	100.0	71.4	21.8	4.3	−5.63
Half Correlations Matrix	ρ_μ	100.0	100.0	94.9	41.5	3.8	−5.30
	τ_μ	100.0	100.0	70.1	20.0	3.8	−5.54
Diagonal Covariance Matrix	ρ_μ	100.0	100.0	95.0	41.1	3.7	−5.27
	τ_μ	100.0	100.0	69.7	19.3	3.6	−5.54
ARCH Variances:							
(1) Estimated Parameters	ρ_μ	100.0	100.0	95.4	43.9	5.3	−5.50
	τ_μ	100.0	100.0	69.8	22.1	4.9	−5.63
(2) Parameter $\alpha_1 = 0.6$	ρ_μ	100.0	100.0	93.7	46.2	7.4	−5.64
	τ_μ	100.0	100.0	70.0	25.3	7.0	−5.77
Average $\hat{k}_1$							
Full Historical Matrix		.886	.934	.957	.970	.981	
Half Correlations Matrix		.886	.935	.958	.970	.981	
Diagonal Covariance Matrix		.886	.935	.958	.971	.981	
ARCH Variances:							
(1) Estimated Parameters		.886	.934	.957	.970	.981	
(2) Parameter $\alpha_1 = 0.6$		.884	.933	.957	.970	.981	

Table I.[10] Since the ARCH effect is relatively small for this monthly data set, we also investigated a process with α_1 arbitrarily set at 0.6.[11] Next, in the simulations, the error terms are taken to be conditionally distributed as

$$u_{t+1} \mid t \sim N(0, h_{t+1}), \quad h_{t+1} = \alpha_0 + \alpha_1 u_t^2. \tag{9}$$

Given the correlation coefficients in Table I, assumed constant, a new covariance matrix can be computed for each observation t.

Table IV reports the rejection frequencies when the error terms are generated by the two ARCH processes. Again the tests do not seem too sensitive to this

[10] Further details on the estimation procedure can be found for instance in Jorion (1988).

[11] The value of α_0 was then computed as $\sigma^2(1 - \alpha_1)$, where σ^2 is the unconditional variance.

alternative specification.[12] It is only when α_1 equals 0.6, which is a very high value by historical standards, that rejections occur slightly too often.

In summary, it seems that the power increases reported above stem from jointly estimating a system of equations, and are little affected by the precise functional form of the covariance matrix. As a result, such multivariate tests are likely to be much more informative than univariate tests.

III. Empirical Results

The primary source of data is the International Monetary Fund's International Financial Statistics (IFS), which includes month-end exchange rates in local currency per U.S. dollar (IFS line ae), as well as consumer price indices (IFS line 64) proxying for price levels.[13] The data sample contains monthly observations from January 1973 to December 1987, amounting to 180 data points. Ten industrial countries are selected for the analysis.

The IMF data sample period covers the floating exchange rate period with 15 years of monthly data. Given that it may take a number of years for real exchange rates to revert to their PPP values, the analysis is also performed with annual data compiled by Lee (1978). This data set includes annual observations for average exchange rates and wholesale price indices for the U.S. and eight other countries from 1900 to 1972.[14] We hope that annual data covering long time periods will provide the kind of low frequency evidence in favor of PPP.[15]

First, the left panel in Table V presents univariate OLS autoregressions. All the point estimates of k_1 are lower than one, by an amount which is not very different from the small sample bias of about 0.02 discussed above. In addition, the test statistics $\hat{\rho}_\mu$ and $\hat{\tau}_\mu$ are well above their 10% critical levels of -11.1 and -2.57, respectively. These results are similar to the findings of Roll (1979) and Adler and Lehmann (1983), based on first differences instead of levels, who are also unable to reject the random walk hypothesis.

The multivariate regressions are reported in the right panel of Table V. Results are presented for k_1's restricted to be equal across countries. The results of the restricted GLS estimation provide marginal evidence against the random walk hypothesis. For instance, the value of $\hat{\rho}_\mu$ is -4.55, which is still above the 10% critical level of -4.91 reported in Table II. This observed value corresponds to an empirical marginal significance level of 16%. However, as noted in the previous

[12] In fact, Phillips (1987) has shown that the univariate distributions obtained by Dickey and Fuller are also valid in the presence of some heterogeneity in the innovation sequence, provided the innovations are martingale differences.

[13] See Frenkel (1976) who advocates the use of the CPI. The reported price indices are monthly averages, instead of true end-of-month numbers. Korajczyk (1985), however, reports that using proxies for end-of-month price levels does not change the nature of his results.

[14] A few countries had missing observations, which were collected from other sources. There was no data for Germany during the hyperinflation of 1922–24; for these three years, end-of-year exchange rates were collected from the Federal Reserve's *Banking and Monetary Statistics*, and prices were taken from the League of Nation's *Monthly Bulletin of Statistics*. During World War II, missing exchange rates were collected from the Crédit Suisse's *Bulletin Financier*, and missing prices from the United Nations' *Monthly Bulletin*.

[15] We thank one of the referees for this suggestion.

Table V

Autoregressions of the Real Exchange Rate: $e_{t+1} = k_0 + k_1 e_t + u_{t+1}$ Monthly Data: January 1973–December 1987

Under the null hypothesis that $k_1 = 1$, $k_0 = 0$, the one-sided 10% critical levels of $\rho_\mu = T(k_1 - 1)$ and $\tau_\mu = (k_1 - 1)/\sigma\,(k_1)$ are −11.1 and −2.57, respectively, for the OLS tests, and −4.91 and −5.30, respectively, for the GLS tests. However, if the k_0's are set equal to their sample values, the one-sided 10% critical levels of ρ_μ and τ_μ are −3.97 and −4.88, respectively, for the GLS tests.

	OLS Unrestricted			GLS Restricted		
	k_0 (SE)	k_1 (SE)	ρ_μ τ_μ	k_0 (SE)	k_1 (SE)	ρ_μ τ_μ
Belgium	0.0045	0.9882	−2.12	0.0081	0.9747	−4.55
	(0.0041)	(0.0118)	−1.00	(0.0030)	(0.0051)	−4.92
Canada	−0.0029	0.9841	−2.86	−0.0042		
	(0.0019)	(0.0119)	−1.34	(0.0012)		
France	0.0047	0.9846	−2.76	0.0068		
	(0.0039)	(0.0141)	−1.09	(0.0028)		
Germany	0.0075	0.9844	−2.81	0.0053		
	(0.0060)	(0.0139)	−1.12	(0.0024)		
Italy	0.0031	0.9887	−2.03	0.0254		
	(0.0034)	(0.0161)	−0.70	(0.0050)		
Japan	0.0115	0.9914	−1.55	0.0141		
	(0.0132)	(0.0156)	−0.55	(0.0037)		
Netherlands	0.0097	0.9838	−2.92	0.0133		
	(0.0072)	(0.0138)	−1.17	(0.0033)		
Norway	0.0139	0.9734	−4.79	0.0184		
	(0.0081)	(0.0171)	−1.55	(0.0042)		
Switzerland	0.0200	0.9720	−5.03	0.0058		
	(0.0108)	(0.0177)	−1.58	(0.0026)		
Britain	0.0055	0.9767	−4.20	0.0112		
	(0.0037)	(0.0170)	−1.37	(0.0033)		

section, this figure should be properly viewed as an upper limit, since these critical values were derived under the assumption that $k_0 = 0$. For instance, taking the drift terms k_0 equal to their sample averages, Table III reports 5% critical values of −4.49 and −5.24 for ρ_μ and τ_μ, respectively. Under this assumption, the observed statistic $\hat{\rho}_\mu = -4.55$ leads to rejection of the random walk hypothesis at the 5% confidence level.

One criticism of the AR(1) specification is that it restricts the dynamics of real exchange rates to only three possibilities: an explosive process, a random walk, or a monotonic adjustment to a constant value. In order to allow for a more general dynamic specification, lagged values of real exchange rate changes were added to the previous model:

$$e_{t+1} = k_0 + k_1 e_t + \textstyle\sum_{j=1}^{p} b_j(e_{t-j+1} - e_{t-j}) + u_{t+1}. \tag{10}$$

Fuller (1976) has shown that with a unit root, the large sample distribution of $\hat{\tau}_\mu$ is the same as that of the AR(1) univariate regression. With 12 lags, the regression yields $\hat{k}_1 = 0.967$ and $\hat{\tau}_\mu = -5.92$. Thus, using Table II, the τ_μ statistic leads to rejection of the random walk hypothesis at the 5% level of significance.

This evidence suggests that the true value of k_1, which determines the speed at which long-run PPP is reached in the aftermath of a shock, is a number such as 0.99, very close to one for monthly data over the floating exchange rate period. In economic terms, this indicates that a 50% initial deviation from PPP would take 69 months to revert to a 25% deviation. With $k_1 = 0.98$, another plausible figure, this half-life is reduced to 34 months.

Since adjustments seem to take place over a long period of years, we also investigated the random walk hypothesis with annual data over the much longer time period 1901–1972. This period covers the gold standard (1900–1914, and around 1925–1931), a flexible exchange rate period and the gold exchange standard (1945–1971). Table VI contains the results of the OLS univariate and GLS restricted estimation procedures. This time, the OLS-derived Dickey and Fuller tests lead to strong rejections of the null hypothesis for six out of eight countries. This is in contrast with the results of Adler and Lehmann (1983), who use the same data set but are unable to reject the random walk hypothesis, and can be explained by the use of the Dickey-Fuller tests instead of the regression in differences. The GLS restricted results are even stronger: a simulation was performed, as in Table II, for this data sample, and under the null $k_1 = 1$, there was not one instance where these observed statistics were attained out of a thousand experiments. This evidence, much more clear-cut than for the 1973–1987 flexible exchange rate period, clearly shows that long time series of annual

Table VI

Autoregressions of the Real Exchange Rate:
$$e_{t+1} = k_0 + k_1 e_t + u_{t+1}$$
Annual Data: 1901–1972

The one-sided 10% critical levels of $\rho_\mu = T(k_1 - 1)$ and $\tau_\mu = (k_1 - 1)/\sigma(k_1)$ under the null are, for OLS, −10.9 and −2.59, respectively, and for GLS, −5.09 and −4.83, respectively.

	OLS Unrestricted			GLS Restricted		
	k_0 (SE)	k_1 (SE)	ρ_μ τ_μ	k_0 (SE)	k_1 (SE)	ρ_μ τ_μ
Canada	0.1881	0.8081	−13.82*	0.1674	0.8293	−12.29*
	(0.0732)	(0.0750)	−2.56	(0.0234)	(0.0236)	−7.24*
France	0.4041	0.5710	−30.89*	0.1603		
	(0.0935)	(0.0980)	−4.38*	(0.0264)		
Germany	0.1039	0.8955	−7.52	0.1689		
	(0.0540)	(0.0536)	−1.95	(0.0264)		
Italy	0.2250	0.7719	−16.43*	0.1679		
	(0.0781)	(0.0763)	−2.99*	(0.0301)		
Japan	0.3056	0.7344	−19.12*	0.1978		
	(0.0931)	(0.0812)	−3.27*	(0.0296)		
Netherlands	0.2218	0.7619	−17.14*	0.1589		
	(0.0730)	(0.0773)	−3.08*	(0.0249)		
Switzerland	0.1580	0.8633	−9.84	0.1967		
	(0.0678)	(0.0591)	−2.31	(0.0283)		
Britain	0.2163	0.7544	−17.69*	0.1500		
	(0.0683)	(0.0768)	−3.20*	(0.0221)		

* Significant at the 10% level.

data are well suited to detecting low frequency mean-reversion tendencies in real exchange rates.

Let us now analyze the speed of adjustment implied by the annual data. The average OLS slope coefficient is about 0.78. Taking the small sample bias into account, this translates into a most likely value of 0.81, with a half-life of 3.3 years, which is roughly consistent with the 3–5 years implied by the monthly 1973–1987 data.

Finally, for comparative purposes, we performed the same analysis in terms of nominal exchange rates over the period 1973–1987. Several significant differences appear. For the real exchange rate, the GLS restricted coefficient was 0.9747, which we found to be marginal evidence against the random walk hypothesis. For nominal exchange rates, on the other hand, the estimated coefficient is 0.9916, which provides absolutely no evidence that the autoregression coefficient is less than one. Taking the small sample bias into account, this suggests an autoregressive coefficient in excess of unity, which would imply that the nominal exchange rate follows an explosive process. These findings confirm the results of Meese and Singleton (1982), who conclude that the logarithm of the nominal exchange rate has a unit root. Further, similar results hold for the annual data: for the real exchange rate, the autoregressive coefficient is 0.8293, which is significantly less than one, while it is 0.9871 for the nominal exchange rate. None

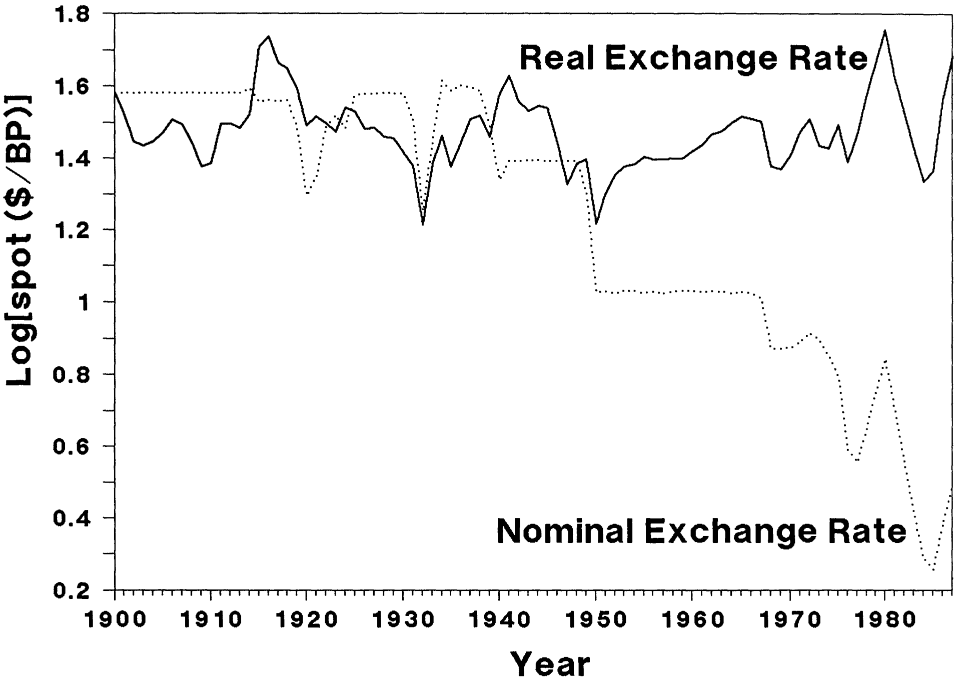

Figure 3. Dollar/pound exchange rate. Time-series plot of the real and nomimal exchange rate of the British pound versus the US dollar ($/BP), based on annual averages. The real exchange rate is obtained by deflating the nominal exchange rate by the ratio of wholesale price indices. The vertical scale is measured as the logarithm of the exchange rate. The real exchange rate is translated so that the 1900 value is equal to that of the nominal exchange rate.

of the test statistics is significant for the latter. These results suggest fundamental differences in the behavior of real and nominal exchange rates, which can only be caused by interactions between price levels and exchange rates. To illustrate this point, Figures 3 and 4 plot real and nominal exchange rates based on annual averages from 1900 to 1987, for the dollar/pound and dollar/French franc, respectively. While nominal exchange rates are clearly nonstationary, shocks to the real exchange rate seem to slowly cancel out over time.

IV. Conclusions

This paper has shown that there is no a priori theoretic reason for the real exchange rate to follow a random-walk, even in the context of efficient financial or commodity speculaton. Since long-run PPP, which is the basis for most exchange rate models, is invalidated if the random walk hypothesis holds true, empirical tests of this assumption carry a particular importance.

To date, most empirical tests of Purchasing Power Parity have been unable to reject the hypothesis that the real exchange rate follows a random walk. In our opinion, these results reflect the poor power of the tests employed rather than evidence against PPP. The contribution of this paper is to refine the econometric tests by pooling the data in a system of univariate autoregressions, and by using the Dickey and Fuller statistics. We demonstrate the improvement in the power of these statistics by Monte-Carlo experiments. Even then, it is hard to reach

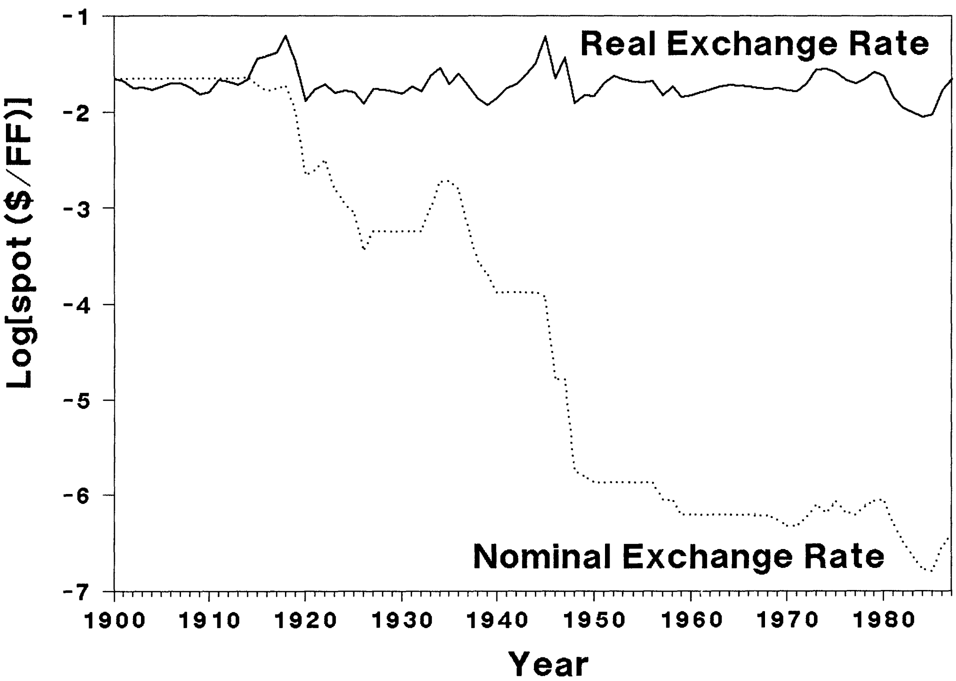

Figure 4. Dollar/French franc exchange rate. Time-series plot of the real and nominal exchange rate of the French franc versus the US dollar ($/FF), based on annual averages.

definite conclusions unless the sample period covers a number of cycles in deviations from PPP, that is a large number of years.

Overall, the empirical results of this paper cast doubt on the hypothesis that the real exchange rate follows a random walk. Our results suggest that the first-order autoregression coefficient approximately equals 0.98–0.99 for monthly data over the floating exchange rate period. If so, a 50 percent over-appreciation of a currency with respect to PPP would take between 3 and 5 years to be cut in half. Similarly, analyzing annual data over the period 1900–1972 reveals that a period of 3 years is needed for such a reversal.

On the other hand, it seems that nominal exchange rates are well approximated by a process with a unit root, which indicates that price levels are the prime reason for the long-term stability in real exchange rates suggested by our results.

In sum, this paper has shown that long-run PPP might indeed hold, although there is no debate that there are substantial, by most reasonable standards, short-term deviations from PPP.

REFERENCES

Adler, Michael and Bruce Lehmann, 1983, Deviations from purchasing power parity in the long run, *Journal of Finance* 38, 1471–1487.

Cumby, Robert E. and Maurice Obstfeld, 1984, International interest rate and price level linkages under flexible exchange rates: A review of recent evidence, in John F. O. Bilson and Richard C. Marston, eds.: *Exchange Rate Theory and Practice* (University of Chicago Press, Chicago, IL).

——— and John Huizinga, 1988, The predictability of real exchange rate changes in the short and long run, Mimeo, University of Chicago.

Dickey, David A. and A. Wayne Fuller, 1979, Distribution of estimators for autoregressive time series with a unit root, *Journal of American Statistical Association* 74, 427–431.

——— and A. Wayne Fuller, 1981, Likelihood ratio statistics for autoregressive time series with a unit root, *Econometrica* 49, 1057–1071.

Dornbusch, Rudiger, 1976, Expectations and exchange rate dynamics, *Journal of Political Economy* 84, 1161–1176.

Evans, G. B. A. and N. E. Savin, 1981, Testing for unit roots: 1, *Econometrica* 49, 753–779.

——— and N. E. Savin, 1984, Testing for unit roots: 2, *Econometrica* 52, 1241–1271.

Frenkel, Jacob A., 1976, A monetary approach to the exchange rate: Doctrinal aspects and empirical evidence, *Scandinavian Journal of Economics* 78, 200–224.

———, 1978, Purchasing power parity: Doctrinal perspective and evidence from the 1920's, *Journal of International Economics* 8, 169–191.

———, 1981, Flexible exchange rates, prices, and the role of 'news': Lessons from the 1970's, *Journal of Political Economy* 89, 665–705.

Fuller, W. A., 1976, *Introduction to Statistical Time Series* (John Wiley and Sons, New York).

Hakkio, Craig S., 1984, A re-examination of purchasing power parity: A multi-country and multi-period study, *Journal of International Economics* 2, 265–278.

———, 1986, Does the exchange rate follow a random walk? A Monte Carlo study of four tests for a random walk, *Journal of International Money and Finance* 4, 221–229.

Huizinga, John, 1987, An empirical investigation of the long-run behavior of real exchange rates, *Carnegie-Rochester Conference on Public Policy* 27, 149–214.

Jorion, Philippe, 1988, On jump processes in the foreign exchange and in the stock markets, *Review of Financial Studies* 2, 427–445.

Kendall, M. G., 1973, *Time-Series* (Griffin, London).

Korajczyk, Robert A., 1985, The pricing of forward contracts for foreign exchange, *Journal of Political Economy* 93, 346–368.

Lee, M. H., 1978, *Purchasing Power Parity* (Dekker, New York).

Meese, Richard and Kenneth Singleton, 1982, On unit roots and the empirical modeling of exchange rates, *Journal of Finance* 37, 1029–1035.

Mishkin, Frederick, S., 1984, Are real interest rates equal across countries? An empirical examination of international parity conditions, *Journal of Finance* 39, 1345–1357.

Mussa, Michael L., 1982, A model of exchange rate dynamics, *Journal of Political Economy* 90, 74–104.

Officer, Lawrence H., 1976, The purchasing power parity theory of exchange rates: A review article, *International Monetary Fund Staff Papers* 23, 1–60.

Phillips, P. C. B., 1987, Time series regression with a unit root, *Econometrica* 55, 277–301.

Roll, Richard, 1979, Violations of purchasing power parity and their implications for efficient international commodity markets, in Marshall Sarnat and Giorgio Szego, eds.: *International Finance and Trade* (Ballinger, Cambridge, MA).

Stulz, René, 1987, An equilibrium model of exchange rate determination and asset pricing with nontraded goods and imperfect information, *Journal of Political Economy* 95, 1024–1040.

Zellner, Arnold, 1962, An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias, *Journal of American Statistical Association* 57, 348–368.