IMF Working Paper

Medium-Term Exchange Rate Forecasting: What Can We Expect?

Guy Meredith

INTERNATIONAL MONETARY FUND

IMF Working Paper

Western Hemisphere Department

Medium-Term Exchange Rate Forecasting: What Can We Expect?

Prepared by Guy Meredith [1]

January 2003

Abstract

The views expressed in this Working Paper are those of the author(s) and do not necessarily represent those of the IMF or IMF policy. Working Papers describe research in progress by the author(s) and are published to elicit comments and to further debate.

The medium-term predictability of exchange rate movements is examined using three models of fundamentals: purchasing power parity, the monetary model, and uncovered interest parity. While the first two approaches yield favorable in-sample results, these largely reflect finite-sample estimation biases. Adjusting for these biases, there is little evidence of predictability, consistent with the lack of systematic improvement in out-of-sample forecasting performance relative to a random walk. Uncovered interest parity fares better at long horizons, but reflects information already embodied in market prices; in this sense, it may not be useful as an indicator of exchange rate misalignment. While more elaborate models of fundamentals might have better medium-term forecasting properties, careful attention must be paid to finite-sample biases in assessing predictability.

.

Contents   Page

Tables

Figures

I. INTRODUCTION

Since the pioneering study of Meese and Rogoff (1983), it has been widely recognized that exchange rate movements are difficult to predict in the short run. There is less agreement, however, on the predictability of longer-horizon changes. If exchange rates are ultimately tied down by economic fundamentals, there is a presumption that they will tend towards levels consistent with such fundamentals over time, even if short-term movements are erratic. Yet attempts to find evidence of the convergence of exchange rates to fundamentals have met with mixed success, and positive results have often proved less robust than initially believed. The question of predictability also has implications for identifying exchange rate misalignment: if exchange rates cannot be forecast, then misalignments (however defined) have no implications for future movements.

The mixed evidence on medium-term exchange rate predictability reflects, in part, econometric problems with estimating forecasting equations in general. In particular, there has been increasing recognition of a tendency toward over-fitting regression models in finite samples in the presence of autocorrelated regressors (as originally discussed by Stambaugh (1986)). Results are then biased toward finding evidence of predictability based on in-sample analysis. In this vein, recent studies of stock market returns have typically overturned earlier evidence in favor of predictability based on fundamentals such as dividend yields and earnings. In the case of exchange rate predictability, the econometric problems are particularly important given the relatively short sample period available under generalized floating — i.e., post 1973.

This paper revisits the issue of medium-term exchange rate predictability, paying close attention to finite-sample biases and differences between in- and out-of-sample performance. The analysis is confined to three relatively simple models of economic fundamentals: purchasing power parity (PPP); the "monetary model" of the exchange rate; and uncovered interest parity (UIP). These models are chosen, both because they have been actively studied in the past, and because their implementation is relatively straightforward and unambiguous. Many other, less parsimonious models have been developed to explain exchange rate movements.[1] They are not considered here, either because they have not been framed as forecasting models, or because their eclectic structure and data demands significantly complicate the assessment of forecasting ability.

To anticipate the conclusions, we find that both PPP and the monetary model do not robustly predict medium-term exchange rate movements: apparent in-sample success does not translate into out-of-sample forecasting performance, consistent with estimation bias. A tendency toward finding greater predictability at medium-term horizons is due to an increase in the magnitude of the bias as the horizon for the exchange rate change increases. UIP fares somewhat better at long horizons of ten years or so. However, UIP simply reflects the forward rate implied by market interest differentials, and thus does not add to existing market

---

[1] Examples include Clarida and others (2001) and Alberola and others (1999).

information on future exchange rate movements. In this sense, even its modest success does not imply an ability to assess exchange rate misalignments, defined as predictable deviations in exchange rate outturns from market expectations.

The paper is structured as follows. The next section discusses the evidence based on PPP, including the econometric issues associated with estimating mean-reverting relationships in finite samples. The third section reviews the monetary model and associated estimation issues, while the fourth assesses the performance of UIP. The last section provides concluding remarks.

II. PURCHASING POWER PARITY

There is a huge literature on the hypothesis that exchange rates return to levels consistent with PPP (see Rogoff (1996) for a survey). To the extent that PPP holds, if not continuously then over the medium term, the current deviation in the real exchange rate from PPP should predict future exchange rate movements. While many studies have covered this ground, we revisit the hypothesis in light of recent critical analysis of the predictive power of such models. Another innovation is that we define PPP in terms of real effective exchange rates (REERs), as opposed to real bilateral rates. While the logic underlying PPP should hold for both, its strength is likely to depend on the importance of trade linkages between countries. Using effective exchange rates weighted according to these linkages should raise the ratio of signal to noise compared with using bilateral rates, increasing the efficiency of estimates and test power.

The tests below generally use CPI-based measures of REERs from the IMF's Information Notice System (INS), as they are often employed to assess competitiveness.[2] Additional tests are performed with the GDP deflator-based measures used in the staff's multilateral exchange rate assessment exercise.[3] The mean-reverting properties of these data are examined for the G-7 currencies, and also for selected currencies of smaller industrial and emerging market countries.[4] Estimation was performed using monthly and annual data to compare the results at different frequencies. Medium-term predictability was tested using two approaches: either iterating forward predictions from the one-year ahead regressions, or directly estimating equations with the five-year change in the REER as the dependent variable. In addition, threshold variables were added to test for nonlinearities in the adjustment process.

---

[2] See Zanello and Desruelles (1997) for a description of these data. The weights are based on both bilateral and third-country competitiveness effects, and reflect trade in manufactures, non-oil commodities, and tourist services.

[3] See Isard and Mussa (1998) for a discussion of the methodology.

[4] Specifying the analysis in terms of mean reversion implies testing relative PPP. Given conceptual and practical problems with measuring absolute international price differences, no tests were performed of absolute PPP. However, Monte Carlo analysis of the implications of testing absolute versus relative PPP for biases in parameter estimates and test sizes are discussed below.

To preview the findings, the coefficients on lagged deviations of REERs from means were generally significant, with implied reversion speeds similar to those found elsewhere. Implied half lives of deviations from PPP were typically 3-4 years. Less favorably, these results can be attributed to finite-sample bias in the OLS estimator of the adjustment coefficient. After accounting for this bias, no evidence of mean reversion was found, as reflected in the model's poor out-of-sample performance. Regressions using five-year changes appeared to perform even better using in-sample data, but the out-of-sample evidence was again unfavorable.

Some evidence was found that threshold effects are important, with the speed and significance of adjustment appearing to increase sharply for deviations in exchange rates from means of greater than about ±10 percent. While this result appears to be larger than can be explained by small-sample bias, as calculated using Monte Carlo studies, there was again no evidence that threshold effects helped improve out-of-sample forecasting performance.

A. Econometric Issues

The basic mean-reversion regression for one-period changes using monthly and annual data was:[5]

$$\ell n(reer_t \,/\, reer_{t-1}) \;=\; \alpha + \beta\,(\ell n(reer_{t-1}) - \ell nreerm) \tag{1}$$

where $\ell nreerm$ is the sample mean of $ln(reer)$.[6] For five-year changes, the regression was:

$$\ell n(reer_t \,/\, reer_{t-5}) \;=\; \alpha + \beta\,(\ell n(reer_{t-5}) - \ell nreerm) \tag{2}$$

In the case of the five-year regressions, the standard errors of the parameters were corrected for MA(4) error terms using the Newey-West (1987) approach.

There are several complications associated with estimating regressions of this type. The first is that the OLS estimate of β is biased in finite samples, leading to estimated mean-reversion speeds that exceed their true value. In particular, as discussed in Stambaugh (1986) and (1999), the least-squares estimator of γ in the AR(1) regression:

$$x_t \;=\; \alpha + \gamma\, x_{t-1} + \varepsilon_t \tag{3}$$

---

[5]   Other specifications with dynamic terms were also tried — including lagged changes in the real exchange rate, and changes in the nominal exchange rate and relative prices separately. While these variables were sometimes significant, they had little effect on the mean reversion parameters reported below or the model's forecasting performance.

[6]   The estimated value of β, of course, would be identical if the mean of *REER* were not included given the presence of an unconstrained constant term.

is biased towards zero in finite samples. The bias can be approximated by:

$$E(\hat{\gamma}-\gamma) \approx -\left(\frac{1+3\gamma}{T}\right),$$ (4)

where T is the sample size. To express equation (3) in the form of the mean-reversion regressions, it can be rewritten as:

$$\begin{aligned} x_t - x_{t-1} &= \alpha + (\gamma-1)x_{t-1} + \varepsilon_t \\ &= \alpha + \beta x_{t-1} + \varepsilon_t \end{aligned}$$ (5)

where $\beta \equiv \gamma - 1$ and the mean of x_t is subsumed in the constant term α. The expected value of the estimated slope parameter in equation (5), $\hat{\beta}$, is then given by:

$$E(\hat{\beta}-\beta) \approx -\left(\frac{4+3\beta}{T}\right)$$ (6)

$$< 0 \quad if \quad \beta > -4/3$$

So $\hat{\beta}$ will show excess mean reversion in finite samples for any plausible value of β. Consider the null hypothesis that x is a random walk and thus $\gamma = 1$ and $\beta = 0$. $E(\hat{\beta})$ is then approximately $-4/T$. In monthly estimation, a sample period of 1980-2001 gives about 250 observations, implying an expected value for the mean-reversion parameter of -0.015 and an implied half-life of deviations from mean of 3.8 years under the null hypothesis of a random walk.[7] In annual estimation, with 21 observations, the expected value of $\hat{\beta}$ is -0.190 and the implied half life is 3.3 years under the null of a random walk.

For regressions using five-year changes in the exchange rate, the small-sample bias has not been derived analytically. We approximate it using Monte Carlo simulations, using data generated on the null hypothesis of a random-walk data-generation process (DGP):

$$x_t = x_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim N(0,1) .$$ (7)

Figure 1 shows resulting regression parameters over samples of different lengths for one-year and five-year ahead regressions, based on 1,000 replications of process (7). The parameters using one-year changes are very close to the analytical results given by formula (6). Those using five-year changes imply greater mean reversion: for a sample length of 17 years, equal to that used here for the 5-year regressions, the estimated parameter is -0.789. To take account of these small-sample biases in the estimates of β, the tables below show both raw parameters and bias-adjusted values. For the one-month and one-year ahead regressions, the

---

[7] The half life is calculated as $\log(0.5)/\log(1+\hat{\beta})$.

bias is calculated using the analytical approximation; for the five-year change regressions, it is based on the Monte Carlo results.
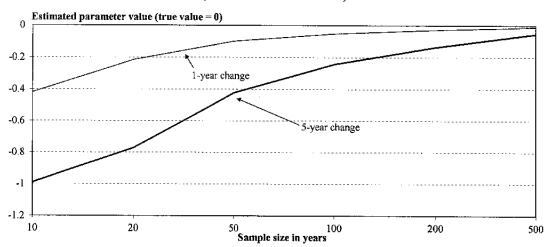
Figure 1. Parameter Bias in Mean-Reversion Regressions
(true DGP is a random walk)

Another perspective on finite-sample bias is given by Monte Carlo tests of how the adjustment parameter varies as the horizon for the exchange rate change increases, holding the length of the sample period fixed. Generating a random walk over a 21-year sample period, again with 1,000 replications, yielded the following results for the mean-reversion regression:

Horizon for exchange rate change	$\hat{\beta}$	$|t|$	$\bar{R}^2$
1 year	-0.213	1.68	0.081
2 years	-0.394	2.35	0.187
3 years	-0.545	2.82	0.273
4 years	-0.671	3.17	0.344
5 years	-0.789	3.36	0.394

The size and significance of the adjustment parameter increases steadily as the horizon of the exchange rate change lengthens, giving the misleading impression that predictability rises at medium-term horizons. Similarly, the $\bar{R}^2$ increases dramatically at longer horizons. These results underscore the dangers of making inferences about predictability based on in-sample tests, especially over medium-term horizons.

Separate from the issue of parameter bias, there is also a problem of size bias in hypothesis tests of mean reversion. As discussed in Engel (2000), if the true DGP does not correspond to the autoregressive specification implicit in the mean-reversion regressions, there will be a tendency to over-reject the hypothesis of a random walk. Engel shows that the size bias in such tests may be huge, even in samples of 100 years, when the exchange rate is driven by a mix of stationary and nonstationary processes. There does not, however, appear to be a straightforward way of assessing this bias without prior knowledge of the true DGP; Berkowitz and Giorgianni (2001) have shown that bootstrap methods, in particular, are not robust in this context.

A third problem in estimating mean-reverting processes is raised by Taylor (2001), specifically that the use of time-averaged data, when the underlying process takes place at a higher frequency, biases downward estimated speeds of adjustment. This effect would work in the opposite direction to the small-sample bias discussed above. Taylor conducts Monte Carlo experiments to calculate the time-averaging bias on the assumption that the true process is daily but estimation is performed using either monthly or annual average data. In either case, he finds that a true daily half-life of 2 years would be biased up by about 50 percent, giving a value using time-averaged data of around 3 years. Furthermore, analytic results show that the ratio of the estimated to the true half life tends to a finite limit exceeding unity as the true half life goes to infinity. Taylor does not provide a limiting value for this ratio, but his tables suggest that it is approximately 1.5. Neither does he analyze the sensitivity of the bias to sample size: although the bias itself is inherently an asymptotic phenomenon, it is not clear how it would interact with the small-sample bias discussed above.

Conceptually, however, there is the issue of how the time-averaging bias relates to the question addressed here — that of predicting exchange rate movements, as opposed to estimating "true" half lives in the underlying data. Taylor's analysis shows that time-averaging data can bias estimates of underlying half lives. Nevertheless, when time-averaged data are all that are available, the best unconditional predictor of the future exchange rate change still should be based on the biased parameter value, not the conceptual underlying value at a higher frequency. This is because the population value of the adjustment parameter changes when the data are time-averaged. So the estimated value is an unbiased measure of the population value using time-averaged data. For the purposes of exchange rate forecasting, then, the time-averaging bias is not relevant.[8]

There could still, however, be biases in the *size* of hypothesis tests that the adjustment parameter is zero when time-averaged data are used. Taylor does not address this issue, but one might expect under-rejection of the hypothesis that the data are a random walk using

---

[8] Similarly, Imbs and others (2002) discuss the effect of aggregation across goods on estimates of PPP, concluding that true adjustment speeds are faster than estimates using aggregated data suggest. As in the case of time aggregation, however, the relevant issue for forecasting purposes is the narrower one of whether the deviation in aggregate measures of the real exchange rate from historical means is of use in predicting exchange-rate movements.

time-averaged data. To explore this issue, Monte Carlo experiments were run where the DGP was mean reverting with a half-life of three years at a daily frequency. The data were replicated 1,000 times over sample lengths of both 21 years (corresponding to that used below in estimation) and of 250 years (assumed to be roughly asymptotic). The mean-reversion regressions were run on both the daily data and time-averaged monthly data, and the average adjustment coefficients, t statistics, and half lives were calculated. The results were as follows:

	Sample length					
	21 years (7560 days)			250 years (90000 days)		
Regression frequency:	$\hat{\beta}$	$\|t\|$	half life (years)	$\hat{\beta}$	$\|t\|$	half life (years)
Daily	-0.0013	2.30	1.5	-0.0007	5.27	2.9
Monthly	-0.0246	1.81	3.2	-0.0132	4.27	4.5

The t statistics on the slope parameters in the monthly regressions are indeed lower than those in the daily regressions, confirming a size bias in testing the null hypothesis that the process is a random walk. The results also shed light on the net bias arising from small-sample and time-averaging problems. In the "short" sample of 21 years, the estimated half life of 3.2 years using monthly data is close to the true half life of 3 years, indicating that the small-sample and time-averaging biases are roughly offsetting. Using daily data, the results are affected only by the small-sample bias, giving a half-life estimate roughly one half of the true value — consistent with the approximation in formula (6). In the 250-year sample, the test size distortion diminishes slightly; the half-life estimate using daily data is close to the true value of 3 years, while that using monthly data rises to about 50 percent above the true value, consistent with Taylor's results.

In practice, it is difficult to adjust for half-life bias arising from time averaging under a general null hypothesis. The approach taken below is to adjust the $\hat{\beta}$s by a constant equal to the asymptotic bias in the constrained GMM estimate. This gives a sense of the effect of time-averaging on the underlying half-life estimates, although many of the estimated adjustment parameters are of a perverse sign when corrected in this way. In any event, as indicated earlier, this adjustment is for presentational purposes, as the best estimate of the future exchange rate change will still reflect the unadjusted parameter.[9]

---

[9] Ideally one would use data that are not time-averaged. While end-of-period exchange rate data are available, however, the price indices needed to construct REERs are generally constructed using data obtained through a period of time, and thus are inherently time-averaged.

Yet another issue, again raised by Taylor (2001), arises from assuming a linear adjustment specification when the true process is nonlinear. A simple example involves threshold levels at which the adjustment speed of the exchange rate changes discontinuously. The adjustment parameter will then reflect a mix of the two processes that occur within and outside of these thresholds, and will be correct in neither case. This bias is addressed in the next section, which introduces nonlinearities to the error-correction specification.

This discussion suggests that attempts to assess predictability on the basis of in-sample evidence are subject to multiple difficulties. Furthermore, adjustments to account for biases in parameter estimates and test size distortions are likely to be sensitive to the specification of the null hypothesis. With these considerations in mind, considerable emphasis is placed on out-of-sample predictive performance in the following analysis.[10]

B. Linear Tests of Mean Reversion

Raw and adjusted estimation results for equation (1) using monthly data on CPI-based REERs are shown in Table 1, along with implied half lives of mean reversion in cases where the slope parameter is negative. The adjustments take into account both the approximate finite-sample bias from formula (6), as well as the ad hoc correction for using time-averaged data. No t-statistics are provided for adjusted parameters, given uncertainties about their distribution under the null hypothesis. All of the raw estimates of β for individual countries are negative, consistent with mean-reverting behavior, although only 3 of the 7 are significant at the 95 percent level. The adjusted R^2s are close to zero, and occasionally negative. Implied half lives range from a low of 2.2 years for the Canadian dollar to 11.2 years for the U.S. dollar. The equations were also estimated as a system using generalized method of moments (GMM), with the constraint imposed that the adjustment parameter was the same across countries (this was not rejected by the data). The common adjustment parameter of -0.015 is highly significant, implying a mean-reversion half life of 3.7 years.

The second column of Table 1 shows the adjustments to the parameters and half lives needed to correct for small-sample bias. Only two of the adjusted parameters remain negative, and the adjustment half lives are almost tripled to 6 years in those cases. The constrained parameter is about zero, suggesting that the series are approximately random walks. Finally, the third column presents estimates corrected for both the small-sample and time-averaging biases. The latter adjustment is crude, however, for the reasons discussed earlier, and amounts to subtracting 0.008 from each of the parameters shown in the second column. In any case, all of the parameters except that for the U.S. become negative again, but imply long half lives; the constrained estimate is negative with an implied half life of 7 years.

The picture using monthly data, then, is that unadjusted estimates of half lives are in the range of those found in other studies. Mean-reverting behavior, though, appears to be an artifact of small-sample bias. After correcting for this bias, the data appear to approximately

---

[10] Inoue and Kilian (2002) argue, in contrast, in favor of in-sample tests. But their results are asymptotic, in which case issues of small-sample bias disappear in any event.

follow a random walk. A further correction for time-averaging bias typically restores negative adjustment parameters, but with long implied half lives. This suggests that there could be mean reversion in the underlying process if it occurs at a higher frequency, although this would still not imply "forecastability" using time-averaged data.

Table 2 shows results using annual data. Again, the unadjusted parameters are all negative and imply half lives in the range found previously. As in the case of the monthly data, the correction for finite-sample bias substantially changes the results, with only three of the seven country-specific parameters remaining negative, and the pooled parameter turning positive. The further adjustment for time averaging again generally restores negative parameters, but with long implied half-lives.

Results using 5-year changes are shown in Table 3. The raw parameters are now centered around roughly minus one, suggesting complete mean reversion at this horizon. The $\bar{R}^2$ s are generally above 0.5, indicating that over half of the variance in real exchange rate changes can be predicted by deviations from PPP over 5-year horizons. After correcting for finite-sample bias, however, the signs of several of the coefficients are reversed, and the remainder become much smaller in absolute value. Nevertheless, two remain negative and statistically significant — those for the U.S. and Canada. The pooled parameter also remains negative and significant. On this basis, there is mild evidence of greater predictability of exchange rate movements at medium-term than short-term horizons.

Out-of-sample tests of predictability are inherently unaffected by finite-sample biases.[11] We assess out-of-sample performance in terms of goodness of fit, defined as the ratio of the root-mean-squared error (RMSE) of forecasts based on rolling regressions to the root-mean-squared deviation (RMSD) of ex post changes in the data. If this ratio (the Theil statistic) exceeds one, the model is out-predicted by a random walk; if less than one, the regression model out-predicts a random walk.[12] If the in-sample estimates are reflective of the out-of-sample properties of the equation, the square of this ratio has an expected value of one minus the $\bar{R}^2$ of the regression.

---

[11] The disadvantage is the limited number of observations available for out-of-sample tests, which reduces their efficiency and power. Given this drawback, it seems desirable to assess the out-of-sample evidence in conjunction with in-sample results adjusted for finite-sample bias.

[12] Diebold and Mariano (1995) compare the performance of various statistics that provide tests for the significance of deviations in this ratio from unity. Their Monte Carlo results, however, indicate potentially substantial size biases in tests based on any of the proposed statistics in samples of the length considered here. Given uncertainties about size biases in this context, no attempt has been made to assess the significance of deviations in this ratio from unity. Rather, we regard a more meaningful picture as being provided by the frequency distribution of positive and negative deviations in the ratio from unity. In other words, if roughly half (or more) of the ratios exceed unity, this suggests that the model does not outperform a random walk.

Table 1. Estimation Results for Monthly Mean-Reversion Regressions

Regression: $\ell n(reer_t / reer_{t-1}) = \alpha + \beta\, \ell n(reer_{t-1} / \overline{reer})$

	Raw estimate			Adjusted for small-sample bias		Adjusted for small-sample and time-averaging biases	
	$\hat{\beta}$	Half life	$\bar{R}^2$	$\hat{\beta}$	Half life	$\hat{\beta}$	Half life
US	-0.005 (0.63)	11.2	-0.00	0.011 (1.30)	...	0.003	...
Japan	-0.013 (1.55)	4.5	0.01	0.003 (0.35)	...	-0.005	11.3
Germany	*-0.012* (2.13)	4.7	0.02	0.004 (0.64)	...	-0.004	13.2
France	-0.012 (1.45)	4.9	-0.00	0.004 (0.52)	...	-0.004	15.1
UK	*-0.025* (2.42)	2.3	0.01	-0.010 (0.92)	6.0	-0.018	3.3
Italy	-0.009 (1.42)	6.5	-0.00	0.007 (1.11)	...	-0.001	54.9
Canada	*-0.025* (1.78)	2.2	0.01	-0.010 (0.69)	5.8	-0.018	3.2
Constrained GMM	*-0.015* (5.67)	3.7	...	0.000 (0.13)	...	-0.008	7.0

Notes:

- Sample period: 1981M1 – 2002M2.
- $\overline{reer}$ is the sample mean of *reer* .
- Absolute values of *t* statistics in parentheses.
- Half lives expressed in years.
- Small-sample bias calculated from formula (6) in text.
- Bold italics indicates a parameter is significant at the 95% level.

Table 2. Estimation Results for Annual Mean-Reversion Regressions

Regression: $\ell n(reer_t \, / \, reer_{t-1}) = \alpha + \beta \, \ell n(reer_{t-1} / \overline{reer})$

	Raw estimate			Adjusted for small-sample bias		Adjusted for small-sample and time-averaging biases	
	$\hat{\beta}$	Half life	$\bar{R}^2$	$\hat{\beta}$	Half life	$\hat{\beta}$	Half life
US	-0.148 (1.13)	4.3	0.014	0.050	...	-0.014	48.3
Japan	-0.187 (1.77)	3.4	0.096	0.004	...	-0.060	11.3
Germany	-0.123 (1.54)	5.3	0.065	0.079	...	0.015	...
France	-0.204 (1.80)	3.0	0.101	-0.016	43.3	-0.080	8.3
UK	*-0.192* (1.91)	3.2	0.118	-0.002	369.7	-0.066	10.2
Italy	-0.144 (1.59)	4.5	0.070	0.055	...	-0.009	73.9
Canada	*-0.338* (1.95)	1.7	0.123	-0.172	3.7	-0.236	2.6
Constrained GMM	*-0.180* (5.21)	3.5	...	0.012	...	-0.052	13.1

Notes:

- Sample period: 1981-2001.
- $\overline{reer}$ is the 1980-2001 mean of $reer$.
- Absolute values of t statistics in parentheses.
- Half lives expressed in years.
- Small-sample bias calculated from formula [(6)] in text.
- Bold italics indicates a parameter is significant at the 95% level.

Table 3. Estimation Results for Five-Year Mean-Reversion Regressions

Regression: $\ell n(reer_t / reer_{t-5}) = \alpha + \beta(\ell n(reer_t) - \ell nreerm)$

	Raw estimate		Adjusted for small-sample bias
	$\hat{\beta}$	$\bar{R}^2$	$\hat{\beta}$
US	*-1.299* (8.22)	0.698	-0.514
Japan	*-0.718* (7.73)	0.649	0.067
Germany	*-0.665* (3.84)	0.611	0.120
France	*-0.976* (4.81)	0.569	-0.191
UK	*-1.016* (5.55)	0.707	-0.231
Italy	*-0.584* (3.24)	0.233	0.201
Canada	*-1.446* (10.82)	0.812	-0.661
Constrained GMM	*-0.862* (117.08)	...	-0.077

Notes:

- Sample period: 1985-2001.
- *lnreerm* is the 1980-2001 mean of $\ell n(reer)$.
- Absolute values of *t* statistics in parentheses.
- Small-sample bias calculated from Monte Carlo results.
- Bold italics indicates a parameter is significant at the 95% level.

Table 4 shows the Theil statistics for the monthly and annual equations. The out-of-sample period for the monthly changes is 1991M1-2002M2; for the annual changes 1991-2001; and for the five-year changes 1995-2001. This provided a 10-year estimation window for the first out-of-sample prediction in each case. The means of the *REER*s for the error-correction terms were based on information known up to the end of the estimation period. Values for the Theil statistics of less than unity are indicated by bold italics. For one-month ahead predictions, the model outperforms a random walk only for Canada. For both the one-year and five-year ahead predictions, just three of the seven statistics are less than unity, indicating that a random walk works better in the majority of the cases. Taken as a whole, the results indicate that the mean-reversion model outperforms a random walk only about one third of the time. These out-of-sample results, then, are consistent with the picture provided by the in-sample estimates after bias correction. Only for the Canadian dollar is there consistent evidence of mean reversion.

The mean-reversion regressions were also run using the GDP deflator-based data over the same sample period. In terms of coverage, these data combine the euro-area countries in a singe aggregate, while including series on several smaller industrial countries (Australia, Denmark, New Zealand, Norway, Sweden, and Switzerland). For the large countries, the estimation results were similar to those obtained with the CPI-based data — the exception is Canada, where the GDP-deflator based series shows less evidence of mean reversion:

	CPI-based *REER*		GDP-deflator based *REER*	
	$\hat{\beta}$	$|t|$	$\hat{\beta}$	$|t|$
US	-0.148	1.13	-0.116	1.03
Japan	-0.187	1.77	-0.247	1.95
UK	-0.192	1.91	-0.197	1.59
Canada	-0.338	1.95	-0.069	0.65

Table 4. Theil Statistics for Out-of-Sample Forecasts from Mean-Reversion Regressions

	1-year changes: 1991M1-2002M2	1-year changes: 1991-2001	5-year changes: 1995-2001
US	1.000	1.073	1.275
Japan	1.001	1.002	*0.656*
Germany	1.010	1.069	1.229
France	1.028	1.465	1.719
UK	1.010	1.045	*0.976*
Italy	1.018	1.131	1.463
Canada	*0.996*	*0.918*	*0.793*

Notes:

- Ratio of the RMSE of the model to the RMSD of $\Delta\ell n(reer)$.
- Bold italics indicate a value below unity, indicating that the model outperforms a random walk.

Pooling the larger and smaller industrial countries into separate groups and estimating the common adjustment parameter within each group revealed a notably larger mean-reversion parameter for the smaller countries, with a $\hat{\beta}$ of -0.314 (and a $|t|$ statistic of 5.31) as opposed to -0.174 (3.26) for the larger countries. Out-of-sample prediction tests for the smaller countries performed no better than for the larger countries. In all cases, the Theil statistics for year-ahead changes for the 1991-2001 period exceeded unity:

	Theil statistic for year-ahead changes: 1991-2001
Australia	1.062
Denmark	1.201
Norway	1.038
New Zealand	1.116
Sweden	1.014
Switzerland	1.159

C. Nonlinear Adjustment and Time Trends

The above specification assumed that the underlying mean-reversion process is log linear. Some recent studies (e.g. Sarno and others (2002), and Sollis and others (2002)) have addressed the issue of whether the speed of adjustment changes as a function of the deviation of the exchange rate from mean. In particular, it is possible that transactions costs, tariffs, and similar frictions create a zone of movement in relative prices within which there is little or no scope for international goods arbitrage. One would expect to see evidence of mean reversion only when price deviations exceed the bounds of this zone.

Such effects can be tested for by adding terms to the standard regression that are either nonlinear functions of the deviations from mean, or that become operative when this gap reaches a certain threshold level (as in threshold autoregression, or TAR, models). One way of implementing the first approach is in the form a Taylor-series expansion of a general nonlinear function:

$$\Delta \ell n(reer_t) = \alpha + \beta\, \ell ngap_{t-1} + \theta\, \ell ngap_{t-1}{}^2\, sign(\ell ngap_{t-1}) + \ldots\ , \qquad (8)$$

where $\ell ngap$ is the log of the ratio of the actual to the mean exchange rate. The significance of the coefficients on the power terms can be tested using standard methods. The threshold approach can be implemented in its simplest form as:

$$\Delta \ell n(reer_t) = \alpha + \beta\, \ell ngap_{t-1} + \theta\, (\ell ngap_{t-1} - \tau)(\ell ngap_{t-1} > \tau \text{ or } \ell ngap_{t-1} < -\tau)\ , \qquad (9)$$

making the assumption that the size of the threshold (τ) and the shift in the adjustment parameter (θ) is the same on both sides of the mean exchange rate.

Preliminary tests of the two specifications indicated that the threshold model was more parsimonious and performed at least as well in terms of goodness-of-fit as the power-series approach. The appropriate size of the threshold was determined on the basis of log likelihood values from GMM estimation of the system of equations for the G-7 currencies, with common slope and threshold parameters imposed. The values of the log likelihood function (LLF) for different threshold levels (expressed in log deviation from mean) using monthly and annual data in estimation are:

Values of Log Likelihood Function for Different Thresholds		
Threshold	Monthly	Annual
none	5495.0	278.5
0.05	5496.5	280.4
0.08	*5497.8*	282.9
0.10	5497.0	284.4
0.12	5496.9	*285.9*
0.15	5495.9	283.6

The highest LLF values are shown in bold italics. They indicate that a threshold of about ±8 percent fits the monthly data best, while one of ±12 percent fits the annual data. In either case, the threshold variable is statistically significant.[13] For simplicity, a common threshold of ±10 percent was used in further analysis at both data frequencies. With a threshold of this size, about 40 percent of the observations at a monthly frequency lie outside the threshold, and 37 percent at an annual frequency.

The in-sample adjustment parameters obtained by estimating equation (9) using all available observations, and imposing common parameters across the G-7 currencies, are:

1-month change		1-year change		5-year change	
$\hat{\beta}$	$\hat{\theta}$	$\hat{\beta}$	$\hat{\theta}$	$\hat{\beta}$	$\hat{\theta}$
-0.009	-0.019	-0.040	-0.439	-1.061	0.405
(2.27)	(2.18)	(0.94)	(5.42)	(12.55)	(2.56)

---

[13] Twice the difference between the LLF value with and without the threshold is distributed $\chi^2(1)$, with a 95 percent critical value of 3.84.

For both one-month and one-year changes in the REERs, the estimated speeds of adjustment are significantly higher outside of the threshold (as reflected in the sum of $\hat{\beta}$ and $\hat{\theta}$) than within ($\hat{\beta}$ only). Indeed, for one-year changes, the within-threshold adjustment parameter becomes insignificant. In contrast, the results for the five-year changes suggest that adjustment is weaker outside the threshold than within.

We focus here on the one-year changes, because they are more favorable to the threshold hypothesis than the five-year changes, and more relevant to medium-term exchange rate forecasting than the one-month changes. The above results were compared with those from Monte Carlo experiments using a random-walk DGP with $N(0,1)$ innovations. The threshold was calibrated to yield the same percentage of observations outside the threshold as in the actual data (i.e. 37 percent). Using 1,000 replications for a 21-year sample, the fitted parameters were $\hat{\beta} = -0.278$ (0.50) and $\hat{\theta} = -0.318$ (1.68). The parameters are similar in size, and the parameter applying to observations outside of the threshold is not statistically significant. These properties contrast with the actual estimation results, suggesting that the finding of threshold effects is not simply a finite-sample anomaly.

The out-of-sample performance of the annual threshold model was analyzed over the 1991-2001 period. In addition, the model was generalized by adding a linear time trend to see whether this improved forecasting performance. Theil statistics for the various versions of the model are shown below, with the lowest values for each currency shown in bold italics. The model with thresholds outperforms the basic linear model in only two out of the seven cases, rising to three when a linear time trend is added. The basic model augmented only by a time trend fails to yield the lowest Theil statistic for any of the seven cases. In fact, the unaugmented basic model yields the best results in four of the seven cases, suggesting that elaborations to the PPP specification are questionable from a forecasting perspective. At the same time, it should be noted that a random walk still unambiguously outperforms any variant of the PPP specification, except for Canada.

	Basic model	With threshold	With time trend	With threshold and time trend
US	1.073	*1.044*	1.941	1.866
Japan	*1.002*	1.073	1.038	1.113
Germany	*1.069*	1.083	1.668	1.612
France	1.465	1.388	1.376	*1.330*
UK	*1.045*	1.112	1.184	1.299
Italy	1.131	*1.097*	1.188	1.173
Canada	*0.918*	0.940	1.098	1.114

D. Absolute Versus Relative PPP

The above are tests of relative PPP, as the model is specified in terms of reversion to a sample mean as opposed to an absolute measure of international price differences. The need to estimate both parameters — the sample mean and the adjustment speed — leads to the finite-sample bias discussed above. Using an exogenous measure of price differences, as opposed to the sample mean, implies estimating just the adjustment speed, potentially reducing inference problems.

While absolute PPP is not directly tested here, it is of interest to examine the theoretical impact of using absolute PPP on finite-sample biases to evaluate how promising this route would be. For this purpose, Monte Carlo studies were performed using a random-walk DGP, but with the constant term in the regression constrained to zero — i.e., the known population value. Regressions were estimated for exchange rate changes from 1 to 5 years over a 21-year sample, based on 1,000 replications. The results for regressions that both included and excluded the constant term are as follows:

Horizon of change in dependent variable in years	With constant term			Without constant term		
	$\hat{\beta}$	$|t|$	Rejection rate (percent)*	$\hat{\beta}$	$|t|$	Rejection rate (percent)*
1	-0.210	1.63	40.7	-0.034	0.68	12.1
2	-0.387	2.26	65.9	-0.065	0.93	29.7
3	-0.532	2.69	72.3	-0.093	1.09	41.9
4	-0.658	3.01	77.2	-0.117	1.19	49.5
5	-0.760	3.23	79.3	-0.134	1.23	55.1
* Rejection rate using a 5 percent nominal test size.						

It is apparent that the slope parameter is still biased (i.e., negative) when the constant term is constrained to its true value of zero, although the bias is reduced to roughly one sixth its original size. Interestingly, however, the bias in the width of confidence intervals for hypothesis tests remains large, especially as the horizon lengthens. In particular, the null hypothesis of a random walk should not be rejected more often than the nominal test size. With a 5 percent test size, however, the null hypothesis was rejected about 12 percent of the time using 1-year changes in the dependent variable, and 55 percent of the time using 5-year changes. So the likelihood of incorrectly rejecting the null hypothesis using absolute measures of PPP may still be inappropriately large, even though the parameter value itself is much less biased. These results, then, suggest that tests of absolute PPP may reduce some problems associated with relative PPP, but are not a complete solution.

III. MONETARY MODEL

Another popular framework for analyzing exchange rate "fundamentals" is the monetary model. In addition to continuous PPP, it assumes uncovered interest parity (UIP), a stable money demand function, and exogeneity of money and real income with respect to the exchange rate, the equilibrium exchange rate at a point in time (e_t^*) can be expressed (in logarithms) as:

$$e_t^* = (m_t - m_t^f) - \gamma(y_t - y_t^f) ,$$ (10)

where $m_t - m_t^f$ is the gap between the logs of the domestic and foreign money supplies, $y_t - y_t^f$ is the gap between real incomes, and γ is the income elasticity of money demand (assumed for simplicity to be the same across countries). Assuming further that deviations in the actual exchange rate from its equilibrium level are temporary, the monetary model implies the following error-correction specification for changes in the exchange rate:

$$e_{t+k} - e_t = \theta(e_t^* - e_t) ,$$ (11)

where k is some horizon over which disturbances to the exchange rate are assumed to decay.

Given the evidence against each of the individual building blocks of the monetary model, it might be surprising if a framework based on their collective performance worked well empirically.[14] Indeed, many subsequent tests of the type initially conducted by Meese-Rogoff concluded that the monetary model was not useful in forecasting near-term exchange rate movements. Some researchers in the mid-1990s, however, appeared to arrive at better results by looking at the medium-term performance of the model. Examples are Mark (1995), Chen and Mark (1995), Mark and Choi (1997), and MacDonald and Taylor (1994), which found that the monetary model contained predictive power that increased over the medium term.[15] Here we focus on the Mark (1995) results, as they are frequently cited. In his study, equation (11) was estimated over a variety of horizons for four currencies — the Canadian dollar, the deutsche mark, the Swiss franc, and the yen — against the U.S. dollar. The sample was based on quarterly data from 1973Q2 to 1991Q4. Bias-adjusted t statistics were calculated to assess the significance of the adjustment parameter θ, bootstrapping was performed to correct for size distortions in hypothesis tests, and out-of-sample Theil statistics were computed to assess goodness of fit relative to a random walk.

---

[14] Others, however, are surprised that it does not work better. Neely and Sarno (2002), for instance, find the model "intuitively appealing," and regard its low predictive power as a puzzle.

[15] Chinn and Meese (1995) obtained more ambiguous results using generalized monetary models — at the most favorable horizon (36 months), the monetary models outperformed a random walk in predicting the sign of the exchange rate change in 9 out of 15 cases, but the associated Theil statistic was less than unity in only 6 of 15 cases.

Mark found that, on either in-sample or out-of-sample criterion, the predictive power of the model increased as the horizon lengthened for three of the four currencies (the exception being the Canadian dollar, where the overall model did not work well at any horizon). At the longest horizon of 16 quarters, the root-mean-squared error (RMSE) of the model was about half that of a random walk. Taken at face value, this suggests that a high proportion — some 75 percent — of the variance in longer-horizon exchange rate movements could be predicted by the monetary model.

Several studies subsequently revealed, however, that Mark's results were not robust, in spite of the apparently sophisticated econometrics. In particular, extending the analysis beyond Mark's observation period revealed that the predictive performance of the model declined dramatically (Groen (1999), Kilian (1999)). Further, as shown in Faust and others (2001), the model only predicted well in a local region around the time that Mark performed his analysis, and its general historical performance deteriorated as the underlying economic data were revised. From an econometric viewpoint, the use of bootstrap methods to enhance statistical inference when the true parameter value is on the boundary of the admissible parameter space has been criticized by Andrews (2000).

Here we revisit the predictive performance for Mark's model beyond the end of his sample period in 1991Q4, based on updated and revised data. Reestimating Mark's equations for the 1973Q1-1991Q4 period yielded the following results at 1-quarter, 4-quarter, and 16-quarter ahead horizons (the t statistics for the 4-quarter and 16-quarter horizons have been adjusted using the Newey-West (1987) approach). As found by Mark, the results for the Canadian dollar are obviously unsatisfactory, with $\hat{\theta}$ having the wrong sign at all horizons. For the other three currencies the results are more favorable, and qualitatively similar to those reported by Mark using his original dataset. Most notably, the adjustment coefficients become larger and much more significant as the horizon extends to 16 quarters.

	1-quarter horizon		4-quarter horizon		16-quarter horizon	
	$\hat{\theta}$	$|t|$	$\hat{\theta}$	$|t|$	$\hat{\theta}$	$|t|$
Canadian dollar	-0.013	1.24	-0.043	1.13	-0.112	0.34
Deutsche mark	0.047	1.15	0.250	1.90	1.463	6.51
Swiss franc	0.084	2.32	0.380	3.13	1.231	13.13
Yen	0.024	0.82	0.178	1.58	0.986	3.28

Regressions with moving end points were then used to generate out-of-sample forecasts for 1992Q1-2001Q4. The resulting Theil statistics are shown below. Theil statistics are also shown for an alternative interpretation of the monetary model, where the dependent variable defined as the future change in the *monetary aggregate* as opposed to the exchange rate. This modified version would be more appropriate than the original model — which assumes that monetary aggregates are exogenous — on the assumption that monetary

aggregates are endogenous in the context of modern financial systems and monetary policy frameworks.

	1-quarter horizon		4-quarter horizon		16-quarter horizon	
	Exchange Rate	Money	Exchange rate	Money	Exchange rate	Money
Canadian dollar	*0.93*	1.35	*0.86*	1.43	*0.69*	2.18
Deutsche mark	0.92	*0.89*	*0.67*	0.98	0.85	*0.63*
Swiss franc	1.00	*0.78*	1.13	*0.83*	1.05	*0.74*
Yen	1.02	*0.43*	1.10	*0.47*	0.92	*0.59*

When the exchange rate is used as the dependent variable, the best prediction results are for the Canadian dollar — ironically, the currency for which the in-sample estimation results were perverse. Inspection of the results revealed that the adjustment parameter gradually changed sign in the mid-1990s, becoming significantly positive by the end of the sample period. While the model for Canada outperforms a random walk over the more recent period, then, there would have been little basis for using it based on the historical evidence. The model also appears to predict movements in the deutschemark somewhat better than a random walk, especially at the 16-quarter horizon. For the Swiss franc and the yen, a random walk generally outperforms the monetary model. In terms of forecasting future changes in monetary aggregates, in contrast, the monetary model uniformly outperforms a random walk, except for Canada. These results contrast with evidence in favor of medium-term exchange rate predictability found by Mark, underscoring the pitfalls of relying on in-sample data and limited out-of-sample observations. In spite of sophisticated attempts to correct for biases in parameter estimates and test sizes, it turns out that in-sample evidence in favor of the monetary model was misleading.

A more recent paper by Mark and Sul (2001) acknowledges the critiques of the earlier studies. Thus, they focus instead on short-horizon results, reestimating the monetary model for OECD countries using data from 1973Q1 to 1997Q1. Using the 1982Q1-97Q1 period for out-of-sample testing, they find that the model outperforms a random walk for 13 of 18 exchange rates at a one-quarter horizon, and for 17 of 18 exchange rates at a 16-quarter horizon. Revisiting this analysis using currently published data for the same group of countries yielded less favorable results — Theil statistics were obtained of less than unity for only 4 out of the same 18 currencies, suggesting that a random walk outperformed the monetary model at the one-quarter horizon.[16] Taking a second out-of-sample period of 1997Q2-2001Q3 yielded Theil statistics of less than unity for 5 of the 8 currencies for which data were available. While slightly better than the earlier findings, these results clearly fail to

---

[16] We did not investigate the performance at 16 quarters, given Mark and Sul's focus on the short-horizon results and the earlier critiques of long-horizon regressions.

provide persuasive support for the monetary model. It is not clear whether these differences in findings are the result of subsequent data revisions or differences in econometric methodology.

More recently, Rapach and Wolnar (2002) have looked for evidence in favor of the monetary model using much longer historical time series than employed in earlier studies — typically annual time series starting the late 1800s or early 1900s up to the 1990s. They obtain the following Theil statistics for out-of-sample, one-year ahead forecasts, using subsets of their overall data series (typically the second half of the series):

Country	Time period	Theil statistic
Belgium	1936-89	0.98
Finland	1954-95	1.02
France	1936-99	1.02
Italy	1939-95	0.94
Portugal	1944-95	1.01
Spain	1949-95	1.03
Switzerland	1939-95	0.99

In only 3 of the 7 cases is the Theil statistic less than unity. In other words, a random walk performs better than the monetary model for a majority of the countries. Rapach and Wolnar view the results — specifically the three cases with Theil statistics of less than unity — as evidence of the success of the monetary model, at least for some countries. Alternatively, the results can be interpreted as showing that the monetary model works no better, and perhaps slightly worse, than a random walk, with Theil statistics distributed around a mean of roughly unity.

In any event, support for the monetary model appears tenuous in spite of several attempts to find a predictive relationship. This is consistent with the empirical shortcomings of the building blocks of the model, perhaps most importantly the assumption that monetary aggregates can be taken as exogenous in predictive regressions. Under modern financial systems and approaches of central banks to monetary policy, it seems much more plausible that money is endogenously determined by other variables in the system.

IV. UNCOVERED INTEREST PARITY

Under the joint hypothesis of risk neutrality and efficient markets (RNEMH), the forward rate should be an unbiased predictor of the future exchange rate; it should also be a "sufficient statistic," in that no other information should add power to the forecast. In this sense, the above frameworks implicitly either assume the absence of RNEMH, or take a reduced-form approach that combines UIP with other relationships in an underlying model.

Past tests have overwhelmingly rejected UIP/RNEMH at short horizons, but the evidence appears more favorable at longer-term horizons (Meredith and Chinn (1998), and

Alexius (2001)). Here we reexamine the medium-term performance of UIP from a forecasting viewpoint. In performing out-of-sample tests, UIP has one striking advantage over the previous models, specifically that there are no parameters to estimate. Conventional estimation of UIP relationships, of course, typically include a constant term and an unconstrained slope coefficient. But, unlike the models looked at above, there are unambiguous theoretical priors about these parameter values under certain assumptions — i.e. zero and one respectively. With these constraints imposed, there is no need to perform estimation. Thus, all available data can be treated as out-of-sample observations.

Taking this approach, we compare the performance of UIP with that of a random walk over both short- and long-term horizons. The interest-rate data are updates of the series described in Meredith and Chinn (1998); exchange rates are from the IMF's *International Financial Statistics*. All data are end-period. Interest rate differentials are defined as the difference between the domestic and U.S. dollar yield on the corresponding instrument, while the exchange rate is the price of U.S. dollars in units of domestic currency.

Theil statistics are shown in Table 5 for UIP forecasts at 1-year, 5-year, and 10-year horizons, although at the longer horizons the country coverage is limited by the availability of interest rate data. At the 1-year horizon, UIP outperforms a random walk for only one out of the six currencies (the Italian lira) consistent with the extensive evidence on its failure at short-term horizons. At the 5-year horizon, results are available for only three currencies – the Canadian dollar has a Theil statistic of less than unity, while those for the deutschemark and the pound are both unity. At the 10-year horizon, in contrast, the Theil statistics are all less than unity using either constant-maturity bonds or benchmark government bond yields, with values that indicate an appreciable increase in forecast accuracy. For instance, the average Theil statistic of 0.74 at this horizon implies a reduction in the variance of the forecast error of 0.74^2, or close to one half.

These results suggest that UIP may be useful in forecasting exchange rate movements at longer horizons. The implications for assessments of exchange rate misalignment are not clear-cut, however. UIP defines a path for the exchange rate that is consistent with the implied forward rate existing in markets. Should the current exchange rate be considered misaligned in situations where its value is forecast to change, but in a way that is consistent with the forward rate in markets? Or should this be viewed as a situation where the exchange rate is in equilibrium, but one that is changing over time? The answer depends on the purpose of identifying misalignments. In particular, if the goal is to anticipate market tensions that will arise in the future from exchange rate misalignments, then one would only want to consider predictable deviations in the future rate from the forward rate. In this case, UIP would not be a useful criterion for identifying misalignment.

Table 5. Theil Statistics for UIP Forecasts

	12-month Eurocurrency (1980-2000)	5-year constant maturity (1978-2000)	10-year constant maturity (1983-2000)	Benchmark government bond yields – assumed 10-year maturity (1983-2000)
Japan	1.05	...	*0.71*	*0.74*
Germany	1.06	1.00	*0.64*	*0.65*
France	1.00	...	...	*0.96*
U.K.	1.05	1.00	*0.78*	*0.80*
Italy	*0.93*	...	...	...
Canada	1.07	*0.87*	*0.72*	*0.66*

Notes:
- Bold italics indicate values less than unity.

V. CONCLUDING REMARKS

The above results suggest that both PPP and the monetary model, in their basic forms, are of little use in forecasting medium-term movements in exchange rates. In-sample evidence in favor of these models reflects parameter bias arising from the use of finite samples for estimation. Because the bias tends to increase as the length of the horizon for exchange rate movements increases, there is a greater likelihood of finding spurious evidence of predictability at medium- than short-term horizons.

From a conceptual perspective, this does not necessarily imply that the fundamentals underlying these models, e.g., movements in international price levels, are not relevant for exchange rate determination. Rather, we conclude that they do not contain leading information that would make them useful in predicting exchange rate movements. Fundamentals could still play a role in determining exchange rates, but not have forecasting content, if other unobserved determinants of exchange rates are nonstationary. In this case, deviations in observed exchange rates from the levels determined by a subset of fundamentals will not tend to decay over time, violating the error-correction behavior underlying both PPP and the monetary model.

More fundamentally, the hypothesis of predictability based on these models sits uncomfortably with the notion that exchange rates are determined in efficient markets by profit-maximizing agents. In this case, models capable of reliably predicting exchange rate movements would yield profit-making opportunities. Their exploitation by market participants would then lead any information embodied in these models to be incorporated in the observed spot value of the exchange rate, as opposed to predictable movements in future rates.

Of course, it is well-known that deviations from the risk-neutral efficient-markets hypothesis must exist, as least at shorter horizons, to explain the biasedness of forward rates as predictors of future spot rates. But the models that have been developed to explain this anomaly, for instance those based on time-varying risk premia, have empirical implications that bear little relationship to the fundamentals-based models described above. In any case, the results for UIP presented above suggest that forward rates become better predictors of exchange rate movements at long as opposed to short horizons. This is consistent with the view that markets incorporate information on current and future fundamentals in the current level of the spot exchange rate.

The conclusions drawn here about medium-term predictability of exchange rate movements are, of course, limited to the relatively simple models discussed above. There may well be augmented versions of these models, or completely different frameworks, that yield more favorable results. The above analysis, however, suggests the need for caution in making inferences about predictability on the basis of in-sample results. Furthermore, interpreting such results and associated parameter biases is likely to become more difficult as the complexity of the model increases, given inherent scope for over-fitting relationships as models are refined. This, in turn, suggests the need to corroborate the results of in-sample analysis with out-of-sample evidence.

REFERENCES

Alberola, Enrique, Susana G. Cervero, Humberto Lopez, and Angel Ubide, 1999, "Global Equilibrium Exchange Rates: Euro, Dollar, Ins, Outs, and Other Major Currencies in a Panel Cointegration Framework," IMF Working Paper 99/175 (Washington: International Monetary Fund).

Alexius, Annika, 2001, "Uncovered Interest Parity Revisited," *Review of International Economics*, Vol. 9, pp. 505-17.

Alexius, Annika, 2002, "UIP, PPP, and the Random Walk Behaviour of Exchange Rates at Really Long Horizons" (unpublished).

Ang, Andrew, and Geert Bekaert, 2001, "Stock Return Predictability: Is it There?" (unpublished).

Andrews, Donald W.K., 2000, "Inconsistency of the Bootstrap When a Parameter is on the Boundary of the Parameter Space," *Econometrica*, Vol. 68, No. 2, March, pp. 399-405.

Berkowitz, Jeremy, and Lorenzo Giorgianni, 2001, "Long-Horizon Exchange Rate Predictability?," *The Review of Economics and Statistics*, Vol. 83, February, pp. 81-91.

Chen, Jian, and Nelson C. Mark, 1995, "Alternative Long-Horizon Exchange-Rate Predictors," *International Journal of Finance Economics*, Vol. 1, pp. 229-50.

Cheung, Yin-Wong, Kon S. Lai, and Michael Bergman, 2001, "Dissecting the PPP Puzzle: The Unconventional Roles of Nominal Exchange Rate and Price Adjustments" (unpublished).

Cheung, Yin-Wong, Menzie D. Chinn, and Antonio Garcia Pascual, 2002, "Empirical Exchange Rate Models of the Nineties: Are Any Fit to Survive?" (unpublished).

Chinn, Menzie D., and Richard A. Meese, 1995, "Banking on currency forecasts: How predictable is the change in money?," *Journal of International Economics*, Vol. 38, pp. 161-78.

Clarida, Richard H., Lucio Sarno, Mark P. Taylor, and Giorgio Valente, 2001, "The Out-of-Sample Success of Term Structure Models as Exchange Rate Predictors: A Step Beyond," NBER Working Paper 8601 (Cambridge, Massachusetts: National Bureau of Economic Research).

Diebold, Francis X., and Robert S. Mariano, 1995, "Comparing Predictive Accuracy," *Journal of Business and Economic Statistics,* Vol. 13, pp. 253-63.

Engel, Charles, 2000, "Long-Run PPP May Not Hold After All," *Journal of International Economics,* Vol. 57, pp. 243-73.

Engel, Charles, and James C. Morley, 2001, "The Adjustment of Prices and the Adjustment of the Nominal Exchange Rate" (unpublished).

Faust, Jon, John H. Rogers, and Jonathan H. Wright, 2001, "Exchange Rate Forecasting: The Errors We've Really Made," U.S. Federal Reserve Board International Finance Discussion Paper No. 714, December.

Groen, Jan J.J., 1999, "Long Horizon Predictability of Exchange Rates: Is It For Real?," *Empirical Economics,* Vol. 24, pp. 451-69.

Imbs, Jean, Mumtaz Haroon, Morten O. Ravn, and Helene Rey, 2002, "PPP Strikes Back: Aggregation and the Real Exchange Rate" (unpublished).

Inoue, Atsushi and Lutz Kilian, 2002, "In-Sample or Out-of-Sample Tests of Predictability: Which One Should We Use?" (unpublished).

Isard, Peter, and Michael Mussa, 1998, "A Methodology for Exchange Rate Assessment," in *Exchange Rate Assessment: Extensions of the Macroeconomic Balance Approach,* IMF Occasional Paper 167, ed. by P. Isard and H. Faruqee eds. (Washington: International Monetary Fund).

Kendall, Maurice G., 1954, "Note on the Bias in the Estimate of Autocorrelation," *Biometrika,* Vol. 41, pp. 403-404.

Kilian, Lutz, 1999, "Exchange Rates and Monetary Fundamentals: What Do We Learn From Long-Horizon Regressions?," *Journal of Applied Econometrics,* Vol. 14, pp. 491-510.

————, and Tao Zha, 1999, "Quantifying the Half-Life of Deviations from PPP: The Role of Economic Priors," Federal Reserve Bank of Atlanta Working Paper 99-21, December.

Lanne, Markku, 2002, "Testing the Predictability of Stock Returns," *The Review of Economics and Statistics,* Vol. 84, No. 3, August, pp. 407-15.

MacDonald, Ronald, and Mark Taylor, 1994, " The Monetary Model of the Exchange Rate: Long-Run Relationships, Short-Run Dynamics and how to Beat the Random Walk," *Journal of International Money and Finance,* Vol. 13, pp. 276-90.

Mark, Nelson C., 1995, "Exchange Rates and Fundamentals: Evidence on Long-Horizon Predictability," *American Economic Review,* Vol. 85, No.1, March, pp. 201-18.

————, and Donggyu Sul, 2001, "Nominal exchange rates and monetary fundamentals: Evidence from a small post-Bretton Woods panel," *Journal of International Economics*, Vol. 53, pp. 29-52.

Meese, Richard, and Kenneth Rogoff, 1983, "Empirical Exchange Rate Models of the 1970s: Do They Fit Out of Sample?," *Journal of International Economics*, Vol. 14, February, pp. 3-24.

Meredith, Guy, and Menzie D. Chinn, 1998, "Long-Horizon Uncovered Interest Rate Parity," NBER Working Paper 6797 (Cambridge, Massachusetts: National Bureau of Economic Research).

Neely, Christopher J., and Lucio Sarno, 2002, "How Well do Monetary Fundamentals Forecast Exchange Rates?" Federal Reserve Bank of St. Louis Review, September/October Vol. 84, No. 5.

Newey, N.K., and K.D. West, 1987, "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, Vol. 55, pp. 703-8.

Rapach, David E., and Mark E. Wohar, 2002, "Testing the Monetary Model of Exchange Rate Determination: New Evidence from a Century of Data," *Journal of International Economics*, Vol. 58, pp. 359-85.

Rogoff, Kenneth, 1996, "The Purchasing Power Parity Puzzle," *Journal of Economic Literature*, Vol. 24, June, pp. 647-68.

Sarno, Lucio, Mark P. Taylor, and Ibrahim Chowdhury, 2002, "Non-Linear Dynamics in Deviations from the Law of One Price: A Broad-Based Empirical Study," CEPR Discussion Paper No. 3377 (London: Centre for Economic Policy Research).

Sollis, Robert, Stephen Leybourne and Paul Newbold, 2002, "Tests for Symmetric and Asymmetric Nonlinear Mean Reversion in Real Exchange Rates," *Journal of Money, Credit, and Banking*, Vol. 34, No. 3, August, pp. 686-700.

Stambaugh, Robert F., 1986, "Bias in Regressions with Lagged Stochastic Regressors," CRSP Working Paper No. 156 (Loughborough, United Kingdom: Centre for Research in Social Policy).

————, 1999, "Predictive Regressions," *Journal of Financial Economics*, Vol. 54, pp. 375-421.

Taylor, Alan M., 2001, "Potential Pitfalls for the Purchasing Power Parity Puzzle? Sampling and Specification Biases in Mean-Reversion Testes of the Law of One Price," *Econometrica*, Vol. 69, No. 2, March, pp. 473-498.

Zanello, Alessandro, and Dominique Desruelles, 1997, "A Primer on the IMF's Information Notice System," IMF Working Paper 97/71 (Washington: International Monetary Fund).