8

What Do We Know about Recent Exchange Rate Models? In-Sample Fit and Out-of-Sample Performance Evaluated

Yin-Wong Cheung, Menzie D. Chinn, and Antonio Garcia Pascual

In contrast to the intellectual ferment that followed the collapse of the Bretton Woods era, the 1990s were marked by a relative paucity of new *empirical* models of exchange rates. The sticky-price monetary model of Dornbusch and Frankel remained the workhorse of policy-oriented analyses of exchange rate fluctuations among the developed economies. However, while no completely new models were developed, several approaches gained increased prominence. Some of these approaches were inspired by new empirical findings, such as the correlation between net foreign asset positions and real exchange rates. Others, such as those based on productivity differences, were grounded in an older theoretical literature but given new respectability by the new international macroeconomics (Obstfeld and Rogoff 1996) literature. None of the empirical models, however, were subjected to rigorous examination of the sort that Frankel (1979) and Meese and Rogoff (1983a, b) conducted in their seminal works.

Consequently, instead of re-examining the usual suspects—the flexible price monetary model, purchasing power parity, and the interest differential[1]—we vary the set of performance criteria and expand the set to include the mean squared error, and the direction-of-change statistic. The later dimension is potentially more important from a market timing perspective, besides serving as another indicator of forecast attributes.

To summarize, in this study, we compare exchange rate models along several dimensions:

• Four models are compared against the random walk. Only one of the structural models—the benchmark sticky-price monetary model of Dornbusch and Frankel—has been the subject of previous systematic analyses. The other models include one incorporating productivity differentials in a fashion consistent with a Balassa-Samuelson formulation, an interest rate parity specification, and a representative behavioral equilibrium exchange rate model.

• The behavior of US dollar-based exchange rates of the Canadian dollar, British pound, German mark, Swiss franc, and Japanese yen are examined. We also examine the corresponding yen-based rates to ensure that our conclusions are not driven by dollar specific results.

• The models are estimated in two ways: in first-difference and error correction specifications.

• In sample fit is assessed in terms of how well the coefficient estimates conform to theoretical priors.

• Forecasting performance is evaluated at several horizons (1-, 4- and 20-quarter horizons), for a recent period not previously examined (post-1992).

• We augment the conventional metrics with a direction-of-change statistic and the "consistency" criterion of Cheung and Chinn (1998).

In accordance with previous studies, we find that no model consistently outperforms a random walk according to the mean squared error criterion at short horizons. However, at the longest horizon we find that the proportion of times the structural models incorporating long-run relationships outperform a random walk is more than would be expected if the outcomes were merely random. Using a 10 percent significance level, a random walk is outperformed 17 percent of the time along a MSE dimension and 27 percent along a direction of change dimension.

In terms of the "consistency" test of Cheung and Chinn (1998), we obtain slightly less positive results. The actual and forecasted rates are cointegrated more often than would occur by chance for all the models. While in many of these cases of cointegration, the condition of unitary elasticity of expectations is rejected; only about 5 percent fulfill all the conditions of the consistency criteria.

We conclude that the question of exchange rate predictability remains unresolved. In particular, while the oft-used mean squared error criterion provides a dismal perspective, criteria other than the conventional ones suggest that structural exchange rate models have some usefulness. Furthermore, structural models incorporating restrictions at long horizons tend to outperform random walk specifications.

8.1 Theoretical Models

The universe of empirical models that have been examined over the floating rate period is enormous. Consequently any evaluation of these models must necessarily be selective. The models we have selected are prominent in the economic and policy literature, and readily implementable and replicable. To our knowledge, with the exception of the sticky-price model, they have also not previously been evaluated in a systematic fashion. We use the random walk model as our benchmark naive model, in line with previous work, but we also select one model—the Dornbusch (1976) and Frankel (1979) model—as a representative of the 1970s vintage models. The sticky-price monetary model can be expressed as follows:

$$s_t = \beta_0 + \beta_1 \hat{m}_t + \beta_2 \hat{y}_t + \beta_3 \hat{i}_t + \beta_4 \hat{\pi}_t + u_t, \tag{1}$$

where s is exchange rate in log, m is log money, y is log real GDP, i and π are the interest and inflation rate, respectively, the caret (ˆ) denotes the intercountry difference, and u_t is an error term.

The characteristics of this model are well known, so we will not devote time to discuss the theory behind the equation. We will observe, however, that the list of variables included in (1) encompasses those employed in the flexible price version of the monetary model, as well as the micro-based general equilibrium models of Stockman (1980) and Lucas (1982).

Second, we assess models that are in the Balassa-Samuelson vein, in that they accord a central role to productivity differentials in explaining movements in real, and hence also nominal, exchange rates (see Chinn 1997). Such models drop the purchasing power parity assumption for broad price indexes and allow the real exchange rate to depend on the relative price of nontradables, itself a function of productivity (z) differentials. A generic productivity differential exchange rate equation is

$$s_t = \beta_0 + \beta_1 \hat{m} + \beta_2 \hat{y} + \beta_3 \hat{i} + \beta_5 \hat{z}_t + u_t. \tag{2}$$

The third set of models we examine we term the "behavioral equilibrium exchange rate" (BEER) approach. We investigate this model as a proxy for a diverse set of models that incorporate a number of familiar relationships. A typical specification is

$$s_t = \beta_0 + \hat{p}_t + \beta_6 \hat{\omega}_t + \beta_7 \hat{r}_t + \beta_8 \hat{g}debt_t + \beta_9 tot_t + \beta_{10} nfa_t + u_t, \tag{3}$$

where p is the log price level (CPI), ω is the relative price of nontradables, r is the real interest rate, *gdebt* is the government debt to GDP ratio, *tot* is the log terms of trade, and *nfa* is the net foreign asset ratio. A unitary coefficient is imposed on $\hat{p}_t$. This specification can be thought of as incorporating the Balassa-Samuelson effect, the real interest differential model, an exchange risk premium associated with government debt stocks, and additional portfolio balance effects arising from the net foreign asset position of the economy.[2] Evaluation of this model can shed light on a number of very closely related approaches, including the macroeconomic framework of the IMF (Isard et al. 2001) and Stein's NATREX (Stein 1999). The empirical determinants in both approaches overlap with those of the specification in equation (3).

Models based on this framework have been the predominant approach to determining the level at which currencies will gravitate to over some intermediate horizon, especially in the context of policy issues. For instance, the behavioral equilibrium exchange rate approach is the model that is most used to determine the long-term value of the euro.

The final specification assessed is not a model per se; rather it is an arbitrage relationship—uncovered interest rate parity:

$$s_{t+k} - s_t = \hat{i}_{t,k}, \tag{4}$$

where $\hat{i}_{t,k}$ is the interest rate of maturity k. Unlike the other specifications, this relation does not need to be estimated in order to generate predictions.

Interest rate parity at long horizons has recently gathered empirical support (Alexius 2001; Chinn and Meredith 2002), in contrast to the disappointing results at the shorter horizons. MacDonald and Nagayasu (2000) have also demonstrated that long-run interest rates can predict exchange rate levels. On the basis of these findings, we anticipate that this specification will perform better at the longer horizons than at the shorter.[3]

8.2 Data and Full-Sample Estimation

8.2.1 Data

The analysis uses quarterly data for the United States, Canada, the United Kingdom, Japan, Germany, and Switzerland over the 1973:2 to

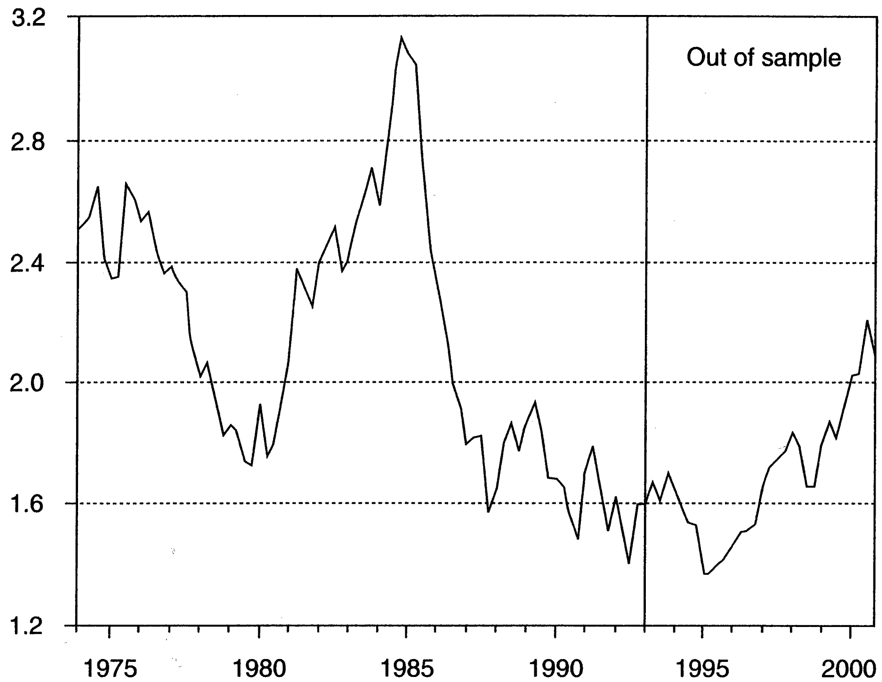

Figure 8.1
German mark–US dollar exchange rate

2000:4 period. The exchange rate, money, price and income variables are drawn primarily from the IMF's *International Financial Statistics*. The productivity data were obtained from the Bank for International Settlements, while the interest rates used to conduct the interest rate parity forecasts are essentially the same as those used in Chinn and Meredith (2002). See appendix A for a more detailed description.

The out-of-sample period used to assess model performance is 1993:1 to 2000:4. Figures 8.1 and 8.2 depict, respectively, the dollar based German mark and yen exchange rates, with the vertical line indicating the beginning of the out-of-sample period. The out-of-sample period spans a period of dollar depreciation and then sustained appreciation.[4]

8.2.2 Full-Sample Estimation

Two specifications of the theoretical models were estimated: (1) an error correction specification, and (2) a first-differences specification. Since implementation of the error correction specification is relatively involved, we will address the first-difference specification to begin

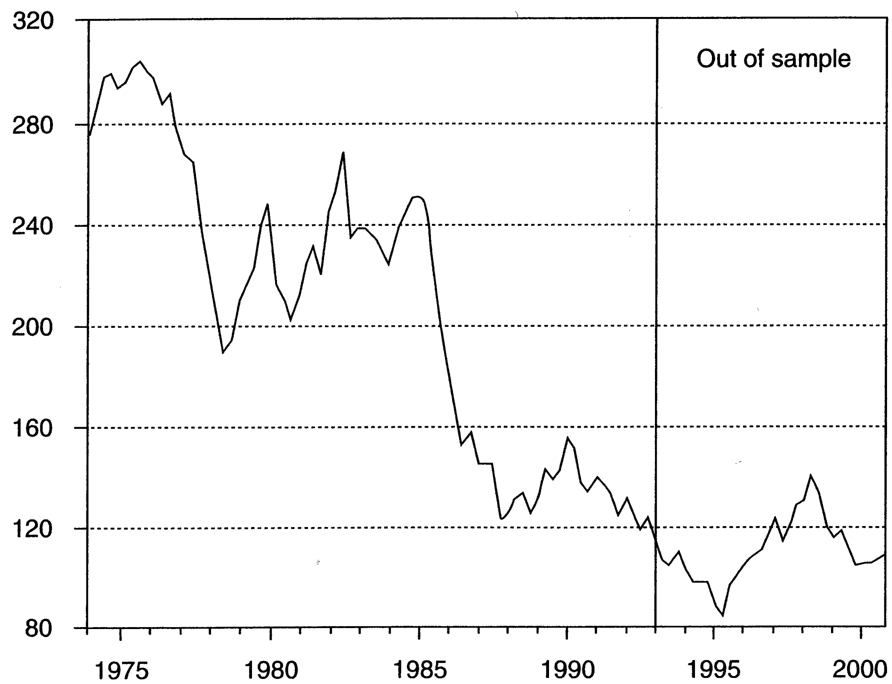

Figure 8.2
Japanese yen–US dollar exchange rate

with. Consider the general expression for the relationship between the exchange rate and fundamentals:

$$s_t = X_t\Gamma + u_t, \tag{5}$$

where X_t is a vector of fundamental variables under consideration. The first-difference specification involves the following regression:

$$\Delta s_t = \Delta X_t\Gamma + u_t. \tag{6}$$

These estimates are then used to generate forecasts one and many quarters ahead. Since these exchange rate models imply joint determination of all variables in the equations, it makes sense to apply instrumental variables. However, previous experience indicates that the gains in consistency are far outweighed by the loss in efficiency, in terms of prediction (Chinn and Meese 1995). Hence we rely solely on OLS.

One exception to this general rule is the UIP model. In this case the arbitrage condition implies a relationship between the change in the exchange rate and the level of the interest rate differential. Since no long-run condition is implied, we simply estimate the UIP relationship as stated in equation (4).

8.2.3 *Empirical Results*

The results of estimating the sticky-price monetary model in levels are presented in panel A of table 8.1. Using the 5 percent asymptotic critical value, we find that there is evidence of cointegration for the dollar-based exchange rates for all currencies save one. The German mark stands out as a case where it is difficult to obtain evidence of cointegration; we suspect that this is largely because of the breaks in the series for both money and income associated with the German reunification. The evidence for cointegration is more attenuated when the finite sample critical values (Cheung and Lai 1993) are used. Then only the Canadian dollar and yen have some mixed evidence in favor of cointegration.

This ambiguity is useful to recall when evaluating the estimates for the British sterling; the coefficient estimates do not conform to those theoretically implied by the model, as the coefficients of money, inflation and income are all incorrectly signed (although the latter two are insignificantly so). Only the interest rate coefficient is significant and correctly signed. In contrast, both the yen and franc broadly conform to the monetary model. Money and inflation are correctly signed, while interest rates enter in correctly only for the yen. Finally, the Canadian dollar presents some interesting results. The coefficients are largely in line with the monetary model, although the income coefficient is wrongly signed, with economic and statistical significance.

The use of the first-difference specification is justified when there is a failure to find evidence of cointegration (the German mark), or alternatively one suspects that estimates of the long-run coefficients are insufficiently precisely estimated to yield useful estimates. In panel B of table 8.1, the results from the first-difference specification are reported. A general finding is that the coefficients do not typically enter with both statistical significance and correct sign. One partial exception is the interest differential coefficient. Higher interest rates, if all else constant is held constant, appear to appreciate the currency in four of five cases, although the yen–dollar rate estimate is not statistically significant. The British sterling–dollar rate estimate is positive (while the inflation rate coefficient is not statistically significant), a finding that is more consistent with a flexible price monetary model than a sticky-price one. Otherwise, the fit does not appear particularly good.

These mixed results are suggestive of alternative approaches; the first we examine is the productivity-based model. Our interpretation

Table 8.1
Full-sample estimates of sticky-price model

	Sign	BP/$	Can$/$	DM/$	SF/$	Yen/$
A. In levels[a]						
Cointegration (asy)		1, 1	3, 1	0, 0	1, 1	1, 1
Cointegration (fs)		0, 0	1, 0	0, 0	0, 0	0, 1
Money	[+]	−2.89*	1.10*	2.14*	3.61*	1.29
		(1.01)	(0.25)	(0.74)	(0.74)	(0.96)
Income	[−]	1.64	9.70*	0.93	−1.10	0.77
		(3.94)	(1.87)	(1.87)	(1.72)	(1.97)
Interest rate	[−]	−19.49*	−6.44*	−5.86	2.09	−17.11*
		(4.01)	(3.27)	(4.14)	(5.73)	(4.72)
Inflation rate	[+]	−7.11	10.74*	24.29*	40.96*	26.56*
		(4.60)	(3.11)	(4.27)	(6.79)	(4.03)
B. In first differences[b]						
Money	[+]	−0.21	−0.00	0.16	−0.02	0.44
		(0.12)	(0.06)	(0.22)	(0.14)	(0.24)
Income	[−]	−2.02*	−0.48	−0.51	0.59	−0.00
		(0.42)	(0.29)	(0.43)	(0.52)	(0.39)
Interest rate	[−]	0.83*	−0.42*	−0.91*	−0.82*	−0.28
		(0.41)	(0.10)	(0.45)	(0.37)	(0.33)
Inflation rate	[+]	−0.15	−0.07	1.26	1.29	0.32
		(0.48)	(0.20)	(1.09)	(0.81)	(0.44)

Note: "Sign" indicates coefficient sign implied by theoretical model. * indicates significantly different from zero at the 5% marginal significance level. Estimates for DM include shift and impulse dummies for German monetary and economic unification.
a. Long-run cointegrating estimates from Johansen procedure (standard errors in parentheses), where the VECM includes two lags of first differences. The number of cointegrating vectors is implied by the trace and maximal eigenvalue statistics, using the 5% marginal significance level; "asy" denotes asymptotic critical values and "fs" denotes finite sample critical values of Cheung and Lai (1993) that are used.
b. OLS estimates (Newey-West standard errors in parentheses, truncation lag = 4).

of the model simply augments the monetary model with a productivity variable. The results for this model are presented in table 8.2. From the asymptotic critical values, the evidence of cointegration in panel A of table 8.2 is comparable to that reported in panel A of table 8.1. For both the British sterling and Canadian dollar, there is evidence of multiple cointegrating vectors. However, in using the finite sample critical values, we find that the number of implied vectors drops to one (or zero) in this case.

In all cases the interest coefficient is correctly signed, and significant in most cases. Furthermore the money and inflation variables are correctly signed in most cases. The productivity coefficients are significant and consistent with the productivity in three cases—the Swiss franc, German mark, and yen. The latter two currencies have previously been found to be influenced by productivity trends.[5]

Estimates of the first-difference specifications do not yield appreciably better results than their sticky-price counterparts. Interest differentials tend to be important, once again, while productivity fails to evidence any significant impact for three of five rates. To the extent that one thinks that productivity is a slowly trending variable that influences the real exchange rate over long periods, this result is unsurprising. While this variable has the correct sign for the German mark–dollar rate, it has the opposite for the sterling–dollar rate.

The Canadian dollar appears to be as resilient to being modeled using this productivity specification as the others. Chen and Rogoff (2002) have asserted that the Canadian dollar is mostly determined by commodity prices; hence it is not surprising that both models fail to have any predictive content.

The BEER model results are presented in table 8.3. There are no estimates for the Swiss franc and the yen because we lack quarterly data on government debt and net foreign assets. Overall, the results are not uniformly supportive of the BEER approach.[6] Although there are some instances of correctly signed coefficients, none show up correctly signed across all three currencies. Moving to a first-difference specification does not improve the results. Besides those on the relative price and real interest rate differentials, very few coefficient estimates are in line with model predictions. For the DM/$ rate, the real interest rate and debt variables possess the correctly signed coefficients, as do the relative price and net foreign assets for the Canadian dollar, but these appear to be isolated instances.[7]

Table 8.2
Full-sample estimates of productivity model

	Sign	BP/$	Can$/$	DM/$	SF/$	Yen/$
A. In levels[a]						
Cointegration (asy)		1, 2	2, 2	0, 0	1, 1	1, 1
Cointegration (fs)		0, 0	1, 0	0, 0	0, 0	0, 1
Money	[+]	0.97*	6.81*	0.62*	2.00*	0.18
		(0.47)	(1.45)	(0.33)	(0.30)	(0.54)
Income	[−]	−4.11*	25.76*	−0.68	−1.04	2.77*
		(1.23)	(6.62)	(0.81)	(0.76)	(1.29)
Interest rate	[−]	−10.63*	−34.53*	−9.35*	3.67	−12.07*
		(1.65)	(11.16)	(2.57)	(2.54)	(2.67)
Inflation rate	[+]	9.86*	70.63*	9.18*	15.36*	12.09*
		(1.63)	(12.00)	(1.85)	2.79	(2.49)
Productivity	[−]	3.56*	16.78*	−5.66*	−4.43*	−2.65*
		(0.68)	(5.60)	(1.11)	(1.46)	(0.76)
B. In first differences[b]						
Money	[+]	0.40*	−0.00	0.16	−0.01	0.43
		(0.16)	(0.06)	(0.22)	(0.14)	(0.24)
Income	[−]	−1.59*	−0.47	−0.51	0.70	0.00
		(0.39)	(0.29)	(0.43)	(0.51)	(0.40)
Interest rate	[−]	−0.57	−0.42*	−0.91*	−0.82*	−0.28
		(0.46)	(0.10)	(0.45)	(0.41)	(0.32)
Inflation rate	[+]	1.10*	−0.08	1.26	1.19	0.37
		(0.50)	(0.20)	(1.09)	(0.81)	(0.45)
Productivity	[−]	1.11*	−0.03	−5.66*	−0.25	−0.32
		(0.21)	(0.15)	(1.11)	(0.21)	(0.31)

Note: "Sign" indicates coefficient sign implied by theoretical model. * indicates significantly different from zero at the 5% marginal significance level. Estimates for DM include shift and impulse dummies for German monetary and economic unification.
a. Long-run cointegrating estimates from Johansen procedure (standard errors in parentheses), where the VECM includes two lags of first differences. The number of cointegrating vectors is implied by the trace and maximal eigenvalue statistics, using the 5% marginal significance level; "asy" denotes asymptotic critical values and "fs" denotes finite sample critical values of Cheung and Lai (1993) that are used.
b. OLS estimates (Newey-West standard errors in parentheses, truncation lag = 4).

Table 8.3
Full-sample estimates of BEER model

	Sign	BP/$	Can$/$	DM/$
A. In levels[a]				
Cointegration (asy)		2, 2	4, 2	1, 1
Cointegration (fs)		1, 2	2, 1	0, 0
Relative price	[−]	1.27*	−1.05*	−9.38*
		(0.38)	(0.34)	(1.36)
Real interest rate	[−]	−3.13*	2.03*	−2.37
		(1.07)	(0.91)	(2.09)
Debt	[+]	−1.06*	−2.62*	0.04
		(0.30)	(0.51)	(0.72)
Terms of trade	[−]	−0.92	0.75*	−0.13
		(0.82)	(0.24)	(1.04)
Net foreign assets	[−]	5.65*	−1.39*	−4.88*
		(0.56)	(0.40)	(0.76)
B. In first differences[b]				
Relative price	[−]	−0.55	−0.44*	−0.38
		(0.56)	(0.17)	(0.59)
Real interest rate	[−]	−0.17	−0.15	−1.04*
		(0.16)	(0.11)	(0.34)
Debt	[+]	−0.38	0.18	1.52*
		(0.27)	(0.22)	(0.64)
Terms of trade	[−]	0.09	0.02	0.59*
		(0.31)	(0.06)	(0.27)
Net foreign assets	[−]	2.61*	−1.19*	3.14*
		(0.49)	(0.25)	(0.72)

Note: "Sign" indicates coefficient sign implied by theoretical model. * indicates significantly different from zero at the 5% marginal significance level. Estimates for DM include shift and impulse dummies for German monetary and economic unification.

a. Long-run cointegrating estimates from Johansen procedure (standard errors in parentheses), where the VECM includes 2 lags of first differences (4 lags for DM). The number of cointegrating vectors is implied by the trace and maximal eigenvalue statistics, using the 5% marginal significance level; "asy" denotes asymptotic critical values and "fs" denotes finite sample critical values of Cheung and Lai (1993) that are used.

b. OLS estimates (Newey-West standard errors in parentheses, truncation lag = 4).

Table 8.4
Uncovered interest parity estimates

		BP/$	Can$/$	DM/$	SF/$	Yen/$
Horizon						
3 month		−2.19*	−0.48*	−0.70	−1.28*	−2.99*
		(1.08)	(0.51)	(1.09)	(1.04)	(0.96)
	Adj R^2	0.04	−0.00	−0.01	0.01	0.06
	SER	0.21	0.08	0.26	0.29	0.28
1 year		−1.42*	−0.61*	−0.58*	−1.05*	−2.60*
		(0.99)	(0.49)	(0.66)	(0.52)	(0.69)
	Adj R^2	0.06	0.03	0.00	0.04	0.17
	SER	0.11	0.04	0.14	0.14	0.13
5 year		0.44	0.24	0.52	−1.18*	1.19
		(0.36)	(0.47)	(0.75)	(0.97)	(0.38)
	Adj R^2	0.02	−0.00	0.02	0.04	0.13
	SER	0.04	0.02	0.06	0.04	0.05

Note: OLS estimates (Newey-West standard errors in parentheses, truncation lag = $k - 1$). SER is standard error of regression. * indicates significantly different from *unity* at the 5 percent marginal significance level.

Although we do not use estimated equations to conduct the forecasting of the UIP model, it is informative to consider how well the data conform to the UIP relationship. As is well known, at short horizons, the evidence in favor of UIP is lacking.[8] The results of estimating equation (4) are reported in table 8.4. Consistent with Chinn and Meredith (2002), the short-horizon data (1 quarter and 4 quarter maturities) provide almost uniformly negative coefficient estimates, in contradiction to the implication of the UIP hypothesis. At the five-year horizon, the results are substantially different for all cases, save the Swiss franc. Now all the coefficients are positive; moreover in no case except the franc is the coefficient estimate significantly different from the theoretically implied value of unity.

8.3 Forecast Comparison

8.3.1 Estimation and Forecasting

We adopt the convention in the empirical exchange rate modeling literature of implementing "rolling regressions." That is, estimates are applied over a given data sample, out-of-sample forecasts produced, then the sample is moved up, or "rolled" forward one observation before the procedure is repeated. This process continues until all the out-of-sample observations are exhausted. This procedure is selected over recursive estimation because it is more in line with previous work, including the original Meese and Rogoff paper. Moreover the power of the test is kept constant as the sample size over which the estimation occurs is fixed, rather than increasing as it does in the recursive framework.

The error correction estimation involves a two-step procedure. In the first step, the long-run cointegrating relation implied by (5) is identified using the Johansen procedure, as described in section 8.2. The estimated cointegrating vector ($\tilde{\Gamma}$) is incorporated into the error correction term, and the resulting equation

$$s_t - s_{t-k} = \delta_0 + \delta_1(s_{t-k} - X_{t-k}\tilde{\Gamma}) + u_t \tag{7}$$

is estimated via OLS. Equation (7) can be thought of as an error correction model stripped of the short-run dynamics. A similar approach was used in Mark (1995) and Chinn and Meese (1995), except for the fact that, in those two cases, the cointegrating vector was imposed a priori.

One key difference between our implementation of the error correction specification and that undertaken in some other studies involves the treatment of the cointegrating vector. In some other prominent studies (MacDonald and Taylor 1994) the cointegrating relationship is estimated over the entire sample, and then out-of-sample forecasting undertaken, where the short-run dynamics are treated as time varying *but the long-run relationship is not*. While there are good reasons for adopting this approach—in particular, one wants to use as much information as possible to obtain estimates of the cointegrating relationships—the asymmetry in the estimation approach is troublesome, and makes it difficult to distinguish quasi–ex ante forecasts from true ex ante forecasts. Consequently our estimates of the long-run cointegrating relationship vary as the data window moves.

It is also useful to stress the difference between the error correction specification forecasts and the first-difference specification forecasts. In the latter, ex post values of the right-hand side variables are used to generate the predicted exchange rate change. In the former, contemporaneous values of the right-hand side variables are not necessary, and the error correction predictions are true ex ante forecasts. Hence we are affording the first-difference specifications a tremendous informational advantage in forecasting.[9]

8.3.2 *Forecast Comparison*

To evaluate the forecasting accuracy of the different structural models, the ratio between the mean squared error (MSE) of the structural models and a driftless random walk is used. A value smaller (larger) than one indicates a better performance of the structural model (random walk). We also explicitly test the null hypothesis of no difference in the accuracy of the two competing forecasts (structural model vs. driftless random walk). In particular, we use the Diebold-Mariano statistic (Diebold and Mariano 1995), which is defined as the ratio between the sample mean loss differential and an estimate of its standard error; this ratio is asymptotically distributed as a standard normal.[10] The loss differential is defined as the difference between the squared forecast error of the structural models and that of the random walk. A consistent estimate of the standard deviation can be constructed from a weighted sum of the available sample autocovariances of the loss differential vector. Following Andrews (1991), a quadratic spectral kernel is employed, together with a data-dependent bandwidth selection procedure.[11]

We also examine the predictive power of the various models along different dimensions. One might be tempted to conclude that we are merely changing the well-established "rules of the game" by doing so. However, there are very good reasons to use other evaluation criteria. First, there is the intuitively appealing rationale that minimizing the mean squared error (or relatedly mean absolute error) may not be important from an economic standpoint. A less pedestrian motivation is that the typical mean squared error criterion may miss out on important aspects of predictions, especially at long horizons. Christoffersen and Diebold (1998) point out that the standard mean squared error criterion indicates no improvement of predictions that take into account cointegrating relationships vis à vis univariate predictions. But surely any reasonable criteria would put some weight on the tendency for predictions from cointegrated systems to "hang together."

Hence, our first alternative evaluation metric for the relative forecast performance of the structural models is the direction-of-change statistic, which is computed as the number of correct predictions of the direction of change over the total number of predictions. A value above (below) 50 percent indicates a better (worse) forecasting performance than a naive model that predicts the exchange rate has an equal chance to go up or down. Again, Diebold and Mariano (1995) provide a test statistic for the null of no forecasting performance of the structural model. The statistic follows a binomial distribution, and its studentized version is asymptotically distributed as a standard normal. Not only does the direction-of-change statistic constitute an alternative metric, it is also an approximate measure of profitability. We have in mind here tests for market-timing ability (Cumby and Modest 1987).[12]

The third metric we used to evaluate forecast performance is the consistency criterion proposed in Cheung and Chinn (1998). This metric focuses on the time series properties of the forecast. The forecast of a given spot exchange rate is labeled as consistent if (1) the two series have the same order of integration, (2) they are cointegrated, and (3) the cointegration vector satisfies the unitary elasticity of expectations condition. Loosely speaking, a forecast is consistent if it moves in tandem with the spot exchange rate in the long run. Cheung and Chinn (1998) provide a more detailed discussion on the consistency criterion and its implementation.

8.4 Comparing the Forecast Performance

8.4.1 The MSE Criterion

The comparison of forecasting performance based on MSE ratios is summarized in table 8.5. The table contains MSE ratios and the *p*-values from five dollar-based currency pairs, four structural models, the error correction and first-difference specifications, and three forecasting horizons. Every cell in the table has two entries. The first one is the MSE ratio (the MSEs of a structural model to the random walk specification). The entry underneath the MSE ratio is the *p*-value of the hypothesis that the MSEs of the structural and random walk models are the same. Because of the lack of data, the behavioral equilibrium exchange rate model is not estimated for the dollar–Swiss franc, dollar–yen exchange rates, and all yen-based exchange rates. Altogether there are 153 MSE ratios. Of these 153 ratios, 90 are computed from the error correction specification and 63 from the first-difference one.

Note that in the tables only "error correction specification" entries are reported for the interest rate parity model. This model is not

Table 8.5
MSE ratios from the dollar-based and yen-based exchange rates

Specification	Horizon	S–P	IRP	PROD	BEER	S–P	IRP	PROD
Panel A		*BP/$*				*BP/yen*		
ECM	1	1.0469	1.0096	1.0795	1.1597	0.9709	1.0421	1.0266
		0.3343	0.6613	0.1827	0.0909	0.5831	0.6269	0.7905
	4	1.0870	0.7696	1.1974	1.5255	1.1466	1.0008	1.4142
		0.5163	0.3379	0.2571	0.0001	0.3889	0.9975	0.3171
	20	0.4949	0.9810	0.7285	1.2841	1.2020	0.7611	1.7493
		0.1329	0.9581	0.5225	0.4016	0.1302	0.5795	0.0295
FD	1	1.0357		1.1678	1.8876	0.9655		1.0000
		0.7095		0.4255	0.0092	0.7175		1.0000
	4	1.2691		1.3830	3.7789	1.1191		1.1114
		0.3260		0.1038	0.0004	0.6543		0.6886
	20	6.0121		2.2029	18.370	4.5445		4.7881
		0.0000		0.0021	0.0000	0.0000		0.0000
Panel B		*CAN$/$*				*CAN$/yen*		
ECM	1	1.0365	1.0849	1.0537	1.2644	0.9617	1.0096	0.9948
		0.3991	0.0316	0.3994	0.0018	0.2537	0.8710	0.9269
	4	1.0681	1.0123	1.1194	1.5570	0.9716	1.0045	1.1185
		0.2531	0.9592	0.2015	0.0002	0.7037	0.9814	0.4038
	20	0.6339	0.1881	1.0204	1.7609	1.1694	0.6462	4.8827
		0.0248	0.0001	0.9276	0.0302	0.2747	0.4125	0.1130
FD	1	1.0474		1.0842	0.5424	1.0106		0.9827
		0.6214		0.3971	0.1544	0.9144		0.8456
	4	0.9866		1.0519	1.2907	1.1578		1.1663
		0.9531		0.8232	0.5046	0.5751		0.5827

	20	0.2051		0.2937	4.7274	12.181		12.12
		0.0318		0.1018	0.0000	0.0000		0.0000
Panel C		*DM/$*				*DM/yen*		
ECM	1	0.9990	1.0705	0.9867	1.0810	1.0447	0.9662	0.9983
		0.5440	0.0383	0.5858	0.1951	0.3200	0.4790	0.0528
	4	0.9967	1.2090	0.9298	1.0484	1.0006	0.8571	1.0003
		0.5861	0.0694	0.2956	0.3109	0.5779	0.3238	0.7265
	20	1.0242	1.0073	1.0410	0.6299	1.0034	0.5485	0.9921
		0.0004	0.9354	0.0030	0.0891	0.6003	0.0480	0.1126
FD	1	1.0354		1.1208	0.4649	1.0227		1.0060
		0.3020		0.1959	0.0009	0.7181		0.9219
	4	1.1184		1.1782	0.3331	1.0859		1.0045
		0.2019		0.0029	0.0059	0.1849		0.9625
	20	2.0817		1.9828	1.2906	0.9521		0.8569
		1.1915		0.0000	0.2550	0.7217		0.3572
Panel D		*SF/$*				*SF/yen*		
ECM	1	0.9784	1.1101	1.1200		0.9961	0.9985	1.0515
		0.7773	0.0692	0.1614		0.9333	0.9522	0.2892
	4	0.8864	1.2871	1.0409		1.0627	0.9276	1.0140
		0.4152	0.0689	0.7438		0.2595	0.3983	0.7786
	20	1.2873	1.4894	0.9651		0.8331	0.9031	0.9216
		0.1209	0.0000	0.8684		0.2925	0.4856	0.1019
FD	1	1.3115		1.3891		0.9350		0.9338
		0.1641		0.1734		0.1643		0.1765
	4	1.6856		1.8437		1.0114		0.9666
		0.0774		0.0713		0.8595		0.7366
	20	5.6773		5.9918		0.9208		0.8852
		0.0000		0.0000		0.0000		0.0001

Table 8.5
(continued)

Specification	Horizon	S–P	IRP	PROD	BEER	S–P	IRP	PROD
Panel E		*Yen/$*						
ECM	1	0.9821	1.0681	0.9973				
		0.8799	0.2979	0.9647				
	4	0.8870	1.2047	0.9460				
		0.6214	0.2862	0.7343				
	20	0.8643	0.9824	0.8500				
		0.4299	0.9661	0.3856				
FD	1	1.0022		0.9456				
		0.9840		0.4427				
	4	1.0240		1.0624				
		0.8207		0.5342				
	20	2.7132		2.2586				
		0.0000		0.0001				

Note: The results are based on dollar-based and yen-based exchange rates and their forecasts. Each cell has two entries. The first is the MSE ratio (the MSEs of a structural model to the random walk specification). The entry underneath the MSE ratio is the p-value of the hypothesis that the MSEs of the structural and random walk models are the same (Diebold and Mariano 1995). The notation used in the table is ECM: error correction specification; FD: first-difference specification; S–P: sticky-price model; IRP: interest rate parity model; PROD: productivity differential model; and BEER: behavioral equilibrium exchange rate model. The forecasting horizons (in quarters) are listed under the heading "horizon." The forecasting period is 1993:1 to 2000:4. Due to data unavailability, the BEER model was not estimated for the Japanese yen and Swiss franc.

estimated; rather the predicted spot rate is calculated using the uncovered interest parity condition. To the extent that long-term interest rates can be considered the error correction term, we believe this categorization is most appropriate.

Overall, the MSE results are not favorable to the structural models. Of the 153 MSE ratios, 109 are not significant (at the 10 percent significance level), and 44 are significant. That is, for the majority of the cases one cannot differentiate the forecasting performance between a structural model and a random walk model. For the 44 significant cases, there are 32 cases in which the random walk model is significantly better than the competing structural models and only 11 cases in which the opposite is true. As 10 percent is the size of the test and 12 cases constitute less than 10 percent of the total of 153 cases, the empirical evidence can hardly be interpreted as supportive of the superior forecasting performance of the structural models. One caveat is necessary, however. When one restricts attention to the long-horizon forecasts, it turns out that those incorporating long-run restrictions outperform a random walk more often than would be expected to occur randomly: five out of 30 cases, or 17 percent, using a 10 percent significance level.

Inspecting the MSE ratios, one does not observe many consistent patterns, in terms of outperformance. It appears that the BEER model does not do particularly well except for the DM/$ rate. The interest rate parity model tends to do better at the 20-quarter horizon than at the 1- and 4-quarter horizons—a result consistent with the well-known bias in forward rates at short horizons.

In accordance with the existing literature, our results are supportive of the assertion that it is very difficult to find forecasts from a structural model that can consistently beat the random walk model using the MSE criterion. The current exercise further strengthens the assertion as it covers both dollar- and yen-based exchange rates and some structural models that have not been extensively studied before.

8.4.2 *The Direction-of-Change Criterion*

Table 8.6 reports the proportion of forecasts that correctly predicts the direction of the exchange rate movement and, underneath these sample proportions, the *p*-values for the hypothesis that the reported proportion is significantly different from 0.5. When the proportion statistic is significantly larger than 0.5, the forecast is said to have the ability to predict the direct of change. On the other hand, if the statistic is

Table 8.6
Direction-of-change statistics from the dollar-based and yen-based exchange rates

Specification	Horizon	S–P	IRP	PROD	BEER	S–P	IRP	PROD
Panel A		*BP/$*				*BP/yen*		
ECM	1	0.5312	0.4849	0.5313	0.4062	0.5625	0.4546	0.6563
		0.7236	0.8618	0.7237	0.2888	0.4795	0.6015	0.0771
	4	0.5862	0.5455	0.4483	0.3448	0.5517	0.6364	0.5517
		0.3531	0.6015	0.5775	0.0946	0.5774	0.1172	0.5775
	20	0.8461	0.7273	0.7692	0.3846	0.5384	0.5758	0.2308
		0.0125	0.0090	0.0522	0.4053	0.7815	0.3841	0.0522
FD	1	0.5937		0.4688	0.4062	0.5937		0.4375
		0.2888		0.7237	0.2888	0.2888		0.4795
	4	0.5517		0.5172	0.3448	0.6551		0.5862
		0.5774		0.8527	0.0946	0.0946		0.3532
	20	0.3076		0.1539	0.3076	0.0000		0.0000
		0.1655		0.0126	0.1655	0.0000		0.0000
Panel B		*CAN$/$*				*CAN$/yen*		
ECM	1	0.4062	0.3939	0.3438	0.3125	0.5937	0.4849	0.6250
		0.2888	0.2230	0.0771	0.0338	0.2888	0.8618	0.1573
	4	0.4827	0.4242	0.4828	0.1724	0.6206	0.5758	0.5172
		0.8526	0.3841	0.8527	0.0004	0.1936	0.3841	0.8527
	20	0.7692	1.0000	0.4615	0.0769	0.5384	0.7273	0.2308
		0.0522	0.0000	0.7815	0.0022	0.7815	0.0090	0.0522
FD	1	0.5312		0.5625	0.6250	0.5000		0.4375
		0.7236		0.4795	0.1573	1.0000		0.4795
	4	0.7586		0.7241	0.5862	0.5172		0.4828
		0.0053		0.0158	0.3531	0.8526		0.8527

	20	1.0000		1.0000	0.0000	0.3076		0.3077
		0.0000		0.0000	0.0000	0.1655		0.1655
Panel C		*DM/$*				*DM/yen*		
ECM	1	0.5000	0.3030	0.3750	0.5625	0.6250	0.5152	0.5000
		1.0000	0.0236	0.1573	0.4795	0.1573	0.8618	1.0000
	4	0.5517	0.3030	0.3103	0.4827	0.4137	0.6667	0.3793
		0.5774	0.0236	0.0411	0.8526	0.3531	0.0555	0.1937
	20	0.0769	0.5152	0.2308	0.2307	0.6923	0.8485	0.6154
		0.0022	0.8618	0.0522	0.0522	0.1655	0.0001	0.4054
FD	1	0.5000		0.4063	0.8125	0.4687		0.5000
		1.0000		0.2888	0.0004	0.7236		1.0000
	4	0.3448		0.2759	0.7931	0.4827		0.4483
		0.0946		0.0158	0.0015	0.8526		0.5775
	20	0.0769		0.0769	0.3076	0.3076		0.4615
		0.0022		0.0023	0.1655	0.1655		0.7815
Panel D		*SF/$*				*SF/yen*		
ECM	1	0.5625	0.3030	0.5625		0.6562	0.6061	0.4688
		0.4795	0.0236	0.4795		0.0771	0.2230	0.7237
	4	0.5517	0.3636	0.5517		0.4827	0.5758	0.4138
		0.5774	0.1172	0.5775		0.8526	0.3841	0.3532
	20	0.5384	0.4546	0.6923		0.5384	0.5000	0.6154
		0.7815	0.6698	0.1655		0.7815	1.0000	0.4054
FD	1	0.4062		0.4375		0.5937		0.6875
		0.2888		0.4795		0.2888		0.0339
	4	0.4137		0.5172		0.5517		0.5862
		0.3531		0.8527		0.5774		0.3532
	20	0.2307		0.2308		0.5384		0.6154
		0.0522		0.0522		0.7815		0.4054

Table 8.6
(continued)

Specification	Horizon	S–P	IRP	PROD	BEER	S–P	IRP	PROD
Panel E		*Yen/$*						
ECM	1	0.6562	0.3636	0.5625				
		0.0771	0.1172	0.4795				
	4	0.5517	0.5152	0.4828				
		0.5774	0.8618	0.8527				
	20	0.7692	0.5152	0.6923				
		0.0522	0.8618	0.1655				
FD	1	0.6875		0.6563				
		0.0338		0.0771				
	4	0.6551		0.6207				
		0.0946		0.1937				
	20	0.0000		0.0000				
		0.0000		0.0000				

Note: The table reports the proportion of forecasts that correctly predict the direction of the dollar-based and yen-based exchange rate movements. Under each direction-of-change statistic, the *p*-values for the hypothesis that the reported proportion is significantly different from 0.5 is listed. When the statistic is significantly larger than 0.5, the forecast is said to have the ability to predict the direct of change. If the statistic is significantly less than 0.5, the forecast tends to give the wrong direction of change. The notation used in the table is ECM: error correction specification; FD: first-difference specification; S–P: sticky-price model; IRP: interest rate parity model; PROD: productivity differential model; and BEER: behavioral equilibrium exchange rate model. The forecasting horizons (in quarters) are listed under the heading "horizon." The forecasting period is 1993:1 to 2000:4. Due to data unavailability, the BEER model was not estimated for the Japanese yen and Swiss franc.

significantly less than 0.5, the forecast tends to give the wrong direction of change. If a model consistently forecasts the direction of change incorrectly, traders can derive a potentially profitable trading rule by going against these forecasts. Thus, for trading purposes, information regarding the significance of "incorrect" prediction is as useful as the one of "correct" forecasts. However, in evaluating the ability of the model to describe exchange rate behavior, we separate the two cases.

There is mixed evidence on the ability of the structural models to correctly predict the direction of change. Among the 153 direction-of-change statistics, 23 (27) are significantly larger (less) than 0.5 at the 10 percent level. The occurrence of the significant outperformance cases is slightly higher (15 percent) than the one implied by the 10 percent level of the test. The results indicate that the structural model forecasts *can* correctly predict the direction of the change, although the proportion of cases where a random walk outperforms the competing models is higher than what one would expect if they occurred randomly.

Let us take a closer look at the incidences in which the forecasts are in the right direction. About half of the 23 cases are in the error correction category (12). Thus it is not clear if the error correction specification—which incorporates the empirical long-run relationship—is a better specification for the models under consideration.

Among the four models under consideration, the sticky-price model has the highest number (10) of forecasts that give the correct direction-of-change prediction (18 percent of these forecasts), while the interest rate parity model has the highest proportion of correct predictions (19 percent). Thus, at least on this count, the newer exchange rate models do not significantly edge out the "old fashioned" sticky-price model save perhaps the interest rate parity condition.

The cases of correct direction prediction appear to cluster at the long forecast horizon. The 20-quarter horizon accounts for 10 of the 23 cases while the 4-quarter and 1-quarter horizons have, respectively, 6 and 7 direction-of-change statistics that are significantly larger than 0.5. Since there have been few studies utilizing the direction-of-change statistic in similar contexts, it is difficult to make comparisons. Chinn and Meese (1995) apply the direction-of-change statistic to three-year horizons for three conventional models, and find that performance is largely currency-specific: the no-change prediction is outperformed in the case of the dollar–yen exchange rate, while all models are outperformed in the case of the dollar–sterling rate. In contrast, in our study at the 20-quarter horizon, the positive results appear to be concentrated in the

Table 8.7
Cointegration between exchange rates and their forecasts

Specification	Horizon	S–P	IRP	PROD	BEER	S–P	IRP	PROD
Panel A		*BP/$*				*BP/yen*		
ECM	1	2.12	14.25*	2.41	19.26*	8.70	5.35	5.06
	4	4.88	5.72	6.98	18.13*	26.54*	3.99	7.26
	20	9.69*	8.71	16.45*	6.54	6.27	5.25	4.02
FD	1	8.51		19.05*	7.66	15.85*		5.50
	4	8.30		7.32	4.53	5.34		5.38
	20	2.78		7.73	1.87	8.77		8.80
Panel B		*CAN$/$*				*CAN$/yen*		
ECM	1	6.74	6.03	3.41	6.32	6.94	6.59	7.77
	4	6.31	5.87	1.97	5.80	2.85	4.18	1.13
	20	6.58	7.03	8.96	4.53	7.22	9.51	4.29
FD	1	14.42*		15.60*	12.53*	15.07*		13.87*
	4	10.97*		7.22	6.22	5.64		4.20
	20	3.87		4.08	1.93	6.31		6.50
Panel C		*DM/$*				*DM/yen*		
ECM	1	2.78	11.18*	3.11	8.38	2.43	5.71	5.57
	4	4.74	11.72*	2.83	6.42	14.77*	4.39	9.50
	20	1.17	1.01	11.09*	3.30	7.12	13.97*	6.45
FD	1	14.99*		7.21	7.63	14.28*		16.37*
	4	8.37		7.36	3.02	42.41*		3.58
	20	1.37		1.20	5.17	5.55		5.84

Panel D		*SF/$*				*SF/yen*		
ECM	1	1.08	6.88	3.24	—	5.12	2.76	10.31*
	4	22.52*	6.84	34.23*	—	1.57	108.57*	3.25
	20	0.69	6.93	0.49	—	4.05	4.72	6.39
FD	1	2.73		1.02	—	4.40		47.89*
	4	5.21		1.65	—	1.81		3.10
	20	2.90		2.78	—	7.83		7.01
Panel E		*Yen/$*						
ECM	1	14.82*	12.20*	4.84	—			
	4	5.73	10.93*	5.33	—			
	20	14.99*	1.05	13.16*	—			
FD	1	20.48*		25.39*	—			
	4	5.61		42.86*	—			
	20	15.06*		13.17*	—			

Note: The table reports the Johansen maximum eigenvalue statistic for the null hypothesis that a dollar-based (or a yen-based) exchange rate and its forecast are not cointegrated. * indicates the 10% marginal significance level. Tests for the null of one cointegrating vector were also conducted, but in all cases the null was not rejected. The notation used in the table is ECM: error correction specification; FD: first-difference specification; S–P: sticky-price model; IRP: interest rate parity model; PROD: productivity differential model; and BEER: behavioral equilibrium exchange rate model. The forecasting horizons (in quarters) are listed under the heading "horizon." The forecasting period is 1993:1 to 2000:4. The dash indicates that the statistics were not generated due to unavailability of data.

yen–dollar and Canadian dollar–dollar rates.[13] It is interesting to note that the direction-of-change statistic works for the interest rate parity model almost only at the 20-quarter horizon, thus mirroring the MSE results. This pattern is entirely consistent with the finding that uncovered interest parity holds better at long horizons.

8.4.3 *The Consistency Criterion*

The consistency criterion only requires the forecast and actual realization comove one-to-one in the long run. One could argue that the criterion is less demanding than the MSE and direction-of-change metrics. Indeed, a forecast that satisfies the consistency criterion can (1) have a MSE larger than that of the random walk model, (2) have a direction-of-change statistic less than 0.5, or (3) generate forecast errors that are serially correlated. However, given the problems related to modeling, estimation, and data quality, the consistency criterion can be a more flexible way to evaluate a forecast. In assessing the consistency, we first test if the forecast and the realization are cointegrated.[14] If they are cointegrated, then we test if the cointegrating vector satisfies the $(1, -1)$ requirement. The cointegration results are reported in table 8.7. The test results for the $(1, -1)$ restriction are reported in table 8.8.

Thirty-eight of 153 cases reject the null hypothesis of no cointegration at the 10 percent significance level. Thus 25 percent of forecast series are cointegrated with the corresponding spot exchange rates. The error correction specification accounts for 20 of the 38 cointegrated cases and the first-difference specification accounts for the remaining 18 cases. There is no evidence that the error correction specification gives better forecasting performance than the first-difference specification.

Interestingly the sticky-price model garners the largest number of cointegrated cases. There are 54 forecast series generated under the sticky-price model. Fifteen of these 54 series (i.e., 28 percent) are cointegrated with the corresponding spot rates. Twenty-six percent of the interest rate parity and 24 percent of the productivity model are cointegrated with the spot rates. Again, we do not find evidence that the recently developed exchange rate models outperform the "old" vintage sticky-price model.

The yen–dollar has 10 out of the 15 forecast series that are cointegrated with their respective spot rates. The Canadian dollar–dollar pair, which yields relatively good forecasts according to the direction-of-change metric, has only 4 cointegrated forecast series. Evidently the forecasting performance is not just currency specific; it also depends on the evaluation criterion. The distribution of the cointegrated cases across forecasting horizons is puzzling. The frequency of occurrence is inversely proportional to the forecasting horizons. There are 19 of 51 one-quarter ahead forecast series that are cointegrated with the spot rates. However, there are only 11 of the 4-quarter ahead and 8 of the 20-quarter ahead forecast series that are cointegrated with the spot rates. One possible explanation for this result is that there are fewer observations in the 20-quarter ahead forecast series, and this effects the power of the cointegration test.

The results of testing for the long-run unitary elasticity of expectations at the 10 percent significance level are reported in table 8.8. The condition of long-run unitary elasticity of expectations, that is, the $(1, -1)$ restriction on the cointegrating vector, is rejected by the data quite frequently. The $(1, -1)$ restriction is rejected in 33 of the 38 cointegration cases. That is 13 percent of the cointegrated cases display long-run unitary elasticity of expectations. Taking both the cointegration and restriction test results together, 3 percent of the 153 cases meet the consistency criterion.

8.4.4 Discussion

Several aspects of the foregoing analysis merit discussion. To begin with, even at long horizons, the performance of the structural models is less than impressive along the MSE dimension. This result is consistent with those in other recent studies, although we have documented this finding for a wider set of models and specifications. Groen (2000) restricted his attention to a flexible price monetary model, while Faust et al. (2001) examined a portfolio balance model as well; both remained within the MSE evaluation framework.

Expanding the set of criteria does yield some interesting surprises. In particular, the direction-of-change statistics indicate more evidence that structural models can outperform a random walk. However, the basic conclusion that no economic model is consistently more successful than the others remains intact. This, we believe, is a new finding.

Even if we cannot glean from this analysis a consistent "winner," it may still be of interest to note the best and worst performing combinations of model/specification/currency. The best performance on the MSE criterion is turned in by the interest rate parity model at the

Table 8.8
Results of (1, −1) restriction test

Specification	Horizon	S–P	IRP	PROD	BEER	S–P	IRP	PROD
Panel A		*BP/$*				*BP/yen*		
ECM	1	—	39.66	—	0.32	—	—	—
		—	0.00	—	0.57	—	—	—
	4	—	—	—	19.99	49.55	—	—
		—	—	—	0.00	0.00	—	—
	20	445.3	—	458.91	—	—	—	—
		0.00	—	0.00	—	—	—	—
FD	1	—		1.56	—	24.73		—
		—		0.21	—	0.00		—
	4	—		—	—	—		—
		—		—	—	—		—
	20	—		—	—	—		—
		—		—	—	—		—
Panel B		*CAN$/$*				*CAN$/yen*		
ECM	1	—	—	—	—	—	—	—
		—	—	—	—	—	—	—
	4	—	—	—	—	—	—	—
		—	—	—	—	—	—	—
	20	—	—	—	—	—	—	—
		—	—	—	—	—	—	—
FD	1	16.58		15.73	1263	17.17		28.50
		0.00		0.00	0.00	0.00		0.00
	4	132.5		—	—	—		
		0.00		—	—	—		

	20	—		—	—	—		
		—		—	—	—		
Panel C		*DM/$*				*DM/yen*		
ECM	1	—	164.5	—	—	—	—	—
		—	0.00	—	—	—	—	—
	4	—	392.97	—	—	11.20	—	—
		—	0.00	—	—	0.00	—	—
	20	—	—	535.13	—	—	5.06	—
		—	—	0.00	—	—	0.02	—
FD	1	6.73		—	—	3.40		3.40
		0.00		—	—	0.06		0.07
	4	—		—	—	3.88		—
		—		—	—	0.04		—
	20	—		—	—	—		—
		—		—	—	—		—
Panel D		*SF/$*				*SF/yen*		
ECM	1	—	—	—		—	—	4.56
		—	—	—		—	—	0.03
	4	3.34	—	9.77		—	313.12	—
		0.06	—	0.00		—	0.00	—
	20	—	—	—		—	—	—
		—	—	—		—	—	—
FD	1	—		—		—		31.07
		—		—		—		0.00
	4	—		—		—		—
		—		—		—		—
	20	—		—		—		—
		—		—		—		—

Table 8.8
(continued)

Specification	Horizon	S–P	IRP	PROD	BEER	S–P	IRP	PROD
Panel E		*Yen/$*						
ECM	1	62.10	209.36	—				
		0.00	0.00	—				
	4	—	33.58	—				
		—	0.00	—				
	20	876.4	—	1916				
		0.00	—	0.00				
FD	1	0.582		1.03				
		0.445		0.31				
	4	—		1.14				
		—		0.29				
	20	436.4		289.22				
		0.00		0.00				

Note: The likelihood ratio test statistic for the restriction of $(1, -1)$ on the cointegrating vector and its p-value are reported. The test is only applied to the cointegration cases present in table 8.3. The notation used in the table is ECM: error correction specification; FD: first-difference specification; S–P: sticky-price model; IRP: interest rate parity model; PROD: productivity differential model; and BEER: behavioral equilibrium exchange rate model. The forecasting horizons (in quarters) are listed under the heading "horizon." The forecasting period is 1993:1 to 2000:4.

20-quarter horizon for the Canadian dollar–yen exchange rate, with a MSE ratio of 0.19 (p-value of 0.0001). The worst performances are associated with first-difference specifications; in this case the highest MSE ratio is for the first differences specification of the sticky-price exchange rate model at the 20-quarter horizon for the Canadian dollar–US dollar exchange rate. However, the other catastrophic failures in prediction performance are distributed across first-difference specifications of the various models, so the key determinant in this pattern of results appears to be the difficulty in estimating *stable* short-run dynamics. (We take here into account the fact that these predictions utilize ex post realizations of the right-hand side variables.)

Overall, the inconstant nature of the parameter estimates appears to be closely linked with the erratic nature of the forecasting performance. This applies to the variation in long-run estimates and reversion coefficients, but perhaps most strongly to the short-run dynamics obtained in the first-differences specifications.

8.5 Concluding Remarks

In this chapter we systematically assess the in-sample fit and out-of-sample predictive capacities of models developed during the 1990s. These models are compared along a number of dimensions, including econometric specification, currencies, and differing metrics.

Our investigation does not reveal that any particular model or any particular specification fit the data well, in terms of providing estimates in accord with theoretical priors. Of course, this finding is dependent on a very simple specification search, and we used theory to discipline variable selection and information criteria to select lag lengths.

On the other hand, some models seem to do well at certain horizons, for certain criteria. Indeed, it may be that one model will do well for one exchange rate and not for another. For instance, the productivity model does well for the mark–yen rate along the direction-of-change and consistency dimensions (although not by the MSE criterion), but that same conclusion cannot be applied to any other exchange rate.

Similarly we fail to find any particular model or specification that out-performed a random walk on a consistent basis. Again, we imposed the disciplining device of using a given specification, and a given out-of-sample forecasting period. Perhaps most interestingly, there is little apparent correlation between how well the in-sample estimates accord with theory and out-of-sample prediction performance.

The only link between in-sample and out-of-sample performance is an indirect one, for the interest parity condition. It is well known that interest rate differentials are biased predictors of future spot rate movements at short horizons. However, the improved predictive performance at longer horizons does accord with the fact that uncovered interest parity is more likely to hold at longer horizons than at short horizons.

In sum, while the results of our study have been fairly negative regarding the predictive capabilities of newer empirical models of exchange rates, in some sense we believe the findings pertain more to difficulties in estimation, rather than the models themselves. And this may point the direction for future research avenues.[15]

Appendix A: Data

Unless otherwise stated, we use seasonally adjusted quarterly data from the *IMF International Financial Statistics* ranging from the second quarter of 1973 to the last quarter of 2000. The exchange rate data are end of period exchange rates. Money is measured as narrow money (essentially M1), with the exception of the United Kingdom, where M0 is used. The output data are measured in constant 1990 prices. The consumer and producer price indexes also use 1990 as base year.

The three-month, annual, and five-year interest rates are end-of-period constant maturity interest rates and are obtained from the IMF country desks. See Meredith and Chinn (1998) for details. Five-year interest rate data were unavailable for Japan and Switzerland; hence data from Global Financial Data http://www.globalfindata.com/ were used, specifically, five-year government note yields for Switzerland and five-year discounted bonds for Japan.

The productivity series are labor productivity indexes, measured as real GDP per employee, converted to indexes (1995 = 100). These data are drawn from the Bank for International Settlements database.

The net foreign asset (NFA) series is computed as follows. Using stock data for year 1995 on NFA (Lane and Milesi-Ferretti 2001) at http://econserv2.bess.tcd.ie/plane/data.html, and flow quarterly data from the IFS statistics on the current account, we generated quarterly stocks for the NFA series (with the exception of Japan, for which there is no quarterly data available on the current account).

To generate quarterly government debt data, we follow a similar strategy. We use annual debt data from the IFS statistics, combined with quarterly government deficit (surplus) data. The data source for Canadian government debt is the Bank of Canada. For the United Kingdom, the IFS data are updated with government debt data from the public sector accounts of the UK Statistical Office (for Japan and Switzerland, we have very incomplete data sets, and hence no behavioral equilibrium exchange rate models are estimated for these two countries).

Appendix B: Evaluating Forecast Accuracy

The Diebold-Mariano statistics (Diebold and Mariano 1995) are used to evaluate the forecast performance of the different model specifications relative to that of the *naive* random walk.

Given the exchange rate series x_t and the forecast series y_t, the loss function L for the mean square error is defined as

$$L(y_t) = (y_t - x_t)^2. \tag{A1}$$

Testing whether the performance of the forecast series is different from that of the naive random walk forecast z_t is equivalent to testing whether the population mean of the loss differential series d_t is zero. The loss differential is defined as

$$d_t = L(y_t) - L(z_t). \tag{A2}$$

Under the assumptions of covariance stationarity and short-memory for d_t, the large-sample statistic for the null of equal forecast performance is distributed as a standard normal, and can be expressed as

$$\frac{\bar{d}}{\sqrt{2\pi \sum_{\tau=-(T-1)}^{(T-1)} l(\tau/S(T)) \sum_{t=|\tau|+1}^{T} (d_t - \bar{d})(d_{t-|\tau|} - \bar{d})}}, \tag{A3}$$

where $l(\tau/S(T))$ is the lag window, $S(T)$ is the truncation lag, and T is the number of observations. Different lag-window specifications can be applied, such as the Barlett or the quadratic spectral kernels, in combination with a data-dependent lag-selection procedure (Andrews 1991).

For the direction-of-change statistic, the loss differential series is defined as follows: d_t takes a value of one if the forecast series correctly predicts the direction of change, otherwise it will take a value of zero.

Hence a value of $\bar{d}$ significantly larger than 0.5 indicates that the forecast has the ability to predict the direction of change; on the other hand, if the statistic is significantly less than 0.5, the forecast tends to give the wrong direction of change. In large samples, the studentized version of the test statistic,

$$\frac{\bar{d} - 0.5}{\sqrt{0.25/T}}, \tag{A4}$$

is distributed as a standard normal.

Notes

We thank, without implicating, Jeff Frankel, Fabio Ghironi, Jan Groen, Lutz Kilian, Ed Leamer, Ronald MacDonald, John Rogers, Lucio Sarno, Torsten Slok, Frank Westermann, seminar participants at Academia Sinica, Boston College, UCLA, and participants at CES-Ifo conference on Exchange Rate Modeling: Where Do We Stand? for helpful comments, and Jeannine Bailliu, Gabriele Galati, and Guy Meredith for providing data. The financial support of faculty research funds of the University of California, Santa Cruz is gratefully acknowledged.

1. A recent review of the empirical literature on the monetary approach is provided by Neely and Sarno (2002).

2. See Clark and MacDonald (1999), Clostermann and Schnatz (2000), Yilmaz and Jen (2001), and Maeso-Fernandez et al. (2001) for recent applications of this specification. On the portfolio balance channel, Cavallo and Ghironi (2002) provide a role for net foreign assets in the determination of exchange rates in the sticky-price optimizing framework of Obstfeld and Rogoff (1995).

3. Despite this finding, there is little evidence that long-term interest rate differentials—or equivalently long-dated forward rates—have been used for forecasting at the horizons we are investigating. One exception from the professional literature is Rosenberg (2001).

4. The findings reported below are not very sensitive to the forecasting periods (Cheung, Chinn, and Garcia Pascual 2002).

5. For the pound, the productivity coefficient is incorrectly signed, although this finding is combined with a very large (and correctly signed) income coefficient, which suggests some difficulty in disentangling the income from productivity effects.

6. Overall, the interpretation of the results is complicated by the fact that, for the level specifications, multiple cointegrating vectors are indicated using the asymptotic critical values. The use of finite sample critical values reduces the implied number of cointegrating vectors, as indicated in the second row, to one or two vectors. Hence we do not believe the assumption of one cointegrating vector does much violence to the data.

7. One substantial caveat is necessary at this point. BEER models have almost uniformly been couched in terms of multilateral exchange rates; hence the interpretation of the BEERs in a bilateral context does not exactly replicate the experiments conducted by BEER exponents. On the other hand, the fact that it is difficult to obtain the theoretically

implied coefficient signs suggests that some searching is necessary in order to obtain a "good" fit.

8. Two recent exceptions to this characterization are Flood and Rose (2002) and Bansal and Dahlquist (2000). Flood and Rose conclude that UIP holds much better for countries experiencing currency crises, while Bansal and Dahlquist find that UIP holds much better for a set of non-OECD countries. Neither of these descriptions applies to the currencies examined in this study.

9. We opted to exclude short-run dynamics in equation (7) because, on the one hand, the use of equation (7) yields true ex ante forecasts and makes our exercise directly comparable with, for example, Mark (1995), Chinn and Meese (1995), and Groen (2000), and on the other, the inclusion of short-run dynamics creates additional demands on the generation of the right-hand-side variables and the stability of the short-run dynamics that complicate the forecast comparison exercise beyond a manageable level.

10. In using the DM test, we are relying on asymptotic results, which may or may not be appropriate for our sample. However, generating finite sample critical values for the large number of cases we deal with would be computationally infeasible. More important, the most likely outcome of such an exercise would be to make detection of statistically significant out-performance even more rare, and leaving our basic conclusion intact.

11. We also experimented with the Bartlett kernel and the deterministic bandwidth selection method. The results from these methods are qualitatively very similar. In appendix B we provide a more detailed discussion of the forecast comparison tests.

12. See also Leitch and Tanner (1991), who argue that a direction of change criterion may be more relevant for profitability and economic concerns, and hence a more appropriate metric than others based on purely statistical motivations.

13. Using Markov switching models, Engel (1994) obtains some success along the direction of change dimension at horizons of up to one year. However, his results are not statistically significant.

14. The Johansen method is used to test the null hypothesis of no cointegration. The maximum eigenvalue statistics are reported in the manuscript. Results based on the trace statistics are essentially the same. Before implementing the cointegration test, both the forecast and exchange rate series were checked for the $I(1)$ property. For brevity, the $I(1)$ test results and the trace statistics are not reported.

15. Our survey is necessarily limited, and we leave open the question of whether alternative statistical techniques can yield better results, for example, nonlinearities (Meese and Rose 1991; Kilian and Taylor 2001), fractional integration (Cheung 1993), and regime switching (Engel and Hamilton 1990), cointegrated panel techniques (Mark and Sul 2001), and systems-based estimates (MacDonald and Marsh 1997).

References

Alexius, A. 2001. Uncovered interest parity revisited. *Review of International Economics* 9(3): 505–17.

Andrews, D. 1991. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59: 817–58.

Bansal, R., and M. Dahlquist. 2000. The forward premium puzzle: Different tales from developed and emerging economies. *Journal of International Economics* 51(1): 115–44.

Cavallo, M., and F. Ghironi. 2002. Net foreign assets and the exchange rate: Redux revived. Mimeo NYU and Boston College. February.

Chen, Y.-C., and K. Rogoff. 2003. Commodity currencies. *Journal of International Economics* 60(1): 133–60.

Cheung, Y.-W. 1993. Long memony in foreign exchange rate. *Journal of Business and Economic Statistics* 11(1): 93–102.

Cheung, Y.-W., and K. S. Lai. 1993. Finite-sample sizes of Johansen's likelihood ratio tests for cointegration. *Oxford Bulletin of Economics and Statistics* 55(3): 313–28.

Cheung, Y.-W., and M. Chinn. 1998. Integration, cointegration, and the forecast consistency of structural exchange rate models. *Journal of International Money and Finance* 17(5): 813–30.

Cheung, Y.-W., M. Chinn, and A. G. Pascual. 2002. Empirical exchange rate models of the 1990s: Are any fit to survive? NBER Working Paper 9393.

Chinn, M. 1997. Paper pushers or paper money? Empirical assessment of fiscal and monetary models of exchange rate determination. *Journal of Policy Modelling* 19(1): 51–78.

Chinn, M., and R. Meese. 1995. Banking on currency forecasts: How predictable is change in money? *Journal of International Economics* 38(1–2): 161–78.

Chinn, M., and G. Meredith. 2002. Testing uncovered interest parity at short and long horizons during the post–Bretton Woods era. Mimeo. Santa Cruz, CA.

Christoffersen, P. F., and F. X. Diebold. 1998. Cointegration and long-horizon forecasting. *Journal of Business and Economic Statistics* 16: 450–58.

Clark, P., and R. MacDonald. 1999. Exchange rates and economic fundamentals: A methodological comparison of Beers and Feers. In J. Stein and R. MacDonald, eds., *Equilibrium Exchange Rates*. Boston: Kluwer, pp. 285–322.

Clostermann, J., and B. Schnatz. 2000. The determinants of the euro–dollar exchange rate: Synthetic fundamentals and a non-existing currency. *Konjunkturpolitik* 46(3): 274–302.

Cumby, R. E., and D. M. Modest. 1987. Testing for market timing ability: A framework for forecast evaluation. *Journal of Financial Economics* 19(1): 169–89.

Dornbusch, R. 1976. Expectations and exchange rate dynamics. *Journal of Political Economy* 84: 1161–76.

Diebold, F., and R. Mariano. 1995. Comparing predictive accuracy. *Journal of Business and Economic Statistics* 13: 253–65.

Engel, C. 1994. Can the Markov switching model forecast exchange rates? *Journal of International Economics* 36(1–2): 151–65.

Engel, C., and J. Hamilton. 1990. Long swings in the exchange rate: Are they in the data and do markets know it? *American Economic Review* 80(4): 689–713.

Faust, J., J. Rogers, and J. Wright. 2003. Exchange rate forecasting: The errors we've really made. *Journal of International Economics* 60(1): 35–59.

Flood, R., and A. Rose. 2002. Uncovered interest parity in crisis. *IMF Staff Papers* 49(2): 252–66.

Frankel, J. A. 1979. On the mark: A theory of floating exchange rates based on real interest differentials. *American Economic Review* 69: 610–22.

Groen, J. J. 2000. The monetary exchange rate model as a long-run phenomenon. *Journal of International Economics* 52(2): 299–320.

Isard, P., H. Faruqee, G. R. Kinkaid, and M. Fetherston. 2001. Methodology for current account and exchange rate assessments. IMF Occasional Paper 209.

Kilian, L., and M. Taylor. 2003. Why is it so difficult to beat the random walk forecast of exchange rates. *Journal of International Economics* 60(1): 85–107.

Lane, P., and G. Milesi-Ferretti. 2001. The external wealth of nations: Measures of foreign assets and liabilities for industrial and developing. *Journal of International Economics* 55: 263–94.

Leitch, G., and J. E. Tanner. 1991. Economic forecast evaluation: Profits versus the conventional error measures. *American Economic Review* 81(3): 580–90.

Lucas, R. 1982. Interest rates and currency prices in a two-country world. *Journal of Monetary Economics* 19(3): 335–59.

MacDonald, R., and J. Nagayasu. 2000. The long-run relationship between real exchange rates and real interest rate differentials. *IMF Staff Papers* 47(1): 116–28.

MacDonald, R., and M. P. Taylor. 1994. The monetary model of the exchange rate: Long-run relationships, short-run dynamics and how to beat a random walk. *Journal of International Money and Finance* 13(3): 276–90.

Maeso-Fernandez, F., C. Osbat, and B. Schnatz. 2001. Determinants of the euro real effective exchange rate: A BEER/PEER approach. ECB Working Paper 85.

Mark, N. 1995. Exchange rates and fundamentals: Evidence on long horizon predictability. *American Economic Review* 85: 201–18.

Mark, N., and D. Sul. 2001. Nominal exchange rates and monetary fundamentals: Evidence from a small post–Bretton Woods panel. *Journal of International Economics* 53(1): 29–52.

Meese, R., and K. Rogoff. 1983. Empirical exchange rate models of the seventies: Do they fit out of sample? *Journal of International Economics* 14: 3–24.

Meese, R., and A. K. Rose. 1991. An empirical assessment of non-linearities in models of exchange rate determination. *Review of Economic Studies* 58(3): 603–19.

Neely, C. J., and L. Sarno. 2002. How well do monetary fundamentals forecast exchange rates? *Federal Reserve Bank of St. Louis Review* 84(5): 51–74.

Obstfeld, M., and K. Rogoff. 1996. *Foundations of International Macroeconomics*. Cambridge: MIT Press.

Rosenberg, M. 2001. Investment strategies based on long-dated forward rate/PPP divergence. *FX Weekly* (New York: Deutsche Bank Global Markets Research, April 27), pp. 4–8.

Stein, J. 1999. The evolution of the real value of the US dollar relative to the G7 currencies. In J. Stein and R. MacDonald, eds., *Equilibrium Exchange Rates*. Boston: Kluwer, pp. 67–102.

Stockman, A. 1980. A theory of exchange rate determination. *Journal of Political Economy* 88(4): 673–98.

Yilmaz, F., and S. Jen. 2001. Correcting the US dollar—A technical note. Morgan Stanley Dean Witter (June 1).