ELSEVIER Journal of International Economics 38 (1995) 161–178

Journal of
INTERNATIONAL
ECONOMICS

Banking on currency forecasts: How predictable is change in money?

Menzie D. Chinn[a,*], Richard A. Meese[b]

[a]Department of Economics, University of California, Santa Cruz, CA 95064, USA
[b]Haas School of Business Administration, University of California, Berkeley, CA 94720, USA

Received August 1992, revised version received April 1994

Abstract

The paper examines the predictive performance of four structural exchange rate models using both parametric and nonparametric techniques. Error correction versions of the models are fit so that plausible long-run elasticities can be imposed on the fundamental variables of each model. A variety of model evaluation statistics are reported. Our findings confirm that fundamental exchange rate models forecast no better than a random walk model for short-term prediction horizons. For longer horizons, error correction terms can explain exchange rate movements significantly better than a no change forecast for a subset of the models and currencies we consider.

Key words: Exchange rates; Forecasting; Random walk

JEL classification: F31; F41; F47

* Correspondence to: Menzie Chinn, Department of Economics, University of California, Santa Cruz, Santa Cruz, CA 95065, or Richard Meese, Haas School of Business, University of California at Berkeley, Berkeley, CA 94720.

An earlier version of this paper was presented at the 1991 Summer Meetings of the Econometric Society. We thank Yin-Wong Cheung, Andrew Rose, Douglas Steigerwald, and a co-editor of this Journal for many useful comments. Chinn acknowledges the support of UCSC research funds, and the Group for International and Comparative Economic Studies. Meese acknowledges support from the Berkeley Program in Finance. This is an abridged version of UCSC GICES/Department of Economics Working Paper #264.

SSDI 0022-1996(94)01334-O

1. Introduction

It is widely recognized that current exchange rate models fit poorly on post-Bretton Woods data. Common problems include parameter instability and dismal out-of-sample explanatory power. Meese (1990) provides a recent survey and references. Candidate explanations for the empirical failure of current models include (i) simultaneity problems, (ii) improper modeling of expectations formation, (iii) failure to account for nonlinearities in the data generation mechanism (DGM) of exchange rates, and (iv) over-reliance on the representative agent paradigm, among others.

Much recent theoretical and empirical work on exchange rates has focused on the third explanation for the shortcomings of current models. Plausible rationales for nonlinearities in the DGM of exchange rates include time deformation (Stock, 1987), regime switching models (Engel and Hamilton, 1990), target zone exchange rate models (Krugman, 1991), chaos and fractional integration (Hsieh, 1989; Cheung, 1992), and garden variety mis-specification of the functional form of behavioral relations.

We pay homage to this literature by fitting three representative exchange rate models using a variety of parametric and nonparametric techniques. In principle, the nonparametric technique should capture any important nonlinearities in our structural models. Also, we impose long-run constraints on our representative models and fit them by a variety of techniques in the hope of mitigating difficulties associated with explanations (i), (ii), and (iv) as well. We examine four bilateral rates (Canada, Germany, Japan and the United Kingdom) relative to the U.S. dollar, and the German mark–Japanese yen cross rate, using monthly fundamental data.

The case for imposing structure on our representative models is compelling for the following reasons. First, the short time series of post-Bretton Woods data makes it difficult to obtain precise estimates of long-run elasticities. Hence we fit error correction models (ECMs) with long-run elasticities imposed a priori, using parameter values from both the theoretical and empirical literature on money demand and exchange rate determination. In contrast, short-run dynamics are modeled as flexibly as possible.

Second, the low signal-to-noise ratio in exchange rate data suggests that by utilizing nonsample information in the construction of empirical exchange rate models, we may be able to improve on the dismal out-of-sample performance of the current generation of models.[1] We do not believe that major country exchange rates can move independently of macroeconomic fundamentals over long time horizons. Thus we have decided to impose

[1] See Meese and Rogoff (1983) for an earlier attempt to impose structure on representative models to improve out-of-sample explanatory power.

long-run constraints on our empirical exchange models using an error correction term (ECT). In the models that follow, the error correction term—the difference between the actual and fundamental value of the spot exchange rate—is employed whether or not it passes standard tests for stationarity. Our rationale is that standard unit root tests are known to have low power against borderline stationary alternatives,[2] and previous work in international finance suggests that the speed of adjustment in the goods market can be quite slow.

The low ratio of signal to noise in exchange rate data also suggests that we examine a variety of model evaluation statistics when assessing the value of fundamental information for explaining currency movements. Towards this end we report several test statistics for our representative models so that the reader can draw his/her own conclusions. We use the statistical test of Diebold and Mariano (1993) to assess the statistical significance of forecast accuracy, and compare our candidate models with the nonstructural alternative of a random walk without drift.

The paper is organized as follows. Representative exchange rate models are described in Section 2. Econometric issues of estimation and testing are discussed in Section 3, and forecasting results in Section 4. Given the generally discouraging results in Section 4, an alternative, long-horizon approach is implemented in Section 5. In this section we isolate the explanatory power of ECTs for long-run (up to three-year) changes in exchange rates. For two of the bilateral rates we consider, the out-of-sample explanatory power of the ECTs is statistically significant. Concluding remarks comprise the final section.

2. Representative exchange rate models

The three models[3] we consider are descendants of the original monetary models of Dornbusch (1976a), Frenkel (1976), Bilson (1978) and Mussa (1976). All start with conventional money demand functions for both the domestic and foreign economies, and impose the condition that expected depreciation equal the nominal interest differential plus an exogenous risk

[2] The problem is more severe that we have let on. Faust (1993) has shown that processes with and without a unit root are "nearly observationally equivalent". The implications of this result for classical inference procedures are grim; statistical tests to distinguish trend stationary from different stationary models (it does not matter which is the null model) have power less than or equal to test size.

[3] In the working paper version of the paper, which includes results for data at the quarterly frequency, we also examine a version of an intertemporal model based on Finn (1989) and Chinn (1993), which expresses the spot rate as a function of consumption flows and money stocks. Consumption data are not available at the monthly frequency and so are omitted here.

premium on domestic assets that may or may not be zero. Our three models are subsumed in

$$s = F[(m - m^*),\ (y - y^*),\ (q - q^*),\ (i - i^*),\ (\pi - \pi^*),\ (tb - tb^*)]\,, \tag{1}$$

where s, m, y, and q are the logarithms of the exchange rate (domestic currency per unit of foreign currency), money supply, real income proxied by industrial production, and the relative price of tradables to nontradables, and i, π, and tb are the levels of the nominal interest rate, the CPI inflation rate, and the cumulated real trade balance, respectively. An asterisk denotes a foreign variable.

Model 1 contains only the terms in monies, incomes, and nominal interest rates, and relies on the further assumption that purchasing power parity (PPP) holds up to an exogenous real exchange rate shock. This 'flexible price' monetary model subsumes the Lucas (1982) model since the latter model contains monies and real incomes but no interest rate term.

Model 2, a 'sticky price' monetary model does not assume PPP holds at all times, as in Frankel (1979). Instead it assumes slow adjustment of goods prices relative to asset prices. Our first version of a sticky price model contains all terms except $(q - q^*)$. Cumulated trade balances enter the specification if wealth is included in the money demand equation; see Frankel (1982). Hooper and Morton (1982) provide an additional rationale for the trade balance terms.

Our third model (like model 2) is motivated by the failure of PPP to hold for broad price indices, such as the consumer price index and GNP deflators. One approach is to make an explicit recognition of nontraded goods, and to posit that PPP only holds for tradable goods. This approach received its original impetus from Balassa's (1964) observation that aggregate price indices seemed to yield counter-intuitive conclusions on whether particular exchange rates were under- or overvalued (see, more recently, Marston, 1987).

Dornbusch (1976b) originally introduced nontradables into a monetary model of the exchange rate. Empirical implementation was pursued by Wolff (1987) and others. Suppose that the aggregate price level index can be represented by a Cobb–Douglas function of the individual nontraded and traded price indices (in logs):

$$p = \beta p^{\mathrm{N}} + (1 - \beta)p^{\mathrm{T}}\,, \tag{2}$$

where β is the proportion of nontradable goods in the consumption bundle. Then one can solve in the usual manner for the spot exchange rate, incorporating sticky prices as in the traditional Dornbusch specification. This yields a version of (1) with all terms except $(tb - tb^*)$.

Unfortunately, the proxies available for either the tradables or nontradables price indices are somewhat limited. Clements and Frenkel (1980) used the manufacturing wages for nontradables, and the wholesale price index for tradables. Wolff (1987) used the PPI and the CPI. In order to make the results comparable at monthly frequencies, and across countries, we opted for the PPI and CPI as our proxies.[4]

3. Estimation and testing of the representative models

3.1. Overview

Several features of the quasi-reduced form (1) merit further comment. First, the models can be written in a present value representation, where the exchange rate is the discounted sum of a linear combination of future fundamentals. However, substitution for the expected future spot rate using the uncovered interest parity assumption allows estimation of the models without the usual complications of rational expectations econometrics. A related point is that direct estimation of (1) does not require imposition of a transversality condition that precludes rational bubbles from the DGM for spot rates.

Second, it is generally accepted that exchange rates and their fundamentals are well approximated by unit root processes.[5] There is considerably less consensus on the evidence for cointegration of exchange rates and the set of fundamentals we consider. In the experiments that follow we fit both unconstrained versions of the models in first differences, and ECM versions that constrain the long-run dynamics of the exchange rate and the appropriate set of fundamentals.

3.2. Construction of the error correction terms

As noted above, we impose the cointegrating vector a priori, using coefficient values from the theoretical and empirical literature on money demand and exchange rate determination. In our view, the post-Bretton

[4] There is a slight problem of interpretation, since the CPI includes nontradables goods prices. However, our measure of β only affects the constant term, given our logarithmic specification of the exchange rate.

[5] For a recent set of test results see Diebold (1988) or Meese and Rose (1991), among others. Also, see Cheung (1992) for alternative evidence suggesting that bilateral exchange rates are better characterized by fractionally integrated processes.

Woods era is too short to extract reliable estimates of the long-run elasticities by direct estimation.[6] Attempts to estimate these cointegrating vectors directly, either by the modifications of the conventional Engle–Granger methodology, or using the multivariate analogue forwarded by Johanssen and Juselius (1990) usually yield incorrectly signed coefficients, buttressing our contention.

Hence, we assume that a log-linear version of (1) is appropriate in the long run, and impose a set of coefficient restrictions for each of our candidate models. These values are given in Table 1. For all models, the money supply and income elasticities are the same. We impose a money supply elasticity of unity, and an income elasticity of 0.75. This latter value is bracketed by the estimates of the long-run income elasticity reported in Stock and Watson (1993) for post-War U.S. data. Boughton (1991) also reports higher income elasticities for the countries we consider, from a low of 0.83 in Japan to a high of 3.3 in the United Kingdom.

The coefficients on interest rates, inflation rates, relative prices (q), and cumulated trade balances vary by model, although the coefficients on the former two variables are functions of the interest rate semi-elasticity. We use

Table 1
Imposed long-run elasticities on models

Model 1: $s = (m - m^*) - \phi(y - y^*) + \mu(i - i^*)$

Model 2: $s = (m - m^*) - \phi(y - y^*) + (\mu + 1/\theta)(\pi - \pi^*) - (1/\theta)(i - i^*) - \gamma(tb - tb^*)$

Model 3: $s = (m - m^*) - \phi(y - y^*) + (\mu + 1/\theta)(\pi - \pi^*) - (1/\theta)(i - i^*) + \beta((p^{\mathrm{T}} - p^{\mathrm{N}}) - (p^{\mathrm{T}*} - p^{\mathrm{N}*}))$

Coefficient	Imposed value	Average Engle–Granger	Average Stock–Watson
ϕ	0.75	1.72	3.70
μ	4.50	2.00	3.20
θ	0.50	0.58	0.36
τ	$5.0 * 10^{-4}$	$8.8 * 10^{-4}$	$2.0 * 10^{-3}$
β	0.50	0.64	0.26

Notes: Average Engle–Granger and Average Stock–Watson are the average (over the five bilateral data sets) coefficient values in Engle–Granger and Stock–Watson (dynamic OLS) cointegrating regressions.

[6] We do, however, provide average estimates of the cointegrating vector in Table 1 estimated by both Engle and Granger (1987), and the 'dynamic ordinary least squares' method of Stock and Watson (1993). A further rationale for imposing coefficient constraints a priori comes from the work of Meese and Rogoff (1983). In their long-run forecasting experiments, constrained coefficient models start to out-perform random walk forecasts at prediction horizons of about 18 months. We use median values of the parameter grids considered in Meese and Rogoff in our experiments below.

4.5 for this value; it depends on the average value of interest rates over the period of interest. This value is also in the range of post-War estimates reported in Stock and Watson (1993), roughly (1.6, 9.1) for the United States, and Boughton (1991), roughly (3.01, 5.57) for four of the countries we consider.

The goods market speed of adjustment parameter is taken to be 0.5 on an annual basis; this corresponds to deviations from PPP damping at rate 0.94 for monthly data. Real cumulated trade balances are converted to a common currency and weighted by a wealth elasticity of money demand equal to 0.0005. This is substantially smaller than the Hooper and Morton (1982) estimate of 0.003 over the 1973–1978 period. The choice of 0.0005 is data driven; it is closer to the estimates obtained from a cointegrating regression that controls for nuisance parameters (short-run dynamics) in the exchange rate relation.[7]

The final parameter of interest is the share of nontradables in the CPI. Our choice of this parameter is also data driven; we take β to be 0.5, which is one minus the geometric average of OLS estimates derived from first difference regressions of the log CPI on the log PPI using data from 1973–1990 for the five countries we consider.

3.3. *Stationarity of the error correction terms*

In order for the ECMs to make sense, the ECTs must be stationary. In Table 2 we conduct a formal assessment of stationarity, using the augmented Dickey–Fuller (ADF) test for a regression with constant and trend. The

Table 2
Augmented Dickey–Fuller tests on error correlation terms

	Model 1	Model 2	Model 3
CN	3.376*	3.157*	3.577**
UK	2.066	2.299	2.037
GY	2.423	3.425*	3.504**
JP	3.054	3.567**	2.551
GJ	3.248*	2.731	2.498

Note: The figures under either Model 1 (Frenkel–Bilson), Model 2 (Hooper–Morton) or Model 3 (Balassa) indicate the t-statistic for an augmented Dickey–Fuller test (with trend) on the ECT, using the indicated lag order for the first differences. The lag length is selected using the Akaike Information Criterion; in all cases 0 was chosen.

*(**)[***] denote significance at the 10(5)[1] percent level.

[7] Note that the average values for γ are for the U.S. coefficients only. Incorrectly signed coefficients are excluded from the average. Blundell-Wignall and Browne (1991) also find declines in sensitivity of exchange rates to cumulated trade balances, which they attribute to financial deregulation.

Akaike (1973) Information Criterion (AIC) is used to determine the optimal lag length; in all cases the implied lag order is zero. The results are mixed; however, seven out of the 15 ECTs appear trend stationary, with the greatest number of rejections appearing for the ECTs based on the Hooper–Morton model.

Since unit root tests have poor power characteristics for borderline stationary alternatives, we employ the Kwiatkowsi–Phillips–Schmidt–Shin (KPSS) test of trend stationarity (Kwiatkowski et al., 1992) to view the evidence from the trend stationary null. This is an LM test constructed by comparing the variance of the trend stationary component error with that of the random walk component error. In order to conduct the test, it is necessary to assume a serial correlation lag length to calculate a robust estimate of the variance for the trend stationary error. Hence, for each ECT there are two entries, corresponding to 4 and 14 monthly lags.[8] The results of implementing this test are reported in Table 3. When a high degree of serial correlation is assumed (14 lags)[9] then there are more failures to reject the trend stationary null.

Despite the fact that our empirical investigation into the stationarity of

Table 3
KPSS test on error correction terms

	ℓ	Model 1	Model 2	Model 3
CN	4	0.384***	0.323***	0.175**
	14	0.156**	0.133*	0.076
UK	4	0.585***	0.324***	0.453***
	14	0.228***	0.147**	0.196**
GY	4	0.482***	0.405***	0.355***
	14	0.184**	0.161**	0.210**
JP	4	0.552***	0.375***	0.687***
	14	0.219***	0.161**	0.281***
GJ	4	0.267***	0.526***	0.559***
	14	0.120*	0.226***	0.234***

Notes: The figures under either Model 1 (Frenkel–Bilson), Model 2 (Hooper–Morton) or Model 3 (Balassa) indicate the $\hat{\eta}_r$ statistic for the KPSS test, using the indicated lag order for estimating the serial correlation robust variance estimators.

*(**)[***] denotes significance at the 10(5)[1] percent level.

[8] The lag lengths of $\ell 0 = 0$, and $\ell 4 = \mathrm{INT}[4(T/100)^{1/4}]$ and $\ell 12 = \mathrm{INT}[12(T/100)^{1/4}]$ are based on suggestions due to Schwert (1989). When a lag order of 0 is chosen (implicitly assuming i.i.d. errors for the trend stationary process, which is very unlikely considering the macroeconomic variables under consideration), then the null hypothesis of trend stationary is always rejected.

[9] Monte Carlo simulations by Kwiatkowski (et al., 1992; pp. 171–72) indicate that for a sample size of 200, there is little size distortion even for a lag order of 14, while power is still fairly high.

ECTs is inconclusive,[10] we proceed to use the ECTs in our estimated models. In Section 5 below we provide simulation evidence that the explanatory power of the ECTs for long changes in exchange rates is not attributable to the high degree of serial correlation in each of the series.

3.4. Specification and estimation techniques

The three models are estimated by OLS and instrumental variables (IV) procedures in unconstrained first differences, and in ECMs. Our ECM variants include the ECT lagged once, and the first difference of fundamentals lagged once. Thus all regressors in the error correction models are predetermined, and one-step-ahead forecasts are ex ante forecasts. In Section 5 we also examine a restricted ECM where the k-period change in the exchange rate is regressed only on a constant and the ECT lagged k periods.

Our error correction specifications are also estimated using a nonparametric regression technique called 'locally weighted regression' (LWR) (see Cleveland and Devlin, 1988). This procedure requires stationary, weakly exogenous regressors, so they are most appropriate for a reduced-form model like our ECM, or an unconstrained model using lagged first differences of fundamentals.

This procedure can be briefly described as follows. Consider the problem of estimating

$$s(t) - s(t-1) = H[X(t-1)] + \epsilon(t) , \tag{3}$$

where H is an unknown smooth function, and X is the vector of first differences of fundamental variables which may include the level of the cointegrating error. The nonparametric estimators allow local linear approximation to the function H by estimating the relation in a moving average manner—essentially by running a least squares regression at every point in the sample.[11] The advantage of our nonparametric procedure is that it should account for possible nonlinearity in the data generation mechanism for the spot exchange rate.

[10] Indeed, the analysis in Faust (1993) suggests our attempt to test for unit roots in the ECTs may not be informative.

[11] The procedure uses a 'window' of observations around each observation $X(s)$ in the sample to estimate the curvature of the function H at the point at $X(s)$. Forecasting with LWR requires that we fit the model only for the last available data point. As such, we chose the window size equal to the number of data points available through the end of 1982, our originating sample.

3.5. The forecasting exercise

We evaluate the out-of-sample explanatory power of our representative models over two forecast periods. The first starts with the end of the 1981–82 recession in the United States, and the second corresponds to the period after the Louvre Accord in April 1987. Since our results are not sensitive to the forecast period, we tabulate results for the longer forecast period, 1983–1990, only.

Fixed sample periods are used, where the originating sample period is 1973:03–1982:12 (118 observations). Whenever necessary, forecasts use actual realized values of the right-hand-side variables. As we 'roll' through each forecast period, parameter estimates are updated with the addition of each new data point. We make use of the root mean square forecast error (RMSE) as the forecast evaluation statistic; we present its values in terms of the logarithmic level of the spot rate so it is in approximate percentage terms. This statistic is calculated for each of the forecast horizons. In addition, we use the random walk as a naive benchmark for comparisons of exchange rate model forecasts.

4. Forecasting results

The forecasting performance of our representative models is reported in Table 4; part (a) contains 1-month forecasts and part (b) contains 12-month forecasts. Each panel of Table 4 contains all four structural models, first difference and ECM specifications, and all five exchange rates. The results in the table are generally discouraging (relative to the naive random walk model with or without drift), so only the lowest out-of-sample RMSE estimation technique for the each model is reported.[12] In parentheses, next to the RMSE statistic, we indicate the estimation technique that produced the lowest value of the statistic. We use (1) to denote OLS, (2) to denote (IV), and (3) to denote LWR, for both the differences and ECM forms.

According to the RMSE evaluation metric, the out-of-sample explanatory power of the representative models in parts (a) and (b) of Table 4 is seldom better than the random walk model with or without drift. The predictive performance of the structural models shows some improvement at horizons

[12] This condensation is necessary to prevent visual overload. There are 120 structural model RMSEs *alone* corresponding to the top half of Table 4. All the RMSEs are available from the authors upon request. Rankings by mean absolute errors are only slightly different.

Table 4
(a) %RMSE statistics, all models, 1-month ahead forecasts 1983: 01–1990:11

	\$/Dm	\$/C\$	\$/£	\$/¥	Dm/¥
Random walk	3.46	1.14	3.67	3.32	2.49
Random walk with drift	3.45	1.16	3.70	3.29	2.48
First difference form					
Model 1 (flex-price)	3.47 (3)	1.13 (3)	3.65 (3)	3.35 (3)	2.66 (3)
Model 2 (sticky-price)	3.50 (3)	1.15 (3)	3.72 (3)	3.51 (3)	2.68 (3)
Model 3 (relative price)	3.55 (3)	1.16 (3)	3.71 (3)	3.41 (3)	2.63 (3)
ECM form					
Model 1 (flex-price)	3.53 (3)	1.14 (3)	3.71 (3)	3.45 (3)	2.65 (3)
Model 2 (sticky-price)	3.62 (3)	1.15 (3)	3.81 (3)	3.47 (3)	2.61 (3)
Model 3 (relative price)	3.60 (3)	1.18 (3)	3.80 (3)	3.49 (3)	2.59 (3)

(b) %RMSE statistics, all models, 12-month ahead forecasts 1983: 12–1990:11

	\$/Dm	\$/C\$	\$/£	\$/¥	Dm/¥
Random walk	16.8	4.99	13.3	17.0	10.9
Random walk with drift	16.7	5.59	14.7	16.7	7.43
First difference form					
Model 1 (flex-price)	16.9 (1)	5.06 (3)	13.8 (1)	15.4 (2)	11.2 (3)
Model 2 (sticky-price)	16.2 (3)	5.26 (3)	14.0 (3)	18.2 (1)	11.7 (3)
Model 3 (relative price)	16.4 (1)	5.48 (3)	14.1 (3)	15.9 (1)	12.5 (3)
ECM form					
Model 1 (flex-price)	19.4 (1)	5.22 (3)	17.1 (3)	17.1 (1)	12.9 (1)
Model 2 (sticky-price)	19.3 (3)	5.04 (1)	17.1 (3)	19.0 (1)	11.7 (3)
Model 3 (relative price)	19.7 (1)	5.49 (3)	16.3 (3)	16.9 (1)	11.9 (3)

Notes: First difference refers to a model fit with contemporaneous first differences of exchange rates and fundamentals. The ECMs are fit with a single lag of differenced fundamentals and the lagged level of the ECT. (1) denotes OLS, (2) denotes 2SLS, and (3) denotes LWR estimation procedures. For 2SLS, we use lags 2–4 of the right-hand-side variables, since there is evidence of MA1 serial correlation, possibly due to time aggregation.

beyond 1-month ahead. The lowest RMSE models for the 1-year horizons appear to be split randomly over estimation techniques and theoretical models. It is true, however, that the LWR forecasts produce the largest number of smallest RMSE statistics for the structural models (whether or not they improve on the random walk model).

ECMs with an ECT lagged once do not, in general, produce the best RMSE forecasts at the yearly predictive horizons. This result obtains despite the fact that in full sample estimation of our representative models, ECTs have the correct (negative) sign. Unfortunately, the single lagged ECTs are rarely significant.

5. Long horizon forecasting

5.1. Specification

We next consider whether it might be useful to forecast very far ahead. Mark (1992) has recently demonstrated that long horizon exchange rate changes are explainable using predetermined values of the difference between the actual and a fundamental value of the exchange rate. In our context this corresponds to the estimation of a constrained ECM for the k-period change in the exchange rate using only the ECT lagged k periods. Our results are based on a regression of the form

$$s(t) - s(t-k) = \theta_0 + \theta_1 e(t-k) + \delta(t) , \tag{4}$$

where $e(t)$ is a cointegrating error (the difference of actual spot rate minus the fundamental value from Table 1), k is the forecast step (12, 24, and 36 months in our applications), and the error term, $\delta(t)$, may exhibit serial correlation.

This model is similar to the ECMs estimated in Section 4, stripped of the short-term dynamics.[13] On the basis of in-sample regression estimates, there is some econometric justification for deleting the short-run coefficients, as they are generally insignificant at the conventional levels. Moreover, they often have signs that are not consistent with our priors.

In-sample statistics for Eq. (4) are provided in Table 5. Since the results are similar across horizons, we only provide the results for the 36-month

[13] For example, an $N \times 1$ vector time series y is said to have an error correction representation if

$$A(L)(1-L)y(t) = -Be(t-1) + u(t) ,$$

where L is the lag operator, $A(0) = I$, the $r \times 1$ vector $e(t) = a'y(t)$ is the stationary ECT, and $u(t)$ is a vector white noise (Engle and Granger, 1987). Consider the ith equation in the system with dependent variable $s(t)$. Assume $A(L) = I$, $r = 1$, and that $e(t)$ is an AR(1): $e(t) = \rho e(t-1) + v(t)$, with correlation coefficient ρ less than but near 1. Then Eq. (4) of the text can be obtained by substitution, i.e. for $k = 2$, let the $N \times 1$ vector $B = \{b(i)\}$, then

$$s(t) - s(t-2) = -b(i)(1+\rho)\, e(t-2) + u(t) + u(t-1) - bv(t-1) .$$

Thus the composite error term $\delta(t)$ in Eq. (4) will follow a moving average process of order $(k-1)$. In our contrived example, as k increases, the coefficient on $e(t-k)$, will also increase.

Table 5
Regressions of 36-month changes in restricted ECM 1974.12–1990.12

	\$/Dm		\$/C\$		\$/£		\$/¥		Dm/¥	
	R^2	$\hat{\theta}$	R^2	$\hat{\theta}$	R^2	$\hat{\theta}$	R^2	$\hat{\theta}$	R^2	$\hat{\theta}$
Model 1	0.49	−0.816***	0.64	−0.404***	0.00	−0.056	0.05	−0.208	0.00	−0.063
		[0.170]		[0.072]		[0.269]		[0.198]		[0.260]
		{0.050}		{0.025}		{0.500}		{0.300}		{0.500}
Model 2	0.23	−0.656**	0.61	−0.235***	0.06	−0.264	0.07	−0.227	0.06	0.145
		[0.275]		[0.043]		[0.311]		[0.165]		[0.119]
		{0.200}		{0.100}		{0.400}		{0.300}		{0.300}
Model 3	0.15	−0.517***	0.58	−0.361***	0.00	−0.072	0.07	−0.201	0.13	−0.192**
		[0.187]		[0.078]		[0.259]		[0.133]		[0.097]
		{0.100}		{0.050}		{0.500}		{0.300}		{0.200}

Notes: The entries are the OLS regression coefficients on the error correction term, from Eq. (4). Newey–West standard errors using $k-1$ lag windows are in the square brackets. Figures in braces are simulated marginal significance levels for the null statistic that the slope coefficient equals zero (see text).
*(**)[***] denotes significance at the 10(5)[1] percent level.

horizon. We tabulate both R^2 values and an asymptotic t-ratio for the coefficient on the lagged ECT; the latter is formed by dividing the OLS estimate by a consistent estimate of its standard error.

Since the variables in both sides of Eq. (4) exhibit considerable serial correlation, there is some danger that a finding of a significant coefficient on the ECT is a manifestation of the 'spurious regression' problem (Granger and Newbold, 1977). Thus, in the last row of Table 5 we provide simulated marginal significance levels (p-values) for testing that the slope coefficient in Eq. (4) is zero.

The p-values are simulated as follows. Pseudo exchange rate data are generated as a random walk with standard normal errors (the k-period exchange rate change is thus a moving average of order $k-1$ with unit coefficients), and pseudo ECT data are generated as a first-order autoregression with coefficient 0.95, independent of the pseudo exchange rate. The variance of the ECT relative to the variance of the one-period change in the exchange rate is calibrated using our historical data on both series (variances of the exchange rate changes and ECTs are quite similar for the exchange rates we consider). The OLS coefficient on the ECT regressor is divided by a Newey–West (1987) robust estimate of its standard error, and the distribution of this ratio is tabulated for 1000 replications. p-Values are for a one-sided test, since a positive ECT coefficient has no economic meaning in our context. All regressions include a constant term.

The simulated p-values of the test statistics are in the range of 5–30 percent for Canada, Germany, Japan, and the German–Japanese cross rate. In addition, R^2 values generally rise as the length of the horizon increases (i.e. from one year to three years) for the \$/Dm and \$/C\$; rates, but not for the others. In addition, coefficient values on the lagged cointegrating errors (when they are precisely estimated) indicate that between 20 and 80 percent of the difference between actual and fundamental values of the exchange rate is removed over a two- to three-year period.

5.2. Forecasting results

In Table 6 we report the ratio of the restricted error correction (REC) model (5) RMSE to the random walk RMSE for the 36-month horizon only, since these results are most positive. We also report a test of the equality of REC and random walk model forecast error RMSE using the Diebold and Mariano (1993) procedure, and the proportion of correct forecasts of the direction of change in the exchange rate.

The equality of RMSE statistic is robust to bias, non-normality, and serial correlation in forecast errors. It is asymptotically distributed as a standard

Table 6
Performance of restricted ECM for 36-Month horizon 1985.12–1990.12

	\$/Dm	\$/C\$	\$/£	\$/¥	Dm/¥
Model 1					
Dir. chg.	0.705	0.361	0.197**	0.918***	0.590
RMSE ratio	0.525*	1.165	2.406**	0.877***	1.088
DM stat.	{1.694}	{1.292}	{−2.240}	{3.340}	{−0.805}
p-value	(0.095)	(0.201)	(0.029)	(0.001)	(0.424)
Model 2					
Dir. chg.	0.689	0.557	0.197**	0.770***	0.672
RMSE ratio	0.918*	0.944	1.547***	0.967	1.023
DM stat.	{1.780}	{0.816}	{−3.281}	{1.252}	{−0.147}
p-value	(0.080)	(0.417)	(0.002)	(0.215)	(0.884)
Model 3					
Dir. chg.	0.361	0.377	0.197**	0.885***	0.721
RMSE ratio	1.002*	1.085	1.601***	0.899	1.100
DM stat.	{−0.032}	{−0.594}	{−5.258}	{2.546}	{−0.415}
p-value	(0.975)	(0.555)	(0.000)	(0.013)	(0.679)

Notes: The direction of change statistic is the proportion of correct 36-month ahead forecast changes in 61 months. The RMSE ratio is the ratio of the structural model RMSE to that of the driftless random walk. The entries in the braces are for the Diebold–Mariano (1993) test statistic for the null hypothesis of equality of forecast RMSE and driftless random walk RMSE. The statistic is asymptotically distributed as a standard normal; p-values in parentheses are for a two-tail test. The statistic uses a Newey–West covariance estimator with a truncation lag of 35.
*(**)[***] denotes significance at the 10(5)[1] percent level.

normal; see Diebold and Mariano (1993) for a derivation and discussion of the small sample properties of the statistic.[14]

Our additional test of the null hypothesis of no forecasting ability compares the observed proportion of successes with the expected number under pure chance (50 percent), again using the Diebold and Mariano (1993) procedure to account for overlapping forecasts (and consequent serial correlation) in the difference between observed and expected outcomes.

To generate $k = 12$-, 24- and 36-month ahead forecasts of exchange rates

[14] The average difference, call it d, between two forecast evaluation metrics divided by an estimate of the standard error of the difference is asymptotically distributed as a standard normal. The variance of the statistic is consistently estimated by an appropriately weighted sum of the available sample autocovariances of the d series. Following Diebold and Mariano (1993) we use the Bartlett weight scheme suggested by Newey and West (1987). Finally, West (1994) shows that one can ignore (asymptotically) the sampling variability of estimated coefficients in RMSE comparisons using the Diebold and Mariano (1993) statistic, only when the prediction error is uncorrelated (in population) with the predictor variables. This condition is satisfied by our reduced-form model (5); it would not be satisfied by model forecasts that were generated using the actual future values of predictor variables.

we use recursive OLS estimates of Eq. (4). Over the 12- and 24-month horizons, there is little evidence of improvement vis-à-vis the random walk. However at the 36-month horizon, the direction-of-change statistics indicate forecasting improvement over the no-change alternative for all three models of the \$/¥ rate. The \$/£ rate direction of change statistics are significantly worse than the no-change alternative for all three of our models.

The equality of RMSE statistics for the \$/¥ rate have p-values of 0.1 percent, 21.5 percent, and 1.3 percent for models 1, 2, and 3, respectively. For the \$/Dm rate, models 1 and 2 have equality of RMSE statistic p-values of around 10 percent. The favorable in-sample results for the \$/C\$ and Dm/¥ rates are not corroborated by our out-of-sample forecasting experiments.

6. Concluding remarks

In this paper we confirm the poor short-term forecasting performance of three structural exchange rate models using both parametric and non-parametric estimation techniques. Our results obtain even after imposing economic structure on the models using error correction terms.

For long-term forecast horizons our results are slightly more positive. We corroborate and extend the results in Mark (1992) and Meese and Rogoff (1983), where some fundamental models are shown to have more explanatory power than the random walk model at prediction horizons beyond two years. Our positive results obtain for some (but not all) of our models of the German mark and Japanese yen exchange rates relative to the U.S. dollar. As usual, positive in-sample results must be interpreted cautiously. Out-of-sample forecasting test statistics rarely exhibit the same degree of statistical significance (relative to the naive random walk model) as in-sample tests.

Data appendix

The data are seasonally unadjusted monthly data, drawn from *OECD Main Economic Indicators* (*MEI*) 1988 disks up to the end of 1988, and various issues of *MEI* thereafter. U.K. M1 data are drawn from *MEI* and U.K. Central Statistical Office *Financial Statistics*. Cumulated trade balance data are based on Frankel's (1984) calculations of wealth. All data are period averages, except for spot exchange rates, which are end of period. A more detailed description is contained in the working paper version.

References

Akaike, H., 1973, Information theory and an extension of the maximum likelihood principle, in: B. N. Petrov and F. Csaki, eds., 2nd International Symposium on Information Theory (Akademini Kiado, Budapest), 267–281.

Balassa, B., 1964, The purchasing power parity doctrine: A reappraisal, Journal of Political Economy 72, 584–596.

Bilson, J., 1978, The monetary approach to the exchange rate: Some evidence, IMF Staff Papers 25, 503–528.

Blundell-Wignall, A. and F. Browne, 1991, Increasing financial market integration, real exchange rates, and macroeconomic adjustment, Working Paper, Economics and Statistic Department, OECD, Paris.

Boughton, J., 1991, Long run money demand in large industrial countries, IMF Staff Papers 38, 1–32.

Cheung, Y., 1992, Long memory in foreign exchange rates, Journal of Business and Economic Statistics 11, 103–112.

Chinn, M., 1993, Estimation of exchange rate and consumption functions when some agents follow rules-of-thumb, Discussion Paper, Department of Economics, The University of California, Santa Cruz.

Clements, K. and J. Frenkel, 1980, Exchange rates, money, and relative Prices: The dollar–pound in the 1920s, Journal of International Economics 10, 249–262.

Cleveland, W. S. and S. J. Devlin, 1988, Locally weighted regression: An approach to regression analysis by local fitting, Journal of the American Statistical Association 83, 596–610.

Diebold, F. X., 1988, Empirical modeling of exchange rate dynamics (Springer-Verlag, New York).

Diebold, F. X. and R. S. Mariano, 1993, Comparing predictive accuracy, Working Paper, Department of Economics, The University of Pennsylvania.

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

Dornbusch, R., 1976b, The theory of flexible exchange rate regimes and macroeconomic policy, in: J. Frenkel and H. Johnson, eds., The economics of exchange rates (Addison-Wesley, Reading).

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, 689–713.

Engle, R. F. and C. W. J. Granger, 1987, Co-integration and error correction: Representation, estimation and testing, Econometrica 55, 252–276.

Faust, J., 1993, Near observational equivalence and unit root processes: Formal concepts and implications, International Finance Discussion Paper 447, Board of Governors of the Federal Reserve.

Finn, M., 1989, An econometric analysis of the intertemporal general-equilibrium approach to exchange rate and current account determination, Journal of International Money and Finance 8, 467–486.

Frankel, J., 1979, On the mark: A theory of floating exchange rates based on real interest differentials, American Economic Review 69, 610–622.

Frankel, J., 1982, The mystery of the multiplying marks: A modification of the monetary model, Review of Economics and Statistics 64, 515–519.

Frankel, J., 1984, Tests of monetary and portfolio balance models of exchange rate determination, in: J. Bilson and R. Marston, eds., Exchange rate theory and practice (University of Chicago Press, Chicago).

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

Granger, C. W.J. and P. Newbold, 1977, Forecasting economic time series (Academic Press, New York).

Hooper, P. and J. Morton, 1982, Fluctuations in the dollar: A model of nominal and real exchange rate determination, Journal of International Money and Finance 1, 39-56.

Hsieh, D., 1989, Testing for nonlinear dependence in daily foreign exchange rates, Journal of Business 62, 339-368.

Johanssen, S. and K. Juselius, 1990, Maximum likelihood estimation and inference on cointegration—with applications to the demand for money, Oxford Bulletin of Economics and Statistics 52, 169-210.

Krugman, P., 1991, Target zones and exchange rate dynamics, Quarterly Journal of Economics 106, 669-682.

Kwiatkowski, D., P. C. B. Phillips, P. Schmidt and Y. Shin, 1992, Testing the null hypothesis of stationarity against the alternative of a unit root, Journal of Econometrics 54, 159-178.

Lucas, R. E., 1982, Interest rates and currency prices in a two-country world, Journal of Monetary Economics 10, 335-360.

Mark, N. C., 1992, Exchange rates and fundamentals: Evidence on long-horizon predictability and overshooting, Working Paper, The Ohio State University.

Marston, R., 1987, Real exchange rates and productivity growth in the United States and Japan, in: S. Arndt and J. D. Richardson, eds., Real-financial linkages among open economies (MIT Press, Cambridge).

Meese, R. A., 1990, Currency fluctuations in the post Bretton Woods era, Journal of Economic Perspectives 4, 117-134.

Meese, R. and K. Rogoff, 1983, The out-of-sample failure of empirical exchange rate models: Sampling error or misspecification?, in: J. Frenkel, ed., Exchange rates and international macroeconomics (University of Chicago Press, Chicago) 67-105.

Meese, R. and A. K. Rose, 1991, An empirical assessment of nonlinearities in models of exchange rate determination, Review of Economic Studies 58, 603-619.

Mussa, M., 1976, The exchange rate and the balance of payments, Scandinavian Journal of Economics 78, 229-248.

Newey, N. K. and K. D. West, 1987, A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, Econometrica 55, 703-708.

Schwert, G. W., 1989, Tests for unit roots: A Monte Carlo investigation, Journal of Business and Economic Statistics 7, 147-159.

Stock, J., 1987, Measuring business cycle time, Journal of Political Economy 95, 1240-1261.

Stock, J. and M. Watson, 1993, A simple estimator of cointegrating vectors in higher order integrated systems, Econometrica 61, 783-820.

West, K. D., 1994, Asymptotic inference about predictive ability, Working Paper, The University of Wisconsin.

Wolff, C. C. P., 1987, Time-varying parameters and the out-of-sample forecasting performance of structural exchange rate models, Journal of Business and Economic Statistics 5, 87-97.