Journal of International Money and Finance (1984), **3**, 357–368

Testing the Unbiasedness Hypothesis in the Forward Foreign Exchange Market: A Specification Analysis

ALLAN W. GREGORY

Department of Economics, University of Western Ontario, London, Ont., Canada N6A 5C2

AND

THOMAS H. MCCURDY*

Department of Economics, Queen's University, Kingston, Ont., Canada K7L 3N6

This paper evaluates two popular regression methods of testing the unbiasedness hypothesis in the forward foreign exchange market. For the 30-day Canada/United States forward foreign exchange market, the evidence overwhelmingly indicates that it is inappropriate to treat the structure of the systematic and stochastic components of the test relations as constant over time. Hence, conclusions inferred from parameter significance testing based upon full-sample estimation can be very misleading. Accordingly, we argue for a specification analysis of the test relations, and more explicit modelling of market fundamentals.

The widespread adoption of flexible exchange rates in the early 1970s has led to an intensive examination of the efficiency of foreign exchange markets. According to Fama (1970), a market is efficient if prices always 'fully reflect' available information so that they provide accurate signals for resource allocation.[1] The extent to which a price aggregates structural information is particularly important for economy-wide markets such as that for foreign exchange. In the case of the forward foreign exchange market, the forward rate should 'summarize' all the current information which is relevant for predicting the equilibrium future spot rate. Thus the efficiency of the forward market is important not only from an allocative perspective, but also in evaluating the usefulness of forward rates as proxies for spot rate expectations.

While there is general agreement regarding the importance of testing foreign

* The authors would like to thank David Backus, David Longworth, Michael Veall and two referees for helpful comments. The financial support of the Social Sciences and Humanities Research Council of Canada and the Advisory Research Committee of Queen's University is acknowledged.

0261-5606/84/03/0357-12$03.00

exchange market efficiency (the large volume of research would seem to confirm this), there appears to be no widely accepted consensus as to what the appropriate tests are. The absence of a generally agreed upon formulation of the problem has given rise to an assortment of tests which relate to the issue of market efficiency. Most of these tests involve regression analysis. Although the regression framework has provided a convenient structure from which hypothesis tests may be entertained, there are several requirements for valid hypothesis testing that have often been overlooked in application.

The primary purpose of this paper is to illustrate the need for a specification analysis of the test relation. The applied research into market efficiency has concentrated almost exclusively upon parameter significance testing, and very little attention has been given to the statistical adequacy of the test relation itself.[2] This neglect seems to run counter to good econometric practice. That is, before undertaking specific parameter tests, it is important to establish that the postulated test relation is in some sense statistically appropriate. Otherwise, misleading or incorrect conclusions can always occur. In this light, several diagnostic tests, as well as subsample estimation, are especially useful for such specification evaluation (see Pagan and Hall, 1983).

In the present study, Canadian/United States data are used to evaluate the empirical specification of two alternative regression equations which have been extensively used to test the composite null hypothesis of rational expectations and no risk premium in the 30-day forward foreign exchange market. This composite hypothesis, which implies that the forward rate is an optimal predictor of the corresponding future spot exchange rate, is referred to as the unbiasedness hypothesis.[3]

For this data set, specification tests and subsample estimation display a variety of econometric deficiencies. For example, investigation of the robustness of conclusions to subsampling indicates that the results are extremely sensitive to the time period used in estimation. This, of course, implies that hypothesis tests from full-sample estimation can be unreliable. Diagnostic tests reveal that the 'ills' associated with the test relation are not the same over time. For one period there may be one type of specification error and for another period a different one. Although there are many possible explanations for this (such as an omitted variable), our results suggest that more than a 'quick' remedy will be required. In this regard, the current trend towards modelling the underlying market fundamentals would appear worthwhile. Nevertheless, the evidence presented here implies that a careful reassessment of the regression-based tests of the unbiasedness hypothesis is warranted.

In Section I, two alternative methods of testing the unbiasedness hypothesis are compared, particularly with respect to the implications of potential misspecification. In addition, the relationship between the unbiasedness hypothesis and the issue of market efficiency is briefly reviewed. Section II discusses the econometric issues and reports the empirical results. Concluding remarks are given in Section III.

I. Forward Rates as Optimal Predictors of Future Spot Rates

Typically, the optimality of the forward rate as a predictor of the corresponding future spot rate has been tested in the context of regression analysis. It is useful to discuss two regression specifications which have been commonly applied: (i) tests of the unbiasedness of the forward premium (or discount) as a linear predictor of the change in the spot rate; and (ii) orthogonality tests of the forecast errors to various information sets.

I.A. Forward Rates as Unbiased Linear Predictors

Consider the regression equation:

$$\langle 1 \rangle \qquad DS_{t+1} = \alpha + \beta FP_t + u_{t+1},$$

in which

$$DS_{t+1} \equiv \left(\frac{S_{t+1} - S_t}{S_t}\right) \quad \text{and} \quad FP_t \equiv \left(\frac{F_t - S_t}{S_t}\right),$$

where S_{t+1} is the spot exchange rate at time $t+1$, F_t is the forward rate established at time t for period $t+1$, and u_{t+1} is an error term. The hypothesis that the normalized forward premium is an unbiased linear predictor of the rate of change of the corresponding spot rate implies that $\alpha=0$ and $\beta=1$.

The unbiasedness hypothesis could be derived from a model of a competitive market which has no transactions costs, risk-neutral speculators and market expectations which are rational. In that case:

$$\langle 2 \rangle \qquad E_t[DS_{t+1}] = FP_t,$$

where E_t is the mathematical expectation operator conditional upon current information I_t (which includes the hypothesized structural model). The relationship between the test equation ⟨1⟩ and the joint null hypothesis of rational expectations and no risk premium implied by ⟨2⟩ can be seen by decomposing the actual change in the spot rate into two orthogonal components:

$$\langle 3 \rangle \qquad DS_{t+1} = E_t[DS_{t+1}] + (DS_{t+1} - E_t[DS_{t+1}]).$$

Then, using ⟨2⟩, ⟨3⟩ yields ⟨1⟩ under the null. Hence, the interest in the applied literature has focused upon estimating equation ⟨1⟩ and testing whether the estimated coefficients of α and β are significantly different from zero and one respectively.

As is well known, since ⟨2⟩ implies that there is no risk premium or significant transactions costs, a rejection of the unbiasedness hypothesis using ⟨1⟩ does not necessarily imply rejection of the rational expectations hypothesis (REH) or consequently of the efficient markets hypothesis (EMH). The omission of transactions costs is common to most regression-based tests. However, there has been considerable effort recently to investigate the alternative hypothesis that there is a risk premium associated with forward contacts for foreign exchange. While there is an extensive theoretical literature concerning the existence of a risk premium which may be time varying, estimation has typically proceeded under the assumption that it is time invariant.[4] In contrast, Domowitz and Hakkio (1983a), Frankel (1982), Hansen and Hodrick (1983), and Hodrick and Srivastava (1984) have attempted to relax this assumption. However, if the analysis is to be conducted on the basis of the estimated coefficients of equation ⟨1⟩, then some indication should be provided that the test relation is in fact stable. For the most part, the applied literature has neglected these considerations.

If we assume that $I_t \supset I_{t-1} \supset \ldots$, then, under the REH and the EMH, the residual in $\langle 1 \rangle$ should be serially independent.[5] However, any misspecification such as parameter instability or omitted variables (say due to a time-varying risk premium) may cause the error term to be correlated. Earlier studies, which were preoccupied with parameter significance testing, generally failed to give sufficient attention to the error structure. Estimation proceeded on the assumption that the errors were serially independent with very little evidence given to support such a claim. A further shortcoming that characterized these earlier studies was the assumption that the variance of u_{t+1} was constant. As observed by several authors (see, for example, Cumby and Obstfeld, 1983, Domowitz and Hakkio, 1983a, Hodrick and Srivastava, 1984, and Hsieh, 1982), this is not implied by the unbiasedness hypothesis. Certainly, any auxiliary assumptions such as homoscedastic errors should also be tested. A critical part in evaluating any econometric specification is an examination of the properties of the underlying error terms.

These arguments suggest that a comprehensive investigation of both the systematic and stochastic structure of equation $\langle 1 \rangle$ is required. In Section II we outline the econometric strategy followed to evaluate test relation $\langle 1 \rangle$.

I.B. Orthogonality Tests

The orthogonality tests of F_t as an optimal predictor of S_{t+1} examine whether information available at time t can be 'used to explain' the forecast error (see, for example, Fama, 1970, Frankel, 1980, Geweke and Feige, 1979, and Hansen and Hodrick, 1980). For this approach, the (normalized) forcast error is regressed upon some subset of the current information set I_t, for example, the $(1 \times k)$ row vector X_t:

$$\langle 4 \rangle \qquad \frac{S_{t+1} - F_t}{S_t} = X_t \phi + w_{t+1},$$

in which ϕ is a $(k \times 1)$ vector of unknown parameters and w_{t+1} is an error term. Testing the null hypothesis involves testing whether any elements of the estimated column vector ϕ are significantly different from zero.[6] If so, then the forecast error is correlated with information which was available when the forecast was formulated. On the basis of these kinds of tests for the forward foreign exchange market, the null hypothesis that $\phi = 0$ has generally been rejected.

An important consideration of this testing procedure is the interpretation of a rejection of the hypothesis that $\phi = 0$. A rejection of this hypothesis has sometimes been interpreted as foreign exchange market inefficiency since there was information available which rational agents could have 'used' to predict the forecast error (that is, there is a systematic portion to this error based upon current information). Clearly, as is well known, there are alternative hypotheses—for example, the null could be rejected because of the existence of a risk premium which is being proxied by some element of X_t.

The most obvious weakness in conducting the orthogonality tests is that the model underlying the determination of S_{t+1} and F_t is not specified. Therefore, there is an element of arbitrariness about what is to be included in X_t. The choice is left to the researcher to select which variables are 'reasonable'. For example, if X_t contains only the past history of the dependent variable then the test of the null hypothesis that $\phi = 0$ is said to be a weak-form test. Expanding X_t to include publicly available information is called a semi-strong form test, and allowing X_t to contain inside or restricted information is a strong-form test. Thus, since the choice of which variables to include in X_t is arbitrary, there are really no limits placed upon the number of possible variable combinations which may be selected and the number of regression equations which may be estimated. Therefore, with finite data sets (and consequently a given number of forecast error observations) it is invariably possible to find, *ex post*, some information set X_t which is correlated with the forecast error and yields significant t-statistics in a regression equation.[7] This in itself is not sufficient evidence to reject the null hypothesis. It is necessary to obtain a systematic and stable relation between the forecast error and some information set. With this in mind, it is important to present evidence that the empirical specification of a test relation is adequate to test the null hypothesis.

In regard to the stochastic structure of the test relation, there is no reason to presume that the errors of ⟨4⟩ are homoscedastic. However, under the null hypothesis, the forecast error is orthogonal to all available information so that the disturbance terms should be serially independent. A complete investigation would involve testing such restrictions.

In sum, the absence of economic modelling together with the omission of a specification analysis of the test relation would seem to suggest that the results from equation ⟨4⟩ could be viewed more accurately as after-the-fact correlation analysis. As such, the test relation may be particularly susceptible to misspecification, due to missing variables, for example. In that case, hypothesis tests would not be valid.

II. The Data, Specification Tests, and Results

The four-weekly data are obtained from a daily data tape that has been kindly supplied by the Bank of Canada. The forward rate F_t and the spot rate S_t^T are Tuesday closing rates while S_{t+1} is the Thursday closing spot rate four weeks and two business days in the future. This ensures that there are 30 days between F_t and S_{t+1}, while at the same time the forward premium and the lagged forecast error contain rates which are available to market participants. If, instead, those variables had been constructed using the Thursday spot rate, we would be testing the market's performance on the basis of information which market participants could not have known when the forward contracts are made. It is important to conduct the tests on the basis of information actually available to the market. The period of the study is from 1973 to 1981 inclusive which gives a full-sample size of 117, four-weekly, observations.

The hypothesis that the (normalized) forward premium is an unbiased linear predictor of the rate of change of the corresonding spot rate was tested using the regression equation:

$$\langle 1' \rangle \qquad \frac{S_{t+1}-S_t^T}{S_t^T} = \alpha+\beta\left(\frac{F_t-S_t^T}{S_t^T}\right)+u_{t+1}.$$

For the orthogonality tests we used a version of ⟨4⟩ which is similar to that used by Hansen and Hodrick (1980, 1983). That is, the X_t of ⟨4⟩ contains the lagged forecast error and the known forward premium:

$$\langle 4' \rangle \qquad \frac{S_{t+1}-F_t}{S_t^T} = \phi_0+\phi_1\left(\frac{S_t^T-F_{t-1}}{S_t^T}\right)+\phi_2\left(\frac{F_t-S_t^T}{S_t^T}\right)+w_{t+1}.$$

If a well-defined theory is available, it is often possible to release that theory from many restrictions it otherwise imposes on the data and then seek to establish, one by one in an orderly way, whether the restrictions are 'acceptable' to the data. This is a powerful and appealing procedure since estimation is invariably under the unrestricted form of the model (the alternative hypothesis), and hence, in this sense is prejudiced against the null. Failure to reject the null is then seen as a strong achievement. The methodology of this approach is to begin by considering the largest scale of the model consistent with the weakest or most 'open' form of the theory.

For the problem of testing unbiasedness of the forward rate or orthogonality of forecast errors to various information sets, it is not obvious what the most 'open' form of the theory is. Since the underlying behavioral theory determining the spot and forward foreign exchange rate is not specified, relaxing the 'theory' for this testing exercise is tantamount to introducing arbitrary explanatory variables with arbitrary lag structures. Such a procedure would inevitably produce an endless series of results with little or no interpretative value. A more meaningful exercise, in the spirit of relaxing restrictions of the theory, is to examine the behavior of the estimated test relations over time. This investigation entails a detailed specification analysis of the equation over various subperiods and should provide evidence of structural change in either the systematic or stochastic components.

For each subperiod we calculate a number of diagnostic tests. Although each test is designed primarily to detect a particular specification error, it may, nevertheless, pick up other kinds of problems (see Pagan and Hall, 1983). For each diagnostic test a marginal significance level is calculated without making any allowance for the fact that many tests are being jointly considered. Pagan and Hall (1983) discuss in some detail the necessary conditions under which the diagnostic tests are independent. For instance, many of the tests used in the present paper are additive under the assumption of normality and the absence of a lagged dependent variable. Unfortunately for our purposes, there is no reason to expect that the errors are normally distributed. Also, the forward premium could be regarded as a lagged dependent variable. Therefore, given the complexity of determining the joint probabilities for all the tests considered, we treat each diagnostic test as if it is calculated in isolation.

Perhaps the simplest kinds of diagnostic tests to calculate are those based upon the Lagrange multiplier principle of testing. The test relations $\langle 1' \rangle$ and $\langle 4' \rangle$ can be estimated by ordinary least squares and various tests can be constructed to check the restrictions imposed in estimation. For each subperiod, we calculate: (i) a test for fourth order serial correlation (AUTO) outlined in Godfrey (1978); (ii) a test for fourth order serial correlation which is valid in the presence of heteroscedasticity (AUTO_{H}) developed in Domowitz and Hakkio (1983b); (iii) a general test for heteroscedasticity (H) from White (1980); (iv) a test for fourth order autoregressive conditional heteroscedasticity (ARCH) derived in Engle (1982); (v) the Information matrix test (INFO) of White (1982) based upon a calculation suggested by Chesher (1983); and (vi) Ramsey's (1969) RESET text using the square of the fitted values.[8] The information matrix test may be interpreted as a test for parameter constancy (Chesher, 1984) and the RESET test is a general test for misspecification (Pagan and Hall, 1983). Each test statistic is compared against the chi-square distribution.

In the case of testing the unbiasedness hypothesis using equation $\langle 1' \rangle$ or $\langle 4' \rangle$, it is not obvious which subperiods to select. Originally our intention was to approach the problem of structural change in a classical fashion by attempting to isolate various stable regimes. That approach required that we select particular points of time to investigate structural instability. For Canada there were several obvious points such as the adoption of M1 targets in September 1975, the introduction of the anti-inflation policy of October 1975, the dramatic depreciation of the Canadian dollar that began with the election of the Parti-Quebecois in Quebec in November 1976, and the credit crunch in mid-1981. For the United States, an often cited event is the October 1979 change in the Federal Reserve Board operating procedure. However, it is not necessary that structural change be associated with any particular 'big' event since it is certainly possible to have an on-going (albeit unpredictable) process of structural change. Thus, in principle, it is difficult, *a priori*, to justify excluding any period. As a practical matter, we have chosen a selection of our results using time periods based upon the calendar year. Although these subperiods are arbitrary, we believe that the results adequately portray the issues and difficulties in testing the unbiasedness hypothesis.

TABLE 1. Forward rates as unbiased linear predictors.

Year	α	β	W	AUTO	$AUTO_H$	H	ARCH	INFO	RESET	R^2
1973–1976	0.000272 (0.0017)	−0.563 (0.80)	0.135	0.0315	0.0902	0.205	0.0485	0.000403	0.298	0.009
1974–1977	0.00282 (0.0024)	−0.477 (1.097)	0.366	0.0709	0.00329	0.248	0.287	0.0591	0.333	0.004
1975–1978	0.00661 (0.0027)	−1.853 (1.27)	0.0387	0.225	0.142	0.355	0.900	0.0907	0.755	0.04
1976–1979	0.00387 (0.0026)	−0.704 (1.29)	0.295	0.441	0.178	0.439	0.770	0.218	0.384	0.006
1977–1980	0.00350 (0.0018)	0.206 (1.12)	0.135	0.705	0.389	0.400	0.287	0.317	0.790	0.0007
1978–1981	0.00134 (0.0019)	−1.931 (1.07)	0.0219	0.297	0.290	0.129	0.039	0.0711	0.150	0.06
1973–1978	0.00254 (0.0015)	−0.245 (0.78)	0.137	0.0713	0.0241	0.0796	0.531	0.00319	0.243	0.001
1976–1981	0.00282 (0.0019)	−1.078 (0.88)	0.0470	0.407	0.150	0.182	0.574	0.0405	0.215	0.02
1973–1981	0.00173 (0.0012)	−0.756 (0.63)	0.0171	0.235	0.094	0.028	0.284	0.00100	0.186	0.01

Note:
Standard errors are given in parentheses. W is the Wald test of the joint hypothesis that $\alpha=0$ and $\beta=1$, AUTO is a test for fourth order serial correlation (Godfrey, 1978), $AUTO_H$ is a heteroscedasticity-robust test for fourth order serial correlation (Domowitz and Hakkio, 1983b), H is a general test of heteroscedasticity (White, 1980). ARCH is a test for fourth order conditional heteroscedasticity (Engle, 1982), INFO is the information matrix test (White, 1982) based upon the calculation of Chesher (1983), RESET is Ramsey's (1969) misspecification test using the square of the fitted values and R^2 is the coefficient of determination. Marginal significance levels are reported for each statistic.

We shall commence with a discussion of the results concerning the forward premium as an unbiased linear predictor of the future depreciation of the spot rate. In Table 1 we present the result of estimating equation ⟨1′⟩ by ordinary least squares for several different (overlapping) subperiods over the period 1973–1981. Features of these results are:

(i) Less than 6% of the variation in actual spot rate changes is explained for all the subsamples.
(ii) There is a considerable movement in the estimated coefficients of α and β and their respective standard errors. This is dramatically reflected by the fact that the Wald (W) test of the joint hypothesis that $\alpha=0$ and $\beta=1$ can either be rejected or retained at the 5% level of confidence by simply adding and subtracting one year of observations in either direction (rejected for 1975–1978 but retained for 1976–1979 and 1974–1977).
(iii) Tests for serial correlation AUTO and $\mathrm{AUTO_H}$ (robust to heteroscedasticity) indicate that the hypothesis of serial independence is rejected only for the earlier part of the sample.
(iv) At the 5% level, using White's (1980) H test, the null hypothesis of homoscedastic errors is rejected for the whole sample but is not rejected for any subperiod.
(v) ARCH errors occur at the beginning and end of the sample period but appear to be absent in the middle years and for full sample estimation.[9]
(vi) Interpreting the information matrix tests as a test for parameter constancy implies a rejection of parameter stability for the larger data sets (1973–1978, 1976–1981, and 1973–1981).
(vii) According to the RESET test there is no evidence of misspecification for any estimation period.

The tendency for the information matrix test to be rejected for the larger data sets, combined with the variable results concerning the validity of the joint null hypothesis that $\alpha=0$ and $\beta=1$, provides strong evidence against the constant structure assumption in the test relation. From a statistical point of view, the estimating equation is satisfactory for the period 1976–1980. In fact, searching over subperiods which use the policy events mentioned above as boundary dates, the period October 1975 to June 1981 'passes' all of the tests for misspecification and, interestingly enough, this corresponds exactly to a period of time for which the null hypothesis of unbiasedness of the forward premium cannot be rejected.

In Table 2 the results for the orthogonality tests appear. To summarize our findings:

(i) The coefficient of determination is again very low, with less than 15% of the variation being explained.
(ii) There are substantial fluctuations in the estimated coefficients and their standard errors, as illustrated by the changing conclusions obtained from the Wald (W) test that all the coefficients are jointly zero.
(iii) For the most part, AUTO and $\mathrm{AUTO_H}$ tests strongly reject the hypothesis of serial independence.
(iv) The H and ARCH tests do not reject the hypothesis of constant variance. (Although these tests are not valid in the presence of serial correlation, see Domowitz and Hakkio, 1983b.)

TABLE 2. Tests for orthogonality of forecast errors.

Year	ϕ_0	ϕ_1	ϕ_2	W	AUTO	$AUTO_H$	H	ARCH	INFO	RESET	R^2
1973–1976	0.000420 (0.0017)	−0.207 (0.14)	−1.751 (0.80)	0.101	0.018	0.003	0.749	0.337	0.003	0.086	0.11
1974–1977	0.00311 (0.0024)	−0.150 (0.14)	−1.632 (1.10)	0.212	0.006	0.000002	0.799	0.886	0.070	0.030	0.06
1975–1978	0.00752 (0.0027)	−0.221 (0.13)	−3.210 (1.26)	0.022	0.038	0.008	0.836	0.910	0.061	0.282	0.14
1976–1979	0.00439 (0.0026)	−0.199 (0.13)	−1.905 (1.28)	0.190	0.058	0.001	0.927	0.886	0.484	0.151	0.08
1977–1980	0.00367 (0.0018)	−0.0627 (0.14)	−0.964 (1.19)	0.245	0.135	0.087	0.810	0.257	0.676	0.968	0.01
1978–1981	0.00184 (0.0018)	−0.318 (0.13)	−3.977 (1.11)	0.002	0.124	0.021	0.475	0.131	0.096	0.010	0.22
1973–1978	0.00287 (0.0015)	−0.163 (0.11)	−1.408 (0.786)	0.100	0.001	0.0000007	0.433	0.828	0.003	0.057	0.06
1976–1981	0.0036 (0.0018)	−0.257 (0.11)	−2.606 (0.86)	0.007	0.006	0.00006	0.542	0.683	0.080	0.080	0.14
1973–1981	0.00212 (0.0012)	−0.210 (0.091)	−2.143 (0.64)	0.003	0.0006	0.00001	0.185	0.506	0.005	0.030	0.10

Note:
Standard errors are given in parentheses. W is the Wald test of the joint hypothesis that $\phi_0=\phi_1=\phi_2=0$, AUTO is a test for fourth order serial correlation (Godfrey, 1978), $AUTO_H$ is a heteroscedasticity-robust test for fourth order serial correlation (Domowitz and Hakkio, 1983b), H is a general test of heteroscedasticity (White, 1980), ARCH is a test for fourth order conditional heteroscedasticity (Engle, 1982), INFO is the information matrix test (White, 1982) based upon the calculation of Chesher (1983), RESET is Ramsey's (1969) misspecification test using the square of the fitted values and R^2 is the coefficient of determination. Marginal significance levels are reported for each statistic.

(v) Both the INFO and RESET tests imply that it is not appropriate to combine the subsamples.

As before, the test relation for the 1977–1980 period appears to be statistically adequate, and again for this period the unbiasedness hypothesis that all coefficients are not significantly different from zero cannot be rejected. Thus the general tendency for both regression-based tests is that whenever the test relation appears to be statistically appropriate, we cannot reject the unbiasedness hypothesis.

III. Conclusion

The purpose of this paper has been to evaluate critically and empirically two popular regression methods of testing the unbiasedness hypothesis for the forward foreign exchange market. An important consideration in this investigation is a specification analysis of the test relations themselves. The evidence based upon Canadian and American 30-day forward foreign exchange data overwhelmingly indicates that it is inappropriate to treat the structure of the systematic and stochastic components of the test relations as constant over time. Diagnostic tests reveal that the 'ills' associated with the test relations vary from period to period. Since conclusions are not robust to subsampling, parameter significance tests on the basis of full-sample estimation can be very misleading.

Tests for unbiasedness using $\langle 1' \rangle$ show that there are periods of time (for example, October 1975 to June 1981) for which it is possible to characterize the 30-day forward exchange rate premium as an unbiased linear predictor of the corresponding rate of change in the spot rate. However, there are also subperiods for which this hypothesis is strongly rejected (although for these periods there is strong evidence that the equation is misspecified). If we are going to use an equation such as $\langle 1' \rangle$ to test the relationship between the forward premium and the rate of change in the spot rate, it is clear that we should not impose, *a priori*, an invariant structure on the test relation.

Results from diagnostic tests on the orthogonality test relation $\langle 4' \rangle$ provide ample evidence that the equation is misspecified. The proposition that the estimated coefficients are unstable seems to be well supported by the information matrix tests. Perhaps it is not too surprising that for any arbitrary choice of 'explanatory' variables the resulting parameter estimates are unstable. Obtaining significant t-statistics in a misspecified regression equation does not necessarily imply evidence against (or for) the unbiasedness hypothesis. The estimated coefficients and standard errors may be inconsistent, rendering such parameter significance testing meaningless. Therefore, without a specification analysis of the test relation itself, such findings as significant t-statistics can hardly be compelling.

One possible explanation of the apparent relation between the forecast error and some elements in X_t is the existence of a time-varying risk premium. Hence, some variables in X_t may be correlated with the risk premium giving rise to a misspecified test relation. Accordingly, we advise explicit modelling of market fundamentals in order to determine the relevant information set and to derive specific testable hypotheses. Recently, some authors (Domowitz and Hakkio, 1983a, Frankel, 1982, Hansen and Hodrick, 1983, and Hodrick and Srivastava, 1984) have moved in this direction by testing various hypotheses concerning the existence of a time-varying risk premium. Nevertheless, their empirical results suggest that the risk premia they have modelled may not account for all the misspecification of test relations such as $\langle 1 \rangle$ and $\langle 4 \rangle$.

Finally, an alternative method of investigating the unbiasedness or 'speculative efficiency' hypothesis without imposing, *a priori*, a particular hypothesis about risk, has been to compute the out-of-sample risk-return tradeoff of filter rules (Dooley and Shafer, 1976, 1983) or trading strategies (Bilson, 1981, Hodrick and Srivastava, 1984). This also appears to be a fruitful direction of research to investigate the more general issue of market efficiency.[10] One reason is that the consequences of misspecification, such as parameter instability, are not relevant for this method.

Notes

1. For a discussion of the relation between informational efficiency and allocative efficiency see, for example, Bray (1981), Grossman (1976), Grossman and Stiglitz (1976) and Harris and Purvis (1981, 1982).
2. Recent exceptions are Domowitz and Hakkio (1983a,b) and Hodrick and Srivastava (1984).

3. This has also been called the 'simple efficiency' hypothesis (Hansen and Hodrick, 1980) and the 'speculative efficiency' hypothesis (Bilson, 1981).
4. In which case α in equation $\langle 1 \rangle$ has usually been interpreted as capturing the time-invariant risk premium. Frenkel and Razin (1980) provide theoretical evidence that a stochastic price level prevents that interpretation, while Stein (1980) and Fama (1983) give theoretical reasons for the risk premium also to affect β in $\langle 1 \rangle$.
5. With respect to the particular data set used in this study (see Section II), we note that first order serial correlation could be introducted by the fact that the forward rates are measured on Tuesday and the corresponding future spot rates are from Thursday. However, the test results indicate that whenever serial correlation is present the process is greater than order one.
6. Notice that $\langle 1 \rangle$ and $\langle 4 \rangle$ are equivalent under the null hypothesis of rational expectations and no risk premium—$\alpha=0$, $\beta=1$, and $\phi=0$ respectively. Nevertheless, the test relations $\langle 1 \rangle$ and $\langle 4 \rangle$ are different—in particular, the dependent variables are not the same.
7. Asymptotically, if the null hypothesis is true so that the forecast errors are orthogonal to all available information, the estimated ϕ should not be significantly different from zero. However, as a practical matter, in order to carry out significance testing of the estimated ϕ, we again must first obtain an econometrically satisfactory specification of the unrestricted equation $\langle 4 \rangle$. The results presented below suggest that it is not always that easy to find such a relation.
8. Tests for first through third order AUTO, AUTO_H and ARCH were also calculated (available from the authors upon request). For the most part, test statistics using these orders produced marginal significance levels similar to those based upon a fourth order process. In addition, we found that higher powers of the fitted values for the RESET test were never important.
9. It is interesting to note that, at times, the ARCH test and White's test for heteroscedasticity disagree. This seems to suggest that these tests are sensitive to the form of the heteroscedasticity. In this regard, combining the variables used in the two tests may prove to be worthwhile.
10. A filter rule published by Dooley and Shafer (1976) continued to perform well from 1976 to 1981 (Dooley and Shafer, 1983). Obviously, their rule was chosen out-of-sample. We would like to thank a referee for this reference.

References

Baillie, R.T., R.E Lippens and P.C. McMahon, 'Testing Rational Expectations and Efficiency in the Foreign Exchange Market', *Econometrica*, May 1983, **51:** 553–63.

Bilson, J.F.O., 'The Speculative Efficiency Hypothesis', *Journal of Business*, July 1981, **54:** 435–52.

Bray, M., 'Futures Trading, Rational Expectations, and the Efficient Markets Hypothesis', *Econometrica*, May 1981, **49:** 575–96.

Chesher, A., 'The Information Matrix Test: Simplified Calculation Via a Score Test Interpretation', *Economics Letters*, 1983, **13:** 45–48.

Chesher, A., 'Testing for Neglected Heterogeneity', *Econometrica*, July 1984, **52:** 865–72.

Cumby, R.E. and M. Obstfeld, 'International Interest-Rate and Price-Level Linkages Under Flexible Exchange Rates: A Review of Recent Evidence', in J.F.O. Bilson and R.C. Marston, eds., *Exchange Rates: Theory and Practice*, Chicago: University of Chicago Press, 1983.

Domowitz, I. and C.S. Hakkio, 'Conditional Variance and the Risk Premium in the Foreign Exchange Market', unpublished manuscript, Northwestern University, September 1983. (1983a)

Domowitz, I. and C.S. Hakkio, 'Testing for Serial Correlation in the Presence of Heteroscedasticity with Applications to Exchange Rate Models', unpublished manuscript, Northwestern University, October 1983. (1983b)

Dooley, M.P. and J.R. Shafer, 'Analysis of Short-Run Exchange Rate Behavior: March, 1973 to September, 1975', International Finance Discussion Paper No. 123, Washington: Federal Reserve Board, 1976.

Dooley, M.P. and J.R. Shafer, 'Analysis of Short-Run Exchange Rate Behavior: March, 1973 to November, 1981', in D. Bigman and T. Taya, eds., *Exchange Rate and Trade Instability: Causes, Consequences, and Remedies*, Cambridge, Mass.: Ballinger, 1983.

Engle, R.F., 'Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflations', *Econometrica*, July 1982, **50,** 987–1007.

Fama, E.F., 'Efficient Capital Markets: A Review of Theory and Empirical Work', *Journal of Finance*, May 1970, **25:** 383–417.

Fama, E.F., 'Forward and Spot Exchange Rates', Working Paper No. 112, University of Chicago, December 1983.

FRANKEL, J.A., 'Tests of Rational Expectations in the Forward Exchange Market', *Southern Economic Journal*, April 1980, **46:** 1083–1101.

FRANKEL, J.A., 'In Search of the Exchange Risk Premium: A Six-Currency Test Assuming Mean-Variance Optimization', *Journal of International Money and Finance*, December 1982, **1:** 255–74.

FRENKEL, J.A., 'Flexible Exchange Rates, Prices, and the Role of "News": Lessons from the 1970s', *Journal of Political Economy*, August 1981, **89:** 665–705.

FRENKEL, J.A. AND A. RAZIN, 'Stochastic Prices and Tests of Efficiency of Foreign Exchange Market', *Economics Letters*, 1980, **6:** 165–70.

GEWEKE, J. AND E. FEIGE, 'Some Joint Tests of the Efficiency of Markets for Foreign Exchange', *Review of Economics and Statistics*, August 1979, **61:** 334–41.

GODFREY, L.G., 'Testing for Higher Order Serial Correlation in Regression Equations When the Regressors Include Lagged Dependent Variables', *Econometrica*, November 1978, **46:** 1303–10.

GRAUER, F.L.A., R.H. LITZENBERGER AND R.E. STEHLE, 'Sharing Rules and Equilibrium in an International Capital Market Under Uncertainty', *Journal of Financial Economics*, June 1976, **3:** 233–56.

GROSSMAN, S., 'On the Efficiency of Competitive Stock Markets Where Traders Have Diverse Information', *Journal of Finance*, May 1976, **31:** 573–85.

GROSSMAN, S. AND J. STIGLITZ, 'Information and Competitive Price Systems', *American Economic Review*, May 1976, **66:** 246–53.

HAKKIO, C.S., 'Expectations and the Forward Exchange Rate', *International Economic Review*, October 1981, **22:** 663–78.

HANSEN, L.P. AND R.J. HODRICK, 'Forward Exchange Rates as Optimal Predictors of Future Spot Rates: An Econometric Analysis', *Journal of Political Economy*, October 1980, **88:** 829–53.

HANSEN, L.P. AND R.J. HODRICK, 'Risk Averse Speculation in the Forward Foreign Exchange Market: An Econometric Analysis of Linear Models', in J.A. Frenkel, ed., *Exchange Rates and International Macroeconomics*, Chicago: University of Chicago Press, 1983.

HARRIS, R.G. AND D.D. PURVIS, 'Diverse Information and Market Efficiency in a Monetary Model of the Exchange Rate', *Economic Journal*, December 1981, **91:** 829–47.

HARRIS, R.G. AND D.D. PURVIS, 'Incomplete Information and the Equilibrium Determination of the Forward Exchange Rate', *Journal of International Money and Finance*, December 1982, **1:** 241–53.

HODRICK, R.J. AND S. SRIVASTAVA, 'An Investigation of Risk and Return in Forward Foreign Exchange', *Journal of International Money and Finance*, April 1984, **3:** 5–29.

HSIEH, D.A., 'Tests of Rational Expectations and No Risk Premium in Forward Exchange Markets', NBER Working Paper No. 843, 1982.

LEVICH, R.M., 'On the Efficiency of Markets for Foreign Exchange', in R. Dornbusch and J.A. Frenkel, eds., *International Economic Policy: Theory and Evidence*, Baltimore: The Johns Hopkins University Press, 1979.

LONGWORTH, D., 'Testing the Efficiency of the Canadian-U.S. Exchange Market Under the Assumption of No Risk Premium', *Journal of Finance*, March 1981, **36:** 43–49.

PAGAN, A. AND A.D. HALL, 'Diagnostic Tests as Residual Analysis', *Econometric Reviews*, No. 2, 1983, **2:** 159–218.

RAMSEY, J.B., 'Tests for Specification Errors in Classical Linear Least-Squares Regression Analysis', *Journal of the Royal Statistical Society*, Series B, 1969, **31:** 350–71.

STEIN, J.L., 'The Dynamics of Spot and Forward Prices in an Efficient Foreign Exchange Market with Rational Expectations', *American Economic Review*, September 1980, **70:** 565–83.

STOCKMAN, A.C., 'Risk, Information, and Forward Exchange Rates', in J.A. Frenkel and H.G. Johnson, eds., *The Economics of Exchange Rates: Selected Studies*, Reading: Addison-Wesley, 1978.

WHITE, H., 'A Heteroscedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroscedasticity', *Econometrica*, May 1980, **48:** 817–38.

WHITE, H., 'Maximum Likelihood Estimation of Misspecified Models', *Econometrica*, January 1982, **50:** 1–25.