DISCUSSION PAPER NO. 401

The Term Structure of the Forward Premium

November 15, 1979

by

Craig S. Hakko

Northwestern University

and

National Bureau of Economic Research
I. Introduction

There have been numerous studies of the efficiency of the foreign exchange (see Levich 1979 for a survey of many of these studies). However, most of these studies focus on a single maturity - usually a one month forward exchange rate. We observe that forward contracts of many maturities are simultaneously traded in the foreign exchange market; yet there are surprisingly few studies that examine the implications of several forward contract maturities (see Porter 1971, Giddy 1977 and Brilhembourg 1978). The hypothesis of market efficiency has well known implications for the relation between a forward exchange rate of a given maturity and the subsequently observed spot rate. In addition, the hypothesis of efficiency has implications for the joint behavior of forward exchange rates of various maturities. This paper will theoretically and empirically examine these additional implications.

Section II will propose an equilibrium theory of the term structure of the forward premium. By combining the (certainty equivalence) theory of the term structure of (domestic and foreign) interest rates with the hypothesis of interest rate parity, a simple expression relating the six month forward premium to the expected future one month forward premium can be derived. It will be shown that the six month forward premium can be written as a geometric average of expected future one month forward premiums. In Section III it is shown that a convenient and efficient method to extract the expected one month forward premium can be obtained by assuming that a general (bivariate) stochastic process generates the one and six month forward premiums. The theory developed in Section II will then impose highly nonlinear cross equation restrictions on the parameters of the stochastic process.
The restrictions imposed on the parameters of the model by the economic theory are highly nonlinear. Sections IV and V discuss two methods of testing the validity of the restrictions. Section IV provides a statistical test of the hypothesis that requires only the unrestricted estimates. The rejection region, under the null hypothesis, is derived. The statistical test proposed in Section V requires the restricted parameter estimates. Maximum likelihood methods for estimating the constrained models are discussed and implemented.

II. The Economics of the Term Structure of the Forward Premium

To develop a theory of the term structure of the forward premium, we begin by assuming that interest rate parity holds. (See Porter 1971 for a similar development). There is much empirical evidence in support of this condition (Frenkel and Levich 1975, 1977). Interest rate parity states that the expected rate of depreciation on foreign exchange equals the interest rate differential. We can write this as:

$$\frac{1 + r_{n,t}}{1 + r_{n-1,t}} = \frac{k_{t}S_{t+n}}{S_{t}}$$

(1)

$$1 + r_{n,t} = \frac{(1 + r_{n,t})^n}{(1 + r_{n-1,t})^{n-1}}$$

(2a)

$$(1 + r_{n,t})^n = (1 + r_{1,t})(1 + r_{2,t}) \ldots (1 + r_{n,t})$$

(2b)
where \( i_{n,t} \) = n period rate of interest at period t
\( i_{m,t} \) = implicit one period forward interest rate for period \( t+n \).

Dividing by the foreign country version of (2a), (2a'), yields:

\[
\frac{1 + i_n}{1 + i_n^*} = \frac{1 + i_{n-1}}{1 + i_{n-1}^*} \cdot \frac{1 + i_{n-1}}{1 + i_{n-2}^*} \cdot \frac{1 + i_{n-2}}{1 + i_{n-2}^*} \cdot \ldots \cdot \frac{1 + i_1}{1 + i_1^*} \cdot \frac{1 + i_0}{1 + i_0^*} = \frac{1 + i_{n-1}}{1 + i_{n-1}^*} \cdot \frac{1 + i_{n-2}}{1 + i_{n-2}^*} \cdot \ldots \cdot \frac{1 + i_1}{1 + i_1^*} \cdot \frac{1 + i_0}{1 + i_0^*}
\]

Substituting interest rate parity (1) into (3) and cancelling yields:

\[
\frac{E_{t} S_{t+n}}{E_{t} S_{t+n-1}} = \frac{1 + i_n^*}{1 + i_n} = 1 + E_{t} \tilde{r}_{1,t+n-1}
\]

\( E_{t} \tilde{r}_{1,t+n-1} \) is the implied expected change in the spot exchange rate in period \( t+n \). Dividing (2b) by (2b'), equating to (3) and substituting from (4) yields:

\[
\frac{E_{t} S_{t+n}}{S_{t}} = \frac{(1 + i_n^*)(1 + i_n) \ldots (1 + i_1^*)}{(1 + i_n^*)(1 + i_n) \ldots (1 + i_1)} = (1 + E_{t} \tilde{r}_{1,t+n-1}) \ldots (1 + E_{t} \tilde{r}_{1,t+n-1}).
\]

Define

\[
r_{k,t} = \left( \frac{F_{k,t}}{S_{t}} \right)^{1/k} - 1.
\]

where \( F_{k,t} \) is the \( k \)-month forward exchange rate prevailing at time \( t \), and \( r_{k,t} \) is the \( k \)-month forward premium. If we assume that the forward rate is an unbiased predictor of the future spot rate, \( F_{k,t} = E_t S_{t+k} \), then we can conclude
that the $k$-month forward premium, $r_{k,t}$, equals the expected $k$-month rate of
depreciation, $E_t r_{t+k}$. Therefore, substituting the definition of $r_{k,t}$ from
equation (6), for $k = 1$ and $n$, into equation (5) we obtain

$$(1 + r_{n,t})^{1/n} = (1 + r_{1,t}) \cdots (1 + E_t r_{1,t+n-1})$$

or, as an approximation (using $\ln (1 + x) = x$, for small $x$):

$$r_{n,t} = \frac{1}{n} \left[ r_{1,t} + E_t r_{1,t+1} + \cdots + E_t r_{1,t+n-1} \right]$$

(7)

$E_t$ is the mathematical expectations operator conditional on the set of in-
formation available to economic agents at time $t$, $\Omega_t$. We will assume that

$\Omega_t \supseteq \Omega_{t-1} \supseteq \cdots$, and that $\Omega_t$ contains at least all current and lagged
values of $(r_{1,t}, r_{n,t})$. The following derivation follows Sargent's (1979)
analysis of the term structure of interest rates.

To add empirical content to equation (7), we must specify how
expectations are formed and what variables belong in $\Omega_t$. First, notice
that the conditional expectations operator in (7), $E_{t|\omega}$, has $\Omega_t$ as the
conditioning set, where $\Omega_t$ includes all relevant information for calculating
expectations of future one-month rates of depreciation. For convenience in
deriving testable implications based on (7), let us write $E_{t|\omega}$ as

$E_t r_{1,t|\omega}$, and so rewrite (7) as

$$r_{n,t} = \frac{1}{n} \left[ r_{1,t} + E_t r_{1,t+1|\omega} + \cdots + E_t r_{1,t+n-1|\omega} \right]$$

(7')

Let $\theta_t$ be any subset of $\Omega_t$, such that $\theta_t$ includes at least current and
lagged values of $r_{1,t}$ and $r_{n,t}$. Now, take expectations of both sides of
(7'), conditional on the smaller information set $\theta_t$, to obtain

$$r_{n,t} = \frac{1}{n} \left[ r_{1,t} + E_{t|\theta_t} r_{1,t+1|\theta_t} + \cdots + E_{t|\theta_t} r_{1,t+n-1|\theta_t} \right]$$

(7'')
where we used the law of iterated projections that states that $E(y|x,z) = E(E(y|x,z)|z)$, where $x$, $y$, $z$ are normal random variables. Notice that $(7'')$ and $(7')$ are of the same form. In particular, if we leave important variables out of $\theta_t$ (that were in $\Theta$) we will not invalidate the tests reported. Now, take first differences of $(7'')$:

$$ r_{n,t} - r_{n,t-1} = \frac{1}{n} \{ (r_{1,t} - r_{1,t-1}) + [E(r_{1,t+1}|\theta_t) - E(r_{1,t}|\theta_{t-1})] + \ldots + [E(r_{t,n+1}|\theta_t) - E(r_{t,n+2-1}|\theta_{t-1})] \} \tag{8} $$

Write $\Delta r_{n,t} = r_{n,t} - r_{n,t-1}$ and $\Delta r_{1,t} = r_{1,t} - r_{1,t-1}$. Take expectations of both sides of (8), conditional on $\theta_{t-1}$, using the law of iterated projections, to get

$$ E(\Delta r_{n,t} | \theta_{t-1}) = \frac{1}{n} (E(\Delta r_{1,t} | \theta_{t-1}) + E(\Delta r_{1,t+1} | \theta_{t-1}) + \ldots + E(\Delta r_{t,n+1} | \theta_{t-1})). \tag{8'} $$

We must now specify exactly which variables to include in $\theta_t$. We shall restrict $\theta_t$ to include only current and lagged values of $\Delta r_{1,t}$, $\Delta r_{n,t}$, that is $\theta_t = (\Delta r_{1,t}, \Delta r_{1,t-1}, \ldots, \Delta r_{n,t}, \Delta r_{n,t-1}, \ldots)$. Given this information set, we can easily calculate the conditional expectations in (8'). We shall report two methods of calculating these expectations, and the restrictions implied by (8').

III. The Empirical Implications of the Term Structure of the Forward Premium

Assuming that $(\Delta r_{1,t}, \Delta r_{n,t})$ is a linearly indeterministic co-
variance stationary stochastic process, we can use the Wold Decomposition

Theorem to write (letting $R_{1,t} = \Delta r_{1,t}$ and $R_{n,t} = \Delta r_{n,t}$):

$$R_{1,t} = \alpha(L) w_t + \beta(L) v_t$$

$$R_{n,t} = \gamma(L) w_t + \delta(L) v_t$$

where $\alpha(L)$, $\beta(L)$, $\gamma(L)$ and $\delta(L)$ are one-sided polynomials in the lag operator $L$.

$$w_t = R_{1,t} - E(R_{1,t} | e_{t-1})$$
$$v_t = R_{n,t} - E(R_{n,t} | e_{t-1})$$

$$E w_{t-k} = \begin{cases} 0 & k = 0 \\ \sigma^2 & k 
eq 0 \end{cases}$$

$$E v_{t-k} = \begin{cases} 0 & k = 0 \\ \sigma_v^2 & k 
eq 0 \end{cases}$$

$$E w_{t-k} v_{t-k} = 0$$

$a(0) = \delta(0) = 1$ and $\beta(0) = \gamma(0) = 0$.

The Weiner-Kolmogorov prediction formulas allow us to write the conditional expectations in (8') in a simple fashion:

$$E_{t-1} R_{t+k} = A(L) w_{t-1} + B(L) v_{t-1}$$

where $[ ]_+$ means "ignore negative powers of $L$." Substituting expression (10) into (8') and rearranging, yields:

$$E_{t-1} R_{t} = \left[ \frac{a(L)}{L} + \frac{a(L)}{L^2} + \ldots + \frac{a(L)}{L^n} \right] w_{t-1}$$
\[
\begin{align*}
\frac{1}{n} \left[ \frac{b(L)}{L} + \frac{b(L)}{L^2} + \ldots + \frac{b(L)}{L^n} \right] v_{t-1} \\
= \frac{1}{n} \left[ \frac{a(L)}{L} \left( \frac{1-L^{-1}}{1-L} \right) \right] w_{t-1} + \frac{1}{n} \left[ \frac{b(L)}{L} \left( \frac{1-L^{-n}}{1-L} \right) \right] v_{t-1}.
\end{align*}
\]

But, we can also use the Wiener-Kolmogorov prediction formula to write the left hand side of (11) as
\[
E_{t-1} R_{t+n | t} = \left[ \frac{a(L)}{L} \right] w_{t-1} + \left[ \frac{b(L)}{L} \right] v_{t-1}.
\]  

Equating terms in (11) and (12) yield a set of cross equation restrictions on the parameters of the bivariate moving average representation of \((R_{1,t}, R_{n,t})\) in (9) implied by the theory of the term structure of the forward premium:
\[
\begin{align*}
\left[ \frac{a(L)}{L} \right] &= \frac{1}{n} \left[ \frac{a(L)}{L} \left( \frac{1-L^{-n}}{1-L} \right) \right] \\
\left[ \frac{b(L)}{L} \right] &= \frac{1}{n} \left[ \frac{b(L)}{L} \left( \frac{1-L^{-n}}{1-L} \right) \right].
\end{align*}
\]

It is possible to estimate equation (9) subject to (13) and so test the validity of the restrictions embodied in (13). However, it is very difficult to estimate constrained bivariate moving averages and so we use an alternative representation of \((R_{1,t}, R_{n,t})\).

Equation (8') is a restriction across the systematic part of \((R_{1,t}, R_{n,t})\), imposed by rational expectations. By our assumptions of stationarity we can write \((R_{1,t}, R_{n,t})\) as a vector autoregression (where the \(a_i's, b_i's, \gamma_i's\) and \(\delta_i's\) in (14) are different than those in (9)), the \(w_t\) and \(v_t\) are the same).
\[ R_{i,t} = \sum_{i=1}^{M} \alpha_i R_{i,t-i} + \sum_{i=1}^{M} \beta_i n_{i,t-i} + \nu_t \] (14a)

\[ R_{n,t} = \sum_{i=1}^{M} \gamma_i R_{i,t-i} + \sum_{i=1}^{M} \beta_i n_{i,t-i} + \nu_t \] (14b)

where

\[ E_{\text{v}} R_{i,t-i} = E_{\text{v}} R_{i,t-i} = E_{\text{v}} R_{n,t-i} = 0, \text{ for } i = 1, 2, \ldots, M \]

\[ E_{\text{v}} \nu_t = \begin{cases} 0 & i \neq 0 \\ \nu_t & i = 0 \end{cases} \]

\{\nu_t, \nu_t\} is the innovation in the \((R_{1,t}, R_{n,t})\) process; the errors are contemporaneously correlated, but uncorrelated at all lags. Equation (14) can be rewritten as:

\[ x_t = \delta x_{t-1} + \alpha_t \] (15)

where

\[
\begin{bmatrix}
R_{1,t} \\
R_{1,t-1} \\
\vdots \\
R_{1,t-M+1} \\
R_{n,t} \\
R_{n,t-1} \\
\vdots \\
R_{n,t-M+1}
\end{bmatrix} = \begin{bmatrix}
\nu_t \\
0 \\
\vdots \\
0 \\
\nu_t \\
0 \\
\vdots \\
0
\end{bmatrix}
\]

\[ \alpha_t = \begin{bmatrix}
\alpha_t \\
\vdots \\
\alpha_t \\
\end{bmatrix} \]

Equation (15) amounts to rewriting an \(N\)th order difference equation as a vector first order system.
Repeated substitution from (15) yields:

\[ x_{t+1} = Ax_t + a_{t+1} = A^{2}x_{t-1} + Aa_t + a_{t+1} \]  
\[ x_{t+j} = A^{j+1}x_{t-1} + A^{j}a_t + \ldots + a_{t+j}. \]  

Since \( E_{t-1}a_{t+k} = 0 \) \((k = 0, 1, \ldots)\) we can write (16) as

\[ E_{t-1}x_{t+j} = A^{j+1}x_{t-1} \quad j = 1, 2, \ldots \]  

Letting \( c = (1 \ 0 \ldots \ 0) \) and \( d = (0 \ldots 0 \ 1 \ 0 \ldots \ 0) \), we can write

\[ R_{1,t} = c \ x_t \]  
\[ R_{n,t} = d \ x_t. \]  

Multiply (15) by \( c \) and \( d \) to get

\[ c \ x_t = c \ Ax_{t-1} + c \ a_t \]  
\[ d \ x_t = d \ Ax_{t-1} + d \ a_t. \]
Equating to (18) we get

\[ R_{t,t} = c A X_{t-1} + \varepsilon_{t,1} \quad (29) \]

\[ k_{n,t} = d A X_{t-1} + a_{n,t} \]

Multiply restriction (17) by \( c \) to get

\[ \tilde{R}_{t-1,1}^{\text{c}} X_{t+1} = c A^{t+1} X_{t-1} \quad j = 1, 2, \ldots \quad (20) \]

Updating (18) by \( +j \) and substituting into (20) yields

\[ \tilde{R}_{t-1,1}^{\text{c}} X_{t+1} = c A^{t+1} X_{t-1} \quad j = 1, 2, \ldots \quad (21) \]

Substitution of (21) into (6) yields

\[ E_{t-1} R_{n,t} = (1/n)(c A X_{t-1} + cA^2 X_{t-1} + \ldots + cA^n X_{t-1}) \]

\[ = \frac{1}{n} c (A + A^2 + \ldots + A^n) X_{t-1}. \quad (22) \]

Taking expectations conditional on \( I_{t-1} \) in (19) yields

\[ E_{t-1} R_{n,t} = d A X_{t-1} \quad (23) \]

Equating equations (22) and (23) yields the following restriction imposed by rational expectations:

\[ d A = (1/n) c (A + A^2 + \ldots + A^n). \quad (24) \]

The intuition behind these restrictions arise from the following observations. We assumed that the \((\tilde{R}_{j,t}, R_{n,t})\) process was generated by a vector autoregression. That is, we regress both \( \tilde{R}_{j,t} \) and \( R_{n,t} \) against lagged values of \((\tilde{R}_{j,t}, R_{n,t})\). Dahl has shown the conditions under which
this is valid (see Whittle 1963). If the economic agents realized this they will use the parameters of the autoregression to generate their forecasts. For the data to be consistent with the model the parameter values must be restricted. These restrictions are summarized in equation (8) and equivalently in equation (24).

IV. Econometric Tests and Results

The restrictions implied by equation (24) are highly nonlinear. There are two basic methods to test the validity of the restrictions implied by the theory. The first method, discussed in detail in this section, was originally proposed by Wald. This method requires obtaining the unrestricted maximum likelihood estimates \( \hat{\psi} \) of the parameter vector \( \psi = (\alpha, \beta, \gamma, \delta) \). Let us write the restrictions implied by (24) in the form \( h(\psi) = 0 \). Wald's method then tests \( h(\hat{\psi}) = 0 \). The second method, discussed in detail in the next section, is based on the likelihood ratio test. This method requires obtaining in addition to the unrestricted estimate \( \hat{\psi} \), the restricted estimate \( \hat{\phi} \). One then compares the likelihood of \( \hat{\psi} \) to \( \hat{\phi} \). A difficulty with this method is obtaining the restricted maximum likelihood estimates. The next section will present two methods of obtaining \( \hat{\phi} \).

Under the assumption that \( \{\varepsilon_t, \nu_t\} \) is bivariate normal, the likelihood function for a sample of \( \{\varepsilon_t, \nu_t\}, t = 1, 2, \ldots, T \) is given by

\[
L(\alpha, \beta, \gamma, \delta) = (2\pi)^{-T/2} |V|^{-T/2} \exp \left( -\frac{1}{2} \sum_{t=1}^{T} \epsilon_t' \psi^{-1} \epsilon_t \right)
\]

where

\[
\epsilon_t = \begin{bmatrix} \nu_t \\ \varepsilon_t \end{bmatrix}, \quad V = \Sigma + \sigma^2 \begin{bmatrix} \nu_t \\ \varepsilon_t \end{bmatrix} \begin{bmatrix} \nu_t \\ \varepsilon_t \end{bmatrix}'
\]

(25)
Maximizing (25) without any restrictions, that is, with all parameters free, is equivalent to estimating (14) by least squares. Wilson (1973) shows that the parameter estimates with an unknown \( \psi \) may be obtained by

\[
\min_{\hat{\psi}} \left| \left| \hat{\psi} - (1/T) \sum_{t=1}^{T} \hat{\epsilon}_t \right| \right|
\]

To test restriction (24) we proceed as follows. Let

\[
\psi = (a', \beta', \gamma', \delta')
\]

\[
\hat{\psi} = \text{OLS (unrestricted) estimate.}
\]

Write restriction (24) as

\[
h(\psi) = \hat{A}(\psi) = (1/n)\hat{c}(\hat{\psi}) + \hat{A}^2(\psi) + \ldots + \hat{A}^n(\psi)
\]

\[
= (0 \ 0 \ \ldots \ 0) = 0
\]

where we write \( A = A(\psi) \) to indicate the dependence of \( A \) on \( \psi \). The test amounts to testing whether the vector \( h(\hat{\psi}) \) is significantly different from the zero vector. The problem is to determine the shape of the rejection regions. The problem is intuitively solved as follows (Silvey 1972, pp. 115-116, or Rao 1973, pp. 418-415):

We expect \( \hat{\psi} \) to be "close to" \( \bar{\psi} \), under the null hypothesis, and we know that \( \hat{A}(\hat{\psi} - \bar{\psi}) \) is approximately \( N(0, B^{-1}) \), where \( B^{-1} \) is the information matrix for the coefficient vector \( \hat{\psi} \), in a single observation.

Expanding \( h(\psi) \) about \( \hat{\psi} \) in a Taylor series, to linear terms, we get

\[
h(\hat{\psi}) \approx h(\hat{\psi}) + H_{\hat{\psi}}(\hat{\psi} - \bar{\psi})
\]
where

$$H_\psi = \left( \begin{array}{c} 2 \lambda \left( \frac{\psi}{\psi} \right) \\ \frac{1}{\psi} \end{array} \right).$$

Since $h(\psi) = 0$, under the null hypothesis, we may rewrite (27) as

$$h(\hat{\psi}) = H_\psi' \left( \begin{array}{c} \hat{\psi} - \psi \end{array} \right).$$

Therefore, $\sqrt{n} h(\hat{\psi})$ is approximately $N(0, H_\psi' B^{-1}_\psi H_\psi)$. Letting $x$ be the vector of observations, the rejection region becomes

$$\{ x \mid n[h(\psi(x))]\left( H_\psi' B^{-1}_\psi H_\psi \right)^{-1} h(\psi(x)) > k \}. \quad (29)$$

To actually apply this test one needs an estimate of $H_\psi$ and $B^{-1}_\psi$. For $(1/n)B^{-1}_\psi$ we can use the estimated variance-covariance matrix, obtained from estimating (14). For $H_\psi$, we numerically differentiate the $1 \times 8$ restriction vector $h(\psi)$ (at the OLS estimates) with respect to all sixteen parameters. Calling these estimates $(1/n)B^{-1}_\psi$ and $H_\psi$, the value $W$ is given by

$$W = h(\psi)' [H_\psi (1/n) B^{-1}_\psi H_\psi'] h(\psi). \quad (30)$$

Under the null hypothesis, $h(\hat{\psi}) = 0$, $W$ is approximately distributed chi-square with eight degrees of freedom. Large values of $W$ indicate rejection of the hypothesis.

The OLS (unrestricted) estimates are given in Tables 1 and 2, under the heading "Unrestricted Estimates." Also presented is an estimate of $V$ and $|V|$. At the bottom of Tables 1 and 2 the $W$-statistic is presented, along with its marginal significance level. The marginal significance level is the probability of observing a number greater than the statistic, given that the null hypothesis is true.
The results presented in Table 1 for Germany indicate a failure
to reject the validity of the hypothesis that the pure expectations theory
of the term structure of the forward premium is correct. The W-statistic
of 10.43 is insignificant, as indicated by a marginal significance level of
23.7 percent. The results in Table 2 for Canada indicate rejection of the
null hypothesis. The W-statistic of 21.84 is significant, as indicated by
a marginal significance level of 0.5 percent. The assumption that \((R_{it}, t')^{H_{0}, t}\) is
stationary is equivalent to the assumption that the characteristic roots of the
matrix A are all less than one in modulus (see Sargent [1979a], p. 273). The roots
of A, using OLS estimates, were calculated and all were found to be less than one
in modulus.

V. Econometric Tests and Results, II

In the last section, we presented tests of the validity of restric-
tion (24) based on the unrestricted estimates. In the case of Canada,
we cannot determine the source of rejection. In this section, we shall
estimate the model with the restrictions imposed, and then compare the re-
stricted and unrestricted models using a likelihood ratio test.

The restrictions implied by (24) are highly nonlinear. Sargent
(1979b) proposes two alternative estimation strategies. The first method
requires estimating the first row of A, equation \((14a)\), by least squares.
Then, the \((M+1)st\) row of A, equation \((14b)\), is calculated using an itera-
tive procedure. Form a preliminary estimate of A, call it \(A_{o}\), by setting
row \(M+1\) to a row of zeroes, and all other rows to their known (or
consistent) values. Calculate the \((M+1)st\) row of A, at iteration \(i+1\), as
<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>a_j</td>
<td>-0.3633</td>
<td>0.0726</td>
<td>-0.0856</td>
<td>-0.2098</td>
</tr>
<tr>
<td>b_j</td>
<td>0.5599</td>
<td>-0.5044</td>
<td>0.1862</td>
<td>0.4478</td>
</tr>
<tr>
<td>y_j</td>
<td>-0.0767</td>
<td>0.0642</td>
<td>0.0292</td>
<td>-0.0412</td>
</tr>
<tr>
<td>δ_j</td>
<td>0.2227</td>
<td>-0.1829</td>
<td>0.0304</td>
<td>0.1210</td>
</tr>
</tbody>
</table>

\[
V = \begin{pmatrix}
7.726 \times 10^{-7} & 2.546 \times 10^{-7} \\
1.781 \times 10^{-7} & 1.361 \times 10^{-7}
\end{pmatrix}, \quad \|V\| = 3.442 \times 10^{-14}
\]

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>a_j</td>
<td>-0.3651</td>
<td>-0.1119</td>
<td>-0.1779</td>
<td>-0.1117</td>
</tr>
<tr>
<td>b_j</td>
<td>0.3003</td>
<td>-0.0845</td>
<td>0.2441</td>
<td>0.2275</td>
</tr>
<tr>
<td>y_j</td>
<td>-0.0797</td>
<td>-0.0410</td>
<td>-0.0264</td>
<td>-0.0098</td>
</tr>
<tr>
<td>δ_j</td>
<td>0.0784</td>
<td>0.0447</td>
<td>0.0435</td>
<td>0.0199</td>
</tr>
</tbody>
</table>

\[
V = \begin{pmatrix}
7.959 \times 10^{-7} & 2.668 \times 10^{-7} \\
1.352 \times 10^{-7} & 1.352 \times 10^{-7}
\end{pmatrix}, \quad \|V\| = 3.652 \times 10^{-14}
\]

Likelihood ratio statistic = 11.134
Marginal significance level = 0.194
W = 10.42
Marginal significance level = 0.237
<table>
<thead>
<tr>
<th>j</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Unrestricted Estimates</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>-0.2830</td>
<td>-0.0174</td>
<td>-0.0168</td>
<td>-0.1421</td>
</tr>
<tr>
<td>β</td>
<td>0.3743</td>
<td>0.1092</td>
<td>0.1202</td>
<td>0.0028</td>
</tr>
<tr>
<td>γ</td>
<td>0.1745</td>
<td>0.1186</td>
<td>0.0643</td>
<td>0.0171</td>
</tr>
<tr>
<td>δ</td>
<td>-0.1140</td>
<td>-0.0249</td>
<td>-0.0161</td>
<td>0.0001</td>
</tr>
<tr>
<td>ψ</td>
<td>$9.423 \times 10^{-8}$</td>
<td>$4.572 \times 10^{-8}$</td>
<td>$4.084 \times 10^{-8}$</td>
<td>$\sqrt{\psi} \approx 1.757 \times 10^{-15}$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>j</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Restricted Estimates</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>-0.5183</td>
<td>-0.0885</td>
<td>-0.0077</td>
<td>-0.1151</td>
</tr>
<tr>
<td>β</td>
<td>0.5196</td>
<td>0.0389</td>
<td>0.0769</td>
<td>-0.0012</td>
</tr>
<tr>
<td>γ</td>
<td>-0.0673</td>
<td>-0.0227</td>
<td>-0.0135</td>
<td>-0.0098</td>
</tr>
<tr>
<td>δ</td>
<td>0.0580</td>
<td>0.0118</td>
<td>0.0064</td>
<td>-0.0001</td>
</tr>
<tr>
<td>ψ</td>
<td>$9.829 \times 10^{-8}$</td>
<td>$4.958 \times 10^{-8}$</td>
<td>$4.498 \times 10^{-8}$</td>
<td>$\sqrt{\psi} \approx 1.963 \times 10^{-15}$</td>
</tr>
</tbody>
</table>

Likelihood ratio statistic = 20.843
Marginal significance level = 0.005

W = 21.84
Marginal significance level = 0.005
(row Mi)_{i+1} = d_{i+1} = c \left( \frac{1}{n} [A_1 + A_2^2 + \ldots + A_{M-1}^M] \right) \quad (31)

where $A_1$ is the estimate of $A$ on the $i$th iteration. At each step in forming $A_{i+1}$, all rows (except the (Mi)st) are kept equal to the corresponding row of $A_0$. If this procedure converges, it will find a $A$ that satisfies (24). This procedure will converge if the characteristic roots of $A$ are less than one in modulus. Since the elements of row 1 are consistently estimated by least squares, the (Mi)st row will be consistently estimated as a function of the first row of $A$.

Define the solution to the iteration on (31) as the (set) function

$$\gamma, \delta = \phi(a, \beta) \quad (32)$$

$\phi$ maps the $a$'s and $\beta$'s into a set of $\gamma$'s and $\delta$'s that satisfy restriction (24). Hence, one (consistent) estimator of $\gamma, \delta$ is $\phi(a, \beta)$.

Under the restriction (24) the likelihood function in (25),

$$L(a, \beta, \gamma, \delta | (x_0, (i_0)))$$

becomes a function only of the $a$'s and $\beta$'s. As Wilson (1973) argues, maximum likelihood estimates with an unknown $V$ are obtained by minimizing $|\hat{\gamma}|$, with respect to the $a$'s and $\beta$'s, where

$$|\hat{\gamma}| = \frac{1}{T} \sum_{t=1}^{T} e_t(a, \beta) e_t(a, \beta)'$$

where the $e_t(a, \beta)$, the residuals from (14), are functions of the $a$'s and $\beta$'s only, since they were calculated from (14) with (24) imposed.

A derivative free nonlinear minimization routine can be used to estimate (14) under the restriction (24). The IMSL subroutine ZMIN, which uses a quasi-Newton method, was used. Generally, 600 iterations were required to obtain three significant digits. The least
squares estimates of $\alpha$ and $\delta$ were used as starting values.

Tables 1 and 2 report, in addition to the unrestricted estimates, the restricted maximum likelihood estimates. In addition, the likelihood ratio statistic, which is distributed $\chi^2(8)$, is reported along with the marginal significance level.

The results for Germany indicate a failure to reject the null hypothesis. The likelihood ratio statistic of 11.134 is insignificant, as indicated by the marginal significance level of 19.4 percent. The results for Canada, a likelihood ratio statistic of 20.843 (marginal significance level of 0.8 percent), indicate a rejection of the null hypothesis.

Not surprisingly, the results implied by the $\psi$-statistic and likelihood ratio statistic are very close (marginal significance levels of 23.7 percent and 19.4 percent for Germany, and 0.5 percent and 0.8 percent for Canada). This is as it should be, since both tests are asymptotically equivalent (Silvey 1970, p. 118), and the sample size is 188 observations.

Can we use the restricted estimates to locate the reason the null hypothesis is rejected for Canada? Table 3 presents an estimate of the variance-covariance matrix of the restricted parameter estimates ($\alpha$ and $\delta$) for Canada. Standard errors for $\gamma$ and $\delta$ are not reported since $\gamma$ and $\delta$ cannot be analytically solved for in terms of $\alpha$ and $\delta$. The difference between the restricted and unrestricted estimates of $\alpha$ and $\delta$ are insignificant. Using the standard errors of the unrestricted estimates of $\gamma$ and $\delta$ (not reported), the unrestricted estimate of $\gamma_{1}$ is significantly different from the restricted estimate (all other $\gamma$'s and $\delta$'s are insignificantly different). In calculating the Wald statistic we required an estimate of $h(\psi)$—the restriction vector implied by (24). An estimate of that vector, evaluated at the OLS estimate is
TABLE 7
VARIANCE-COVARIANCE MATRIX OF CANADIAN ESTIMATES OF (a, $\beta$)

<table>
<thead>
<tr>
<th></th>
<th>$a_1$</th>
<th>$a_2$</th>
<th>$a_3$</th>
<th>$a_4$</th>
<th>$\beta_1$</th>
<th>$\beta_2$</th>
<th>$\beta_3$</th>
<th>$\beta_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_1$</td>
<td>0.171</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$a_2$</td>
<td>-0.133</td>
<td>0.605</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$a_3$</td>
<td>0.007</td>
<td>0.331</td>
<td>0.386</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$a_4$</td>
<td>0.243</td>
<td>-0.247</td>
<td>0.004</td>
<td>0.523</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>0.035</td>
<td>0.036</td>
<td>-0.005</td>
<td>0.027</td>
<td>0.030</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>-0.228</td>
<td>-0.179</td>
<td>-0.107</td>
<td>-0.271</td>
<td>-0.133</td>
<td>0.683</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_3$</td>
<td>-0.080</td>
<td>0.066</td>
<td>-0.333</td>
<td>-0.116</td>
<td>0.083</td>
<td>-0.072</td>
<td>0.755</td>
<td></td>
</tr>
<tr>
<td>$\beta_4$</td>
<td>0.074</td>
<td>-0.115</td>
<td>-0.070</td>
<td>-0.257</td>
<td>0.028</td>
<td>-0.159</td>
<td>-0.115</td>
<td>0.848</td>
</tr>
</tbody>
</table>

$h = (0.211 \ 0.129 \ 0.079 \ 0.035 \ -0.185 \ -0.052 \ -0.032 \ -0.00)$

$|h - 0| = 0.107$.

It is interesting to note that the first restriction, $h_1(\delta) = 0.211$, is also the largest in absolute value. Again the results are consistent between the Wald test and likelihood ratio test.

The period being examined, April 1973 - May 1975, was a time of adjustment to a new floating rate system. It is possible that the "noisiness" of the system changed over time (Hakkio [1979] presents evidence that the variance had increased for Canada). To allow for this possibility, we split the Canadian
sample in two, and reestimated the restricted and unrestricted version of the model. The results are presented in Tables 4 and 5.

For the first period we obtain a likelihood ratio statistic 21.465, with a marginal significance level of 0.6 percent. So, we again reject the null hypothesis for Canada for the first period. The likelihood ratio statistic for the whole period and the first period are quite close. For the second period we obtain a likelihood ratio statistic of only 13.502, with a marginal significance level of 10.2 percent. Therefore, for the second period we fail to reject the null hypothesis, and therefore conclude that the data are consistent with the pure expectations theory of the term structure of the forward premium—for the second period.

As before, the roots of the A matrix indicate that the process was stationary. As shown in Hakkio (1979), the variance of the forecast error changed over time in the Canadian case. Consequently, the data may not have been stationary and so the econometric methodology may have been suspect. Splitting the sample in two presumably reduced the problem posed by heteroskedasticity.

VI. Conclusions

Most studies that test foreign exchange market efficiency focus on the relation between the spot exchange rate and a single maturity for the forward exchange rate, usually the one month rate. This procedure ignores the fact that more than one maturity currently is being traded. This paper extends the analysis of market efficiency so as to obtain implications concerning the joint movements of the spot exchange rate and the one and six month forward exchange rate. Using the certainty equivalence theory of the term structure of (domestic and foreign) interest rates and the
### Table 4

**Canada: Estimates of Bivariate Autoregression Unrestricted and Restricted, First Period**

<table>
<thead>
<tr>
<th>j</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_j$</td>
<td>-0.5758</td>
<td>-0.2897</td>
<td>-0.3760</td>
<td>-0.4209</td>
</tr>
<tr>
<td>$b_j$</td>
<td>1.0329</td>
<td>0.4575</td>
<td>0.7801</td>
<td>0.0494</td>
</tr>
<tr>
<td>$\gamma_j$</td>
<td>0.0177</td>
<td>-0.0767</td>
<td>-0.1543</td>
<td>-0.1640</td>
</tr>
<tr>
<td>$\delta_j$</td>
<td>0.2066</td>
<td>0.2860</td>
<td>0.3201</td>
<td>0.0760</td>
</tr>
</tbody>
</table>

\[
\hat{V} = \begin{bmatrix}
11.125 \times 10^{-8} & 4.684 \times 10^{-8} \\
2.897 \times 10^{-8}
\end{bmatrix}
\| \hat{V} \| = 1.030 \times 10^{-15}

**Restricted Estimates**

<table>
<thead>
<tr>
<th>j</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_j$</td>
<td>-0.7581</td>
<td>-0.2875</td>
<td>-0.1967</td>
<td>-0.2003</td>
</tr>
<tr>
<td>$b_j$</td>
<td>0.9032</td>
<td>0.1384</td>
<td>0.3603</td>
<td>-0.0012</td>
</tr>
<tr>
<td>$\gamma_j$</td>
<td>-0.0895</td>
<td>-0.0462</td>
<td>-0.0227</td>
<td>-0.0098</td>
</tr>
<tr>
<td>$\delta_j$</td>
<td>0.0880</td>
<td>0.0302</td>
<td>0.0176</td>
<td>-0.0001</td>
</tr>
</tbody>
</table>

\[
\hat{V} = \begin{bmatrix}
12.691 \times 10^{-8} & 5.748 \times 10^{-8} \\
3.631 \times 10^{-8}
\end{bmatrix}
\| \hat{V} \| = 1.304 \times 10^{-15}

Likelihood ratio statistic = 21.465
Marginal significance level = 0.006
<table>
<thead>
<tr>
<th>j</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>a_j</td>
<td>-0.0554</td>
<td>0.0691</td>
<td>0.1566</td>
<td>0.2611</td>
</tr>
<tr>
<td>b_j</td>
<td>0.0727</td>
<td>-0.0466</td>
<td>-0.1590</td>
<td>-0.1747</td>
</tr>
<tr>
<td>c_j</td>
<td>0.3164</td>
<td>0.2144</td>
<td>0.1989</td>
<td>0.3244</td>
</tr>
<tr>
<td>d_j</td>
<td>-0.1859</td>
<td>-0.1703</td>
<td>-0.1966</td>
<td>-0.1570</td>
</tr>
</tbody>
</table>

\[ \hat{V} = \begin{bmatrix} 5.332 \times 10^{-8} & 3.503 \times 10^{-8} \\ 4.736 \times 10^{-8} & 0.0 \end{bmatrix} \quad |\hat{V}| = 1.582 \times 10^{-15} \]

### Restricted Estimates

<table>
<thead>
<tr>
<th>j</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>a_j</td>
<td>-0.1480</td>
<td>0.1212</td>
<td>0.2087</td>
<td>0.2047</td>
</tr>
<tr>
<td>b_j</td>
<td>0.1487</td>
<td>-0.0993</td>
<td>-0.1770</td>
<td>-0.1684</td>
</tr>
<tr>
<td>c_j</td>
<td>0.0518</td>
<td>0.0964</td>
<td>0.0822</td>
<td>0.0437</td>
</tr>
<tr>
<td>d_j</td>
<td>-0.0376</td>
<td>-0.0802</td>
<td>-0.0688</td>
<td>-0.0360</td>
</tr>
</tbody>
</table>

\[ \hat{V} = \begin{bmatrix} 6.013 \times 10^{-8} & 3.649 \times 10^{-8} \\ 5.259 \times 10^{-8} & 0.0 \end{bmatrix} \quad |\hat{V}| = 1.831 \times 10^{-15} \]

Likelihood ratio statistic = 13.302
Marginal significance level = 0.102
hypothesis of interest rate parity, it is possible to write the six month forward premium as a geometric average of the current one month forward premium and expected future one month forward premiums. Testable implications are then derived by assuming that the one and six month forward premiums are generated by a bivariate autoregression.

The hypothesis of rational expectations imposes a set of highly non-linear cross equation restrictions on the parameters of the model. Two different methods were then presented to test the validity of the restrictions. Both methods yielded identical conclusions. It was found that for Germany the data were consistent with the theory of the term structure of the forward premium. For Canada the data were inconsistent with the theory. Hakio (1979) found that the Canadian - U.S. exchange rate system appeared to change over time with respect to the forecast variance. To examine the implications of this, we estimated the model over two subperiods. We then found that although the data were inconsistent with the theory in the first subperiod, the data were consistent with the theory in the second sub-period. Due to the complex nature of the restrictions, the cause of rejection could not be fully ascertained. In addition, although the theory may be rejected for Canada, there is no clear alternative theory that can be accepted.
REFERENCES


