NEW YORK UNIVERSITY
STERN SCHOOL OF BUSINESS
FINANCE DEPARTMENT

Working Paper Series, 1994

The Forward Premium Anomaly: Three Examples in Search of a Solution

David K. Backus, Silverio Foresi, and Chris I. Telmer

FD-94-6

The Forward Premium Anomaly: Three Examples in Search of a Solution*

David K. Backus† Silverio Foresi‡ Chris I. Telmer§

July 1, 1994

Extremely Preliminary
For Discussion Only

Abstract

Perhaps the most puzzling feature of currency prices is the tendency for high interest rate currencies to appreciate, when the expectations hypothesis suggests the reverse. This *forward premium anomaly* has been attributed, by some, to a time-varying risk premium, but theory has yet to produce a risk premium with the requisite properties. We characterize the risk premium in a general theoretical framework and construct three examples that illustrate features a theoretical explanation of the anomaly is likely to have.

JEL Classification Codes: F31, G12, G15.

Keywords: forward and spot exchange rates; risk premium; pricing kernels; bond pricing.

*We thank Martin Evans and especially Stan Zin for helpful comments and suggestions. Backus thanks the National Science Foundation for financial support.

†Stern School of Business, New York University and NBER.

‡Stern School of Business, New York University.

§Graduate School of Industrial Administration, Carnegie Mellon University.

1 Introduction

One of the most puzzling features of international asset prices is the tendency for high interest rate currencies to appreciate, when the expectations hypothesis suggests the opposite: that investors will demand higher interest rates from currencies expected to fall in value. This departure from uncovered interest parity, which we term the *forward premium anomaly*, has been documented in dozens—and possibly hundreds—of studies, and has spawned a second generation of papers attempting to account for it. One of the most influential of these is Fama (1984), who attributed the behavior of forward and spot exchange rates to a time-varying risk premium. Fama showed that the implied risk premium must (i) be negatively correlated with the expected rate of depreciation and (ii) have a greater variance.

We refer to this feature of the data as an anomaly, because asset pricing theory to date has been notably unsuccessful in producing a risk premium with the requisite properties. This includes applications of the capital asset pricing model to currency prices (Frankel and Engel, 1984; Mark, 1988), statistical models relating risk premiums to changing second moments (Cumby, 1988; Domowitz and Hakkio, 1985; Hansen and Hodrick, 1983), and consumption-based asset pricing theories, including departures from time-additive preferences (Backus, Gregory, and Telmer, 1993; Bansal, 1991; Macklem, 1991) and from expected utility (Bekaert, Hodrick, and Marshall, 1992).

We approach the anomaly from the perspective of dynamic asset pricing theory, in which asset prices are governed by a *pricing kernel.* Existence of a pricing kernel is guaranteed in any economy that does not admit pure arbitrage opportunities. We relate the behavior of spot exchange rates to pricing kernels in two currencies and describe, in an abstract theoretical setting, the properties these kernels must have to account for the puzzling behavior of forward and spot exchange rates.

These properties are given concrete form in three examples that we think will lead to a persuasive explanation of the anomaly. The first is a variant of the popular Hamilton model in which conditional moments, including risk premiums, vary across two "regimes." The second is a two-currency version of the Cox-Ingersoll-Ross (1985) model similar to Nielsen and Saá-Requejo (1993) and Saá-Requejo (1994). This model delivers Fama's condition (i) with any permissible values of the parameters, but has difficulty with condition (ii). Our third example combines elements of the first two: expansionary monetary policy and other factors periodically leads to greater uncertainty and an expectation among market participants that the currency will fall in value. This combination of events tends to produce contrary movements in the depreciation rate and the forward premium and thus a solution to the anomaly.

2 The Forward Premium Anomaly

We begin with a review of properties of spot and forward exchange rates for the US dollar versus the remaining G7 currencies. Here and elsewhere, s_t is the logarithm of the dollar price of one unit of foreign currency and f_t is the logarithm of the dollar price of a one-month forward contract: a contract arranged at date t specifying payment of $\exp(f_t)$ dollars at date $t+1$ in return for one unit of foreign currency.

We report summary statistics in Table 1 for the depreciation rate of the dollar, $s_{t+1} - s_t$, and the forward premium, $f_t - s_t$. We see in Panel A of the table that depreciation rates have been small relative to their standard deviations. The largest mean depreciation rate rate is 0.0036 monthly for the yen, or about 4.3 percent per year, but the standard deviation is almost ten times larger. Higher moments, however, exhibit no clear pattern. There is some evidence of skewness for the pound, but little for other currencies. The kurtosis parameters are generally positive, indicating greater likelihood of extreme events than with normal random variables, but they are small in most cases. Similarly, there is little indication that depreciation rates are autocorrelated. The forward premium, in contrast, is highly persistent, with autocorrelations ranging between 0.625 for the franc and 0.884 for the pound.

One way to think about this evidence is to relate it to the *expectations hypothesis*: that forward rates are expected future spot rates. We express this in logarithmic form as $f_t = E_t s_{t+1}$ or $f_t - s_t = E_t s_{t+1} - s_t$. Although we do not observe expected future spot rates, we can get an indication of the accuracy of the expectations hypothesis by comparing mean forward premiums and depreciation rates across currencies. We see in Figure 1 (which uses entries from Table 1) that while the two means are not the same, their differences are small relative to their cross-section variation. Countries with large forward premiums, on average, are also those against which the dollar has depreciated the most.

This sanguine view of the expectations hypothesis changes dramatically when we turn from cross-section to time-series evidence. A huge body of work has established, for the extant flexible exchange rate period, that forward premiums have been negatively correlated with subsequent depreciation rates for exchange rates between most major currencies. Canova and Marrinan (1993), Hodrick (1987), and Levich (1985) provide exhaustive references to the literature. The most common evidence comes from regressions of the form

$$s_{t+1} - s_t = a_1 + a_2(f_t - s_t) + \text{residual}. \tag{1}$$

The expectations hypothesis implies a regression slope $a_2 = 1$, yet most studies estimate a_2 to be negative. Thus we find not only that the expectations hypothesis provides a poor approximation to the data, but that its predictions of future currency movements are in the wrong direction. We report similar evidence in Table 2, where a_2 ranges from -0.461 for the lira to -3.542 for the deutschemark. The t-statistics for these estimates, relative to $a_2 = 1$, range from 1.9 for the lira to 4.3 for the Canadian dollar.

This evidence has motivated, as we noted, a growing number of studies suggesting explanations. Foremost among these is Fama (1984), who labels the difference between the forward rate and the expected future spot rate a risk premium, and proceeds to document its properties. In Fama's interpretation, the forward premium, $f_t - s_t$, includes a risk premium p as well as the expected rate of depreciation q:

$$\begin{aligned} f_t - s_t &= (f_t - E_t s_{t+1}) + (E_t s_{t+1} - s_t) \\ &\equiv p_t + q_t. \end{aligned} \tag{2}$$

The cross-section evidence (Table 1 and Figure 1) suggests that risk premiums are small, on average, but the time series evidence implies they are highly variable. Since the population regression coefficient a_2 is

$$a_2 = \frac{Cov(q, p+q)}{Var(p+q)} = \frac{Cov(q,p) + Var(q)}{Var(p+q)}, \tag{3}$$

it's clear that a constant risk premium p generates $a_2 = 1$. To generate negative values of a_2 we need $Cov(q,p) + Var(q) < 0$. Fama notes that this requires (i) negative covariance between p and q and (ii) greater variance of p than q. These two conditions serve as hurdles that any theoretical explanation of the anomaly must surpass.

We can quantify the components of the forward premium by taking a closer look at the estimates of equation (1) summarized in Table 2. The fitted values of the regression are estimates of expected rates of depreciation q; their means and standard deviations are reported in Table 3. An analogous estimate of the risk premium is

$$p_t = (f_t - s_t) - q_t \cong -a_1 + (1 - a_2)(f_t - s_t),$$

whose means and standard deviations are also reported in Table 3. We see, first, that mean risk premiums are small, as suggested by Figure 1. The means are uniformly smaller than the standard deviations, in some cases by more than an order of magnitude. We also see that the estimates of p and q satisfy the two conditions suggested by Fama. Condition (ii) is apparent in Table 3, where we see that the risk premium has a larger standard deviation than the expected rate of depreciation for each of the six currencies listed. Condition (i) follows from the difference in signs of a_2 and $1 - a_2$: since our estimates of p and q are both linear functions of the forward premium, with slope coefficients a_2 and $1 - a_2$ of opposite sign, the correlation between them is minus one. These properties together allow us to reproduce the slope coefficient a_2 from formula (3). For the deutschemark we have

$$\begin{aligned}
Var(p) &= (0.007980)^2 = 0.6368 \times 10^{-4} \\
Var(q) &= (0.006223)^2 = 0.3872 \times 10^{-4} \\
Cov(p,q) &= -(0.007980)(0.006223) = -0.4966 \times 10^{-4} \\
Var(p+q) &= 0.3087 \times 10^{-5}.
\end{aligned}$$

These imply, by (3), a slope coefficient $a_2 = -3.542$, as we reported in Table 2. These numbers work out simply as a feature of least squares, but they give us a rough idea of the magnitudes involved in conditions (i) and (ii). Since fitted values are projections on a limited information set, we might expect them to be less variable than the conditional means, $q_t = E_t s_{t+1} - s_t$ and $p_t = f_t - E_t s_{t+1}$. The population variances, then, are likely to be larger than we've estimated, and the population correlation is likely to be smaller in absolute value. Finally we note that the strong autocorrelation evident in forward premiums (Table 1) translates, in our estimates, into equal persistence in both components.

In short, Fama's interpretation of the evidence suggests a highly variable risk premium that reverses the sign of the slope parameter a_2 in the forward premium regression relative to what it would be under the expectations hypothesis. But without a theory that generates a risk premium with the required properties, the term risk premium is more a convenient label than an explanation of the anomaly. We consider several potential explanations below.

3 Theoretical Considerations

Before turning to models that might account for the anomaly, we find it useful to consider currency prices in a fairly general theoretical setting. As in the modern dynamic theory of asset pricing, we think of asset prices as being consistent with a *pricing kernel*: a stochastic process governing prices of state-contingent claims. Existence of such a process is guaranteed in any economic environment that precludes arbitrage opportunities. The beauty of this result is its simplicity. It requires only that market prices of traded assets not permit combinations of trades that produce positive payoffs in some states with no initial investment—a departure from covered interest rate parity, for example. In this section, we develop this approach to the pricing of currencies, relate currency variability to the variability of pricing kernels, and examine the relation between the conditional distribution of the kernels and the forward premium anomaly.

3.1 Pricing Kernels

We begin with assets denominated in domestic currency ("dollars"), then move on to those denominated in a foreign currency ("marks"). With respect to dollar assets, consider the dollar value v_t of a claim to the stochastic cash flow of d_{t+1} dollars one period later. If m is a random variable that prices one-period state-contingent claims, the price and cash flow of the asset satisfy the pricing relation,

$$v_t = E_t\left(m_{t+1} d_{t+1}\right), \tag{4}$$

or

$$1 = E_t\left(m_{t+1} R_{t+1}\right), \tag{5}$$

where $R_{t+1} = d_{t+1}/v_t$ is the one-period return on the asset. We refer to the possible realizations of m as state prices, to the stochastic process governing m as the pricing kernel, and to the random variable m as the state-price function or simply the pricing kernel. In economies with a representative agent, m is the nominal intertemporal marginal rate of substitution and (5) is a first-order condition. More generally, there exists a positive random variable m satisfying the pricing relation (5) for returns R on all traded assets if the economy admits no pure arbitrage opportunities. When the economy has a complete set of markets for state-contingent claims, m is the unique solution to (5), but otherwise there is generally a range of choices of m that satisfy the pricing relation for all returns. These issues, and the relevant literature, are reviewed by Duffie (1992).

The pricing kernel, embodied in m, and the pricing relation (5) are the basis of most modern theories of bond pricing: given a pricing kernel, we use (5) to compute prices and yields for bonds of all maturities. Denote by b_t^n the price of an n-period zero-coupon bond: the claim to one dollar at date $t+n$ in all states. Since the return on an $(n+1)$-period bond is b_{t+1}^n/b_t^{n+1}, we can compute bond prices recursively from

$$b_t^{n+1} = E_t\left(m_{t+1} b_{t+1}^n\right), \tag{6}$$

starting with $b_t^0 = 1$ (a dollar today costs a dollar). The price of a one-period bond, for example, is $b_t^1 = E_t m_{t+1}$. Continuously-compounded bond yields y are related to prices by

$$b_t^n = \exp\left(-y_t^n n\right).$$

The short rate r_t is the yield y_t^1 on a one-period bond:

$$r_t = -\log b_t^1 = -\log E_t m_{t+1}. \tag{7}$$

We return to this equation when we consider exchange rates.

When we consider assets with returns denominated in deutschemarks, we might adopt an analogous approach and use a random variable m^* to value them. Alternatively, we could convert mark returns into dollars and value them using m. The equivalence of these two procedures gives us a connection between exchange rate movements and state prices in the two currencies, m and m^*. If we use the first approach, mark returns R_t^* satisfy

$$1 = E_t\left(m_{t+1}^* R_{t+1}^*\right). \tag{8}$$

If we use the second approach, with $S = \exp(s)$ denoting the spot price, in dollars, of one mark, then

$$1 = E_t\left(m_{t+1}(S_{t+1}/S_t)R_{t+1}^*\right).$$

If the mark asset and currencies are both traded, there are obvious arbitrage opportunities unless the return satisfies both conditions:

$$E_t\left(m_{t+1}^* R_{t+1}^*\right) = E_t\left(m_{t+1}(S_{t+1}/S_t)R_{t+1}^*\right).$$

This equality ties the rate of depreciation of the dollar to the random variables m and m^* that govern state prices in dollars and marks. Certainly this relation is satisfied if $m_{t+1}^* = m_{t+1}S_{t+1}/S_t$. This choice is dictated when the economy has a complete set of markets for currencies and state-contingent claims. With incomplete markets, the choices of m and m^* satisfying (5,8) are not unique, but we will see that there is no loss of generality in choosing them to satisfy the same equation.

We summarize the connection between pricing kernels and currency prices in

Proposition 1 *Consider stochastic processes for the depreciation rate, S_{t+1}/S_t, and returns R_{t+1} and R_{t+1}^* on dollar and mark denominated assets. If these processes do not admit arbitrage opportunities, then we can choose the pricing kernels m and m^* for dollars and marks to satisfy both*

$$m_{t+1}^*/m_{t+1} = S_{t+1}/S_t \tag{9}$$

and the pricing relations (5,8).

Proof. Consider dollar returns on the complete set of traded assets, including the dollar returns $(S_{t+1}/S_t)R_{t+1}^*$ on mark-denominated assets. If these returns do not admit arbitrage opportunities, then there exists a positive random variable m_{t+1} satisfying (5) for dollar returns on each asset (Duffie 1992, Theorem 1A and extensions). For any such m, the choice $m_{t+1}^* = m_{t+1}S_{t+1}/S_t$ automatically satisfies (8).

The proposition tells us that of the three random variables, m_{t+1}, m_{t+1}^*, and S_{t+1}/S_t, one is redundant, and can be constructed from the other two. We start with the two pricing kernels, which brings out the essential symmetry between the two currencies. Despite the algebra, the intuition is relatively straightforward: if we know prices of state-contingent claims in dollars and marks, we can compute the implied exchange rate from their ratio. The only ambiguity stems from combinations of state-contingent claims that are not traded.

Before turning to forward rates, we note that the observed properties of exchange rates and other asset prices imply that m and m^* are closely related. To see this, note that (9) implies

$$Var\left(s_{t+1} - s_t\right) = Var\left(\log m^*_{t+1} - \log m_{t+1}\right).$$

For, say, the deutschemark, we estimate the left side to be 0.0338^2—the square of about 3 percent per month (see Table 1). The right side we can estimate from Hansen-Jagannathan bounds. A ballpark estimate of the standard deviation of $\log m$ is about 0.4; see, for example, Bekaert and Hodrick (1992, Table XI), who report estimated lower bounds as high as 0.78. If the variances of $\log m$ and $\log m^*$ are the same, then their correlation, ρ, is the solution to

$$0.0338^2 = 2(1-\rho)(0.4)^2,$$

so that $\rho = 0.996$. Stated more simply: the small variance of the depreciation rate relative to the pricing kernels implies that the two kernels are highly correlated. The correlation does not fall appreciably with smaller estimates of the standard deviation of $\log m$. With a standard deviation of 0.1, for example, the correlation falls only to 0.943. With most estimates of the variance of pricing kernels we have

Remark 1 *Most of the variation in the pricing kernels, m and m^*, is common to both.*

An explicit factor structure is suggestive of how the interdependence of the two pricing kernels might be modeled. Suppose $\log m$ and $\log m^*$ combine a common factor x and idiosyncratic factors z and z^*:

$$\begin{aligned} \log m_t &= x_t + z_t \\ \log m^*_t &= x_t + z^*_t. \end{aligned}$$

If the three factors are independent, and the idiosyncratic factors have the same variance, the numbers in the preceding paragraph imply

$$\begin{aligned} Var(x) &= 0.3993^2 \\ Var(z) &= Var(z^*) = 0.0239^2. \end{aligned}$$

Almost all of the variance in state prices, then, is in the common factor x.

The strong correlation between the two pricing kernels suggests that it may be difficult to detect the relatively small idiosyncratic features that bear on the pricing of currencies: exchange rates depend only on differences between the logarithms of m^* and m, yet most of the variation in these random variables is common and therefore does not affect currency prices.

3.2 Forward Rates and Risk Premiums

Given pricing kernels for two currencies and equation (9) for spot exchange rates, we derive the forward premium and its components from the pricing relation. Consider a forward contract specifying at date t the exchange at $t+1$ of one mark and $F_t = \exp(f_t)$ dollars, with the forward rate F_t determined at date t as the notation suggests. This contract produces net dollar cash flows at date $t+1$ of $F_t - S_{t+1}$. Since it involves no payments at date t, the pricing relation (4) implies

$$0 = E_t\left[m_{t+1}(F_t - S_{t+1})\right].$$

If we divide by S_t and apply Proposition 1 we find

$$(F_t/S_t)E_t(m_{t+1}) = E_t\left(m_{t+1}S_{t+1}/S_t\right) = E_t(m^*_{t+1}),$$

or

$$F_t/S_t = E_t m^*_{t+1}/E_t m_{t+1}.$$

Thus the forward premium is

$$f_t - s_t = \log E_t m^*_{t+1} - \log E_t m_{t+1}. \tag{10}$$

This equation and the definition of the short rate, equation (7), give us

$$f_t - s_t = r_t - r^*_t, \tag{11}$$

the familiar covered interest rate parity condition.

Now consider the components of the forward premium. The expected rate of depreciation is, from (9),

$$q_t \equiv E_t s_{t+1} - s_t = E_t \log m^*_{t+1} - E_t \log m_{t+1}. \tag{12}$$

Thus we see that the first of Fama's components is governed by the means of the logarithms of the pricing kernels. The risk premium is, from (2,10),

$$p_t = \left(\log E_t m^*_{t+1} - E_t \log m^*_{t+1}\right) - \left(\log E_t m_{t+1} - E_t \log m_{t+1}\right), \tag{13}$$

the difference between the "log of the expectation" and the "expectation of the log" of the pricing kernels m and m^*. If this seems overly complicated, we should recall that nonlinearity is the essence of the risk aversion that underlies most of finance. The examples in the following sections will help to give these expressions more concrete form.

With additional structure we can be more specific about the factors that affect the risk premium. Many popular bond pricing models start with conditionally log-normal pricing kernels: $\log m_{t+1}$ and $\log m^*_{t+1}$ are conditionally normal with means (μ_{1t}, μ^*_{1t}) and variances (μ_{2t}, μ^*_{2t}). With this structure, one-period bond prices are

$$\begin{aligned} E_t m_{t+1} &= \exp(\mu_{1t} + \mu_{2t}/2) \\ E_t m^*_{t+1} &= \exp(\mu^*_{1t} + \mu^*_{2t}/2) \end{aligned}$$

and, from (10), the forward premium is

$$f_t - s_t = (\mu^*_{1t} - \mu_{1t}) + (\mu^*_{2t} - \mu_{2t})/2.$$

The first term on the right is the expected rate of depreciation and the second is the risk premium. Fama's conditions require, in this case, (i) negative correlation between differences in conditional means and variances of the two pricing kernels and (ii) greater variation in the one-half the difference in the conditional variances.

If the conditional distributions of $\log m$ and $\log m^*$ are not normal, the risk premium depends on higher moments. For an arbitrary distribution, equation (12) tells us (again) that only the means affect the expected rate of depreciation. The risk premium is given, in general, by (13), but if all of the conditional moments of $\log m$ exist, $\log E_t m_{t+1}$ can be expanded

$$\log E_t m_{t+1} = \sum_{j=1}^{\infty} \kappa_{jt}/j!, \tag{14}$$

where κ_{jt} is the jth cumulant for the conditional distribution of $\log m_{t+1}$. Equation (14) is an expansion of the cumulant generating function (the logarithm of the moment generating function) evaluated at one; see Stuart and Ord (1987, chs 3,4). The cumulants are closely related to moments, as we can see from the first four: $\kappa_{1t} = \mu_{1t}$, $\kappa_{2t} = \mu_{2t}$, $\kappa_{3t} = \mu_{3t}$, and $\kappa_{4t} = \mu_{4t} - 3(\mu_{2t})^2$. The notation is standard, with μ_{jt}, for $j > 1$, denoting the jth central conditional moment of $\log m_{t+1}$. For the normal distribution, the cumulants are zero after the first two, so equation (14) gives us a way of quantifying the impact of departures from normality. If the foreign kernel has a similar representation, we can express the forward premium as

$$f_t - s_t = \sum_{j=1}^{\infty} (\kappa^*_{jt} - \kappa_{jt})/j!,$$

and the risk premium as

$$p_t = \kappa^*_{-1,t} - \kappa_{-1,t}, \tag{15}$$

where

$$\kappa_{-1,t} = \sum_{j=2}^{\infty} \kappa_{jt}/j!, \qquad \kappa^*_{-1,t} = \sum_{j=2}^{\infty} \kappa^*_{jt}/j!.$$

We refer to the sums $\kappa_{-1,t}$ and $\kappa^*_{-1,t}$ generically as "higher-order cumulants" or moments.

With equations (15) and (12) describing risk premiums and expected depreciation, we have

Remark 2 *If conditional moments of all order exist for the logarithms of the two pricing kernels, m and m^*, then Fama's necessary conditions for the forward premium anomaly imply*

(i) negative correlation between differences in conditional means, $\kappa^*_{1t} - \kappa_{1t}$, and differences in higher-order cumulants, $\kappa^*_{-1,t} - \kappa_{-1,t}$; and

(ii) greater variation in the latter.

In short, the forward premium anomaly requires, in the theory, an inverse relation between differences in first moments and high-order moments of logarithms of pricing kernels. In the following sections, we use examples to illustrate how such a relation might be incorporated into a dynamic theory of asset pricing.

4 Example 1: Alternating Regimes

Our first example adapts the popular two-regime model of currency fluctuations used by Engel and Hamilton (1990) and Evans and Lewis (1993) to the theoretical setting of the last section. These two papers find that US dollar depreciation rates against major currencies are approximated well by a model that alternates between two "regimes" with different conditional means and variances of depreciation rates.

We can reproduce these features, and extend them to the forward premium anomaly, in a version of our theoretical framework with conditionally log-normal pricing kernels. Let us say that the behavior of the two pricing kernels varies across two regimes as indicated by a random variable z. The regime is part of agents' information sets. When $z = 1$ the world is in regime 1, and when $z = 2$ the world is in regime 2. The regimes follow a Markov chain with transition probabilities

$$\text{Prob}\{z_{t+1} = j | z_t = i\} = \begin{bmatrix} (1-\theta)\pi + \theta & (1-\theta)(1-\pi) \\ (1-\theta)\pi & (1-\theta)(1-\pi) + \theta \end{bmatrix}, \qquad (16)$$

where π is the unconditional probability of regime 1 and θ is the autocorrelation of z.

To account for the persistent patterns of appreciation and depreciation, we let the first moments of $\log m$ and $\log m^*$ vary across the two regimes. Let us say, to be concrete, that the first moment of the dollar pricing kernel alternates between two values,

$$E_t(\log m_{t+1}) = \begin{cases} \delta_1 & \text{when } z_t = 1 \\ \delta_2 & \text{when } z_t = 2, \end{cases}$$

and that $\log m^*$ has analogous parameters δ_i^*. That gives us, from (12), an expected rate of depreciation

$$q_t = \begin{cases} q_1 = \delta_1^* - \delta_1 & \text{when } z_t = 1 \\ q_2 = \delta_2^* - \delta_2 & \text{when } z_t = 2. \end{cases}$$

Engel and Hamilton (1990) and Evans and Lewis (1993) report estimates of q_1 and q_2.

To account for the anomaly, we need similar variation in higher-order moments across regimes. Let us say that the conditional variance of the dollar kernel alternates between two values,

$$Var_t(\log m_{t+1}) = \begin{cases} \gamma_1 & \text{when } z_t = 1 \\ \gamma_2 & \text{when } z_t = 2, \end{cases}$$

and that $\log m^*$ has analogous conditional variances γ_i^*. This structure delivers a risk premium

$$p_t = \begin{cases} p_1 = (\gamma_1^* - \gamma_1)/2 & \text{when } z_t = 1 \\ p_2 = (\gamma_2^* - \gamma_2)/2 & \text{when } z_t = 2. \end{cases}$$

While both Engel and Hamilton (1990) and Evans and Lewis (1993) report estimates of the conditional variance of the depreciation rate across regimes, this information is not enough to identify p_1 and p_2.

Since both the risk premium and expected depreciation are functions of the regime, in this example, they are related to each other. It's fairly easy to construct examples in which they are negatively correlated and reproduce the anomalous forward premium regressions. Consider an example based on Evans and Lewis's (1993, Table II) monthly estimates for the dollar-deutschemark rate (Engel and Hamilton's quarterly estimates are similar). Their estimates imply that regimes are highly persistent ($\theta = 0.924$) and that regime 1 occurs with greater frequency ($\pi = 0.865$). The expected rate of depreciation alternates between $q_1 = -0.000308 = -0.037/1200$ in regime 1 (the strong dollar regime) and $q_2 = 0.006892 = 0.827/1200$ in regime 2 (the strong mark regime). These values imply $Var(q) = 0.0002461^2$. If we posit a risk premium that is lower in regime 2, and has a greater variance, we can account for the anomaly. An example that replicates our estimate of a_2 from Table 2 is $p_2 = p_1 - 0.0009233$ for any p_1.

The obvious strength of this example is that it builds on the striking trends in spot exchange rates documented by Engel and Hamilton (1990) and Evans and Lewis (1993). Other considerations, however, suggest to us that this is, at present, not a persuasive explanation of the anomaly. First, the standard deviation of the expected rate of depreciation implied by the Evans-Lewis estimates is more than an order of magnitude smaller than we estimated in Section 2. Their estimates for the deutschemark imply a standard deviation that is smaller by a factor of 25 (0.0002461 vs. the estimate of 0.006223 in Table 3). As a result, the standard deviation of the risk premium required to reproduce the regression slope (in the example, 0.0003156) is smaller by the same factor than the estimate in Table 3, as is the standard deviation of the forward premium. This suggests that while there are sharp differences across "regimes," most of the variation in risk premiums and expected depreciation implied by forward premium regressions is missed by these models. Second, our example can reproduce the negative slope coefficients of Table 2 in any sample that includes observations of both regimes, but not within a regime. Yet when we estimate a_2 over shorter samples, or within the estimated regimes, we find that the anomaly remains. Third, the variances that determine the risk premium in our example are not identified by the behavior of the spot exchange rate. The regime models estimate the conditional variance of the spot rate across regimes,

$$Var_t(s_{t+1} - s_t) = Var_t\left(\log m^*_{t+1} - \log m_{t+1}\right)$$

not the difference between the conditional variances,

$$p_t = Var_t\left(\log m^*_{t+1}\right)/2 - Var_t\left(\log m_{t+1}\right)/2.$$

Without additional information, we have no way of knowing whether our choice of risk premiums in the two states is a reasonable one, or simply a fortunate choice. Moreover, other work (Bekaert and Hodrick 1993, Domowitz and Hakkio 1985), has failed to detect the relation this implies between differences in conditional variances of depreciation rates and risk premiums. Finally, this model generates only two different yield curves in each country, one in each regime. In this sense, it fails to utilize the information on pricing kernels that earlier work has derived from bond prices. We correct this oversight in the next section.

5 Example 2: Affine Yield Models

Our second example builds on the modern theory of bond pricing, in particular the affine yield class exemplified by Cox, Ingersoll, and Ross (1985). Similar examples have been studied by Nielsen and Saá-Requejo (1993) and Saá-Requejo (1994). One advantage of models in this class is that they are relatively simple analytically: both interest rates and risk premiums are linear functions of underlying state variables. Another advantage is that we can use the knowledge of pricing kernels accumulated in an extensive empirical literature on bond prices. Finally, we will see that many of the models in this class produce automatically the contrary movements in the conditional mean and variance of pricing kernels that are required by Fama's condition (i) in log-normal settings. Condition (ii), however, is less easily satisfied.

To see how this class of models works, it's useful to start with a two-country, discrete-time version of the Cox-Ingersoll-Ross (1985) model, adapted from Sun (1992). The heart of the model is a state variable z that obeys the "square root process"

$$z_{t+1} = (1-\varphi)\delta + \varphi z_t + \sigma z_t^{1/2} \epsilon_{t+1}, \tag{17}$$

with $0 < \varphi < 1$, $\delta > 0$, and $\{\epsilon_t\} \sim \text{NID}(0,1)$. Despite the nonlinearity of the innovation, this is a first-order autoregression with mean δ and autocorrelation φ. The effect of the square-root term is to reduce the conditional variance of the innovation to zero as z approaches zero. If $(1-\varphi)\delta > \sigma^2/2$, z can only become negative with extreme negative realizations of ϵ. In discrete time this happens with positive probability, but we can make the probability as small as we like by choosing a small time interval, and in the continuous time limit the state variable z is always positive. Given z, the pricing kernel is characterized by

$$-\log m_{t+1} = z_t + \lambda z_t^{1/2} \epsilon_{t+1}. \tag{18}$$

Since ϵ_{t+1} is normal, $\log m_{t+1}$ is conditionally normal.

This structure is an example of the conditionally log-normal pricing kernels described in Section 3. The conditional mean and variance,

$$\begin{aligned} E_t \log m_{t+1} &= -z_t \\ Var_t \log m_{t+1} &= \lambda^2 z_t, \end{aligned}$$

are both linear in the state variable z, and have a perfect negative correlation. The short rate is

$$r_t = -\log E_t m_{t+1} = -\left(E_t \log m_{t+1} + (1/2)\, Var_t \log m_{t+1}\right) = \left(1 - \lambda^2/2\right) z_t. \tag{19}$$

As long as $1 > \lambda^2/2$, the short rate inherits the positive sign of z. Yields on bonds of longer maturities are computed by applying (6) repeatedly. The values $\delta = 0.00728$, $\sigma = 0.0103$, $\varphi = 0.976$, and $\lambda = -.885$ reproduce the mean, standard deviation, and autocorrelation of the short rate and the average slope of the yield curve for US treasury securities (Backus, 1994).

We complete the example by appending to this Cox-Ingersoll-Ross model of dollar bonds a similar model of deutschemark bonds. If the deutschemark kernel has the same structure based on a second state variable z^*, then the deutschemark short rate is $r_t^* = (1 - \lambda^2/2)\, z_t^*$ and the forward premium is $f_t - s_t = (1 - \lambda^2/2)\,(z_t - z_t^*)$. The components of the forward premium are the expected rate of depreciation,

$$q_t = (z_t - z_t^*),$$

and the risk premium,

$$p_t = -(1/2)\lambda^2(z_t - z_t^*).$$

Thus the linearity of the conditional mean and variance translate into forward premium components that are linear functions of the differential $z - z^*$. More important, this structure automatically generates the negative correlation between p and q of Fama's condition (i). Saá-Requejo (1994, p 21) makes the same observation of a similar model.

The negative correlation in this model between the risk premium and expected depreciation is striking, since it indicates that the inverse relation between first and second moments of pricing kernels required by Fama's condition (i) is a standard feature of a popular model of bond pricing. A closer look suggests, however, that this partial success is difficult to extend further. The problem is Fama's condition (ii): for the risk premium to be more variable than expected depreciation we need $\lambda^2/2 > 1$, the reverse of the inequality that produced a positive interest rate. Apparently the model cannot explain the anomaly without generating negative rates of interest. Futhermore, parameter estimates are typically consistent with the positive interest rate restriction. The parameter values listed above, for example, result in positive interest rates, and the implied forward premium regression has slope

$$a_2 = \frac{1}{1 - \lambda^2/2} = 1.644,$$

which is not only positive but greater than one. In short, this two-currency version of the Cox-Ingersoll-Ross model delivers Fama's condition (i), but condition (ii) and the negative regression slope follow only if interest rates are negative.

One possible solution is to extend the theory to the broader class of affine yield models described by Duffie and Kan (1993). Despite the wide range of behavior included in this class, we find that the problem remains if we limit ourselves to symmetric models:

Proposition 2 *In the Duffie-Kan class of affine yield models, suppose (a) pricing kernels are symmetric across currencies and (b) interest rates are strictly positive. Then the model cannot account for the forward premium anomaly.*

A proof is given in Appendix A.

Proposition 2 tells us that the difficulty we had with the two-country Cox-Ingersoll-Ross model extends to the broader class of symmetric affine yield models. There are, however, asymmetric affine models that reproduce the negative regression slopes we reported in Table 2. One of the simplest examples is based on a single state variable z obeying a process like (17), with pricing kernels

$$\begin{aligned} -\log m_{t+1} &= \alpha z_t + (2\beta z_t)^{1/2}\, \epsilon_{t+1} \\ -\log m^*_{t+1} &= \alpha^* z_t + (2\beta^* z_t)^{1/2}\, \epsilon_{t+1}. \end{aligned}$$

This model is asymmetric, by our definition, if $(\alpha, \beta) \neq (\alpha^*, \beta^*)$. In this setting short rates are $r_t = (\alpha - \beta) z_t$ and $r^*_t = (\alpha^* - \beta^*) z_t$, expected depreciation is

$$q_t = (\alpha - \alpha^*) z_t,$$

and the risk premium is

$$p_t = -(\beta - \beta^*) z_t.$$

If $\alpha > \beta$ and $\alpha^* > \beta^*$, interest rates a positive. Moreover, we can generate an inverse relation between the forward premium and the interest differential if $\alpha - \alpha^*$ and $\beta - \beta^*$ have the same sign, and the latter is larger in absolute value. An example is $\alpha = 1$, $\alpha^* = 0.844$, $\beta = 0.4$, and $\beta^* = 0.2$, which generates a regression slope of $a_2 = -3.542$, the estimate for the deutschemark reported in Table 2. The choices of α and β in this example are approximately those implied by the parameter values we cited for the Cox-Ingersoll-Ross model. The foreign currency parameters result in a foreign interest rate that is higher and more variable than the dollar interest rate.

In principle, an asymmetric model might provide a convincing explanation of the anomaly, but examples we have studied to date suggest to us that asymmetries often introduce counterfactual features in other dimensions. One feature is the small mean risk premiums we noted in Figure 1 and Table 3. Symmetric models generate zero mean risk premiums as a matter of course, since the risk premium is proportional to $z - z^*$, which has mean zero by construction. With asymmetric models it is less easy to accomplish the same thing. Our one-factor example has a mean risk premium of

$-(\beta - \beta^*)\delta < 0$. We can imagine correcting this fault by introducing a parameter that shifts the mean of p to zero, as with the modified kernel

$$-\log m^*_{t+1} = \alpha^* z_t + \left[2(\beta - \beta^*)\delta + 2\beta^* z_t\right]^{1/2} \epsilon_{t+1},$$

but we worry that the proliferation of parameters threatens to void the theory of predictive content. This drawback does not constitute an airtight case against asymmetry, and symmetry itself is not something we want to take literally, but they increase the appeal to us of models outside the affine class.

We summarize briefly. The class of affine yield models includes some of the simplest examples that reproduce the inverse relation between expected depreciation and the forward premium, but it is not yet clear whether it can provide a persuasive solution of the anomaly. Certainly symmetric models cannot account for the anomaly without generating negative interest rates. With more work we may be able to reconcile asymmetric affine models with both time-series and cross-section evidence, but this seems to us, at present, to require a fortuitious combination of parameter values. Nevertheless, we use the intuition of this class of models in a related example in the next section.

6 Example 3: Continuous Regimes

Examples 1 and 2 suggest that while we can construct theoretical models that replicate the puzzling negative slope of forward premium regressions, they are neither trivial nor easily reconciled with the behavior of currency prices and interest rates. Our third example is an attempt to retain some of the analytical simplicity of affine models with the nonlinearity of regime models.

We begin with a scenario. Suppose that the US, or some other country, appears to be entering a recession. As events unfold, we see, relative to the rest of the world, (i) a relatively high degree of uncertainty over future events, (ii) expansionary monetary policy, and (iii) expected depreciation of the currency. Uncertainty is a common feature of downturns, where we generally see an increase in the cross-section dispersion of both aggregate quantities and asset prices, so (i) is consistent with other work. The combination of monetary policy and expected depreciation might be attributed, in the business press, to the central bank placing greater emphasis on restoring domestic growth than on defending the currency. In our view, this combination of events seems plausible. The reduction of interest rates in the US in 1992, for example, was widely interpreted as an indication that the dollar would fall relative to the mark (and it did).

These events imply, in our framework, opposite movements in the conditional mean and variance of the pricing kernel. The weak currency implies a decline in the conditional mean of the domestic pricing kernel; see (12). An increase in the conditional variance of the kernel might be a consequence, as we have seen, of a recession, or other domestic turbulence. These two moments push the short rate in opposite directions, but for the short rate to fall we need the variance, or higher moments more generally, to dominate. If so, then the events described imply a rise in the expected rate of depreciation and a larger fall in the risk premium; see equation (15).

One appealing mathematical representation of our scenario relates domestic and foreign pricing kernels to two sets of state variables. Set one is common across countries, set two is country-specific. If the common state variables affect the two kernels symmetrically, then they play no role in currency pricing. This is apparent from equations (11,12,13), where the forward premium and its components are expressed as differences between moments of the two kernels. Since the common state variables have no impact on currency pricing, we focus here on the idiosyncratic factors.

We model the idiosyncratic part of the pricing kernels with equations similar to those of the two-currency Cox-Ingersoll-Ross model of the last section. The difference is that the state variable driving changes in conditional moments is bounded above and below. As a result, we can eliminate the tendency in affine models to generate negative interest rates. The regime model has a similar feature: with a finite number of regimes, moments are bounded (obviously) by their maximum and minimum values.

Let us say, then, that the conditional mean and variance of the pricing kernel are linear functions of a state variable z:

$$-\log m_{t+1} = \mu + z_t + \lambda z_t^{1/2} \epsilon_{t+1}. \tag{20}$$

The only departure from equation (18) is the intercept μ. We bound the state variable below by zero and above by one by postulating a law of motion of the form

$$z_{t+1} = -\varphi_1 z_t + \varphi_2(1 - z_t) + \sigma \left[z_t(1 - z_t)\right]^{1/2} \epsilon_{t+1}, \tag{21}$$

with $0 < \varphi_1, \varphi_2$ and $\{\epsilon_t\} \sim \text{NID}(0,1)$. This equation, like (17), is a discrete-time approximation to a continuous-time process in which z is driven away from its boundaries at zero and one, back into the unit interval; see Karlin and Taylor (1991, pp 239–241).

We have two stories for how a bounded state variable might arise. One is that targeting of interest rates keeps domestic interest rates from deviating too much from world levels. Targeting of interest rates is a useful way of thinking about monetary policy which has proved, in related work by Balduzzi, Bertola, and Foresi (1993), to have significant implications for the term structure of interest rates.

Our second story involves learning about an unobservable regime in the spirit of Lewis (1989). Suppose the world alternates, as in Section 4, between two regimes. If the regimes are not part of agents' information sets, they can learn about them from market prices. This learning could take the form of a subjective probability π_t that the world is in regime 1, which we can think of as a (naturally bounded) state variable. Although our law of motion (21) is not easily put into the form of a Bayesian updating formula, as in Lewis (1989, eq 9) or Gray (1993, eq 11), we think this story helps to motivate bounds on the state variable z.

Given the pricing kernel (20) and law of motion (21), the mathematics of forward and spot exchange rates is much like the last section. If we posit a similar process for the foreign country, short rates are

$$r_t = \mu + \left(1 - \lambda^2/2\right) z_t, \qquad r_t^* = \mu + \left(1 - \lambda^2/2\right) z_t^*,$$

the forward premium is

$$f_t - s_t = \left(1 - \lambda^2/2\right) (z_t - z_t^*),$$

and the forward premium components are

$$\begin{aligned} q_t &= (z_t - z_t^*) \\ p_t &= -(1/2)\lambda^2(z_t - z_t^*). \end{aligned}$$

This reproduces the negative slope of forward premium regressions if $\lambda^2/2 > 1$. Unlike our Cox-Ingersoll-Ross example, this need not be inconsistent with positive interest rates. For $\lambda^2/2 > 1$ the short rate varies between $\mu + 1 - \lambda^2/2$ and μ. If we set $\mu = \lambda^2/2 - 1$, this is always positive. If we think of this as the idiosyncratic factor of a multi-factor model, this device simply adds a nonnegative amount to the short rate implied by the other factors.

This model needs more work before we can regard it as a likely solution to the forward premium anomaly, but it has a number of features that we find appealing. First and foremost, it is capable of reproducing the negative slope of forward premium regressions with strictly positive interest rates and zero mean risk premium. With $\lambda = 1.601$, for example, the regression slope is -3.542, our estimate in Table 2 for the deutschemark, and with $\mu = \lambda^2/2 - 1 = 0.282$ the interest rate is always positive. Second, the model is consistent with the finding of Brenner, Harjes, and Kroner (1993) that while volatility is generally increasing in the level of the short rate, there is a component of volatility that can be high even with low interest rates. If equations (20,21) govern the short rate, then the model implies a counterfactual inverse relation between volatility and the short rate. But if example 3 is the idiosyncratic component of a multi-factor model, with the common factor governed by a Cox-Ingersoll-Ross model, then we would expect to see something very close to their finding. Third, this structure is capable of accounting for the success of regime models of exchange rate movements. If regimes are persistent and we learn relatively quickly which one we are in, the data are likely to be well approximated by a model in which we disregard intermediate probabilities.

7 Final Remarks

We would like to say that we now have the solution to the forward premium anomaly, but such a claim is clearly premature. Nevertheless, we think we have made progress on some fronts.

One front is our characterization of the anomaly in terms of conditional moments of pricing kernels: expected depreciation is governed by conditional means, and the risk premium by higher moments [equation (15)]. For the risk premium to be more variable than the expected depreciation, as Fama (1984) suggests it is in the data, we need more variation in higher moments than in means. This message is uncomfortably close to "things are complicated," but we think it conforms with the growing body of statistical work on nonlinear dynamics in asset prices. For the forward premium anomaly, these nonlinearities are more than a refinement of an approximately linear theory, they are essential.

Another front concerns the common and country-specific factors governing asset returns worldwide. We showed, absent barriers on the international trade of assets, that the variance of spot exchange rate changes for the G7 currencies suggests that most of the variation in state prices is common. The country-specific factors that affect currencies, in other words, are small relative to common "world" factors. This interpretation is a blessing to analysts of fixed income securities, who generally operate with casual disregard of foreign factors. But it is a curse for currency analysts, who might otherwise hope to use information from fixed income studies to justify pricing kernels used to value currencies. The information about pricing kernels needed to characterize currency prices must be gleaned, it seems, from currency prices themselves.

Finally, we used three examples to indicate how dynamic asset pricing theory might be developed to provide a resolution of the anomaly. All three are capable of accounting for the anomaly. Of the three, we think the last is the most interesting, since it combines ingredients of regime models of exchange rates and affine models of bond pricing and reproduces what we feel is a plausible combination of events. Perhaps further work will tell us how well this structure mimics the properties of interest rates and currency prices more generally.

References

Backus, David (1994), "Bond pricing in discrete time," unpublished notes, Stern School of Business, New York University, February.

Backus, David, Allan Gregory, and Chris Telmer (1993), "Accounting for forward rates in markets for foreign currency," *Journal of Finance* 48, 1887-1908.

Balduzzi, Pierluigi, Giuseppe Bertola, and Silverio Foresi (1993), "A model of target changes and the term structure of interest rates," NBER Working Paper No. 4347, April.

Bansal, Ravi (1991), "Can nonseparabilities explain exchange rate movements and risk premia?" unpublished manuscript, Carnegie Mellon University.

Bekaert, Geert, and Robert Hodrick (1992), "Characterizing predictable components in excess returns on equity and foreign exchange markets," *Journal of Finance* 47, 467-509.

Bekaert, Geert, and Robert Hodrick (1993), "On biases in the measurement of foreign exchange risk premiums," *Journal of International Money and Finance* 12, 115-138.

Bekaert, Geert, Robert Hodrick, and David Marshall (1992), "The implications of first-order risk aversion for asset market risk premiums," unpublished manuscript, Northwestern University, December.

Brenner, Robin, Richard Harjes, and Kenneth Kroner (1993), "Another look at alternative models of the short-term interest rate," unpublished manuscript, University of Arizona, September.

Canova, Fabio, and Jane Marrinan (1993), "Predicting excess returns in financial markets," unpublished manuscript, Brown University, January; forthcoming *European Economic Review.*

Cox, John, Jonathan Ingersoll, and Stephen Ross (1985), "A theory of the term structure of interest rates," *Econometrica* 53, 385-407.

Cumby, Robert (1988), "Is it risk? Explaining deviations from uncovered interest parity," *Journal of Monetary Economics* 22, 279-299.

Domowitz, Ian, and Craig Hakkio (1985), "Conditional variance and the risk premium in the foreign exchange market," *Journal of International Economics* 19, 47-66.

Duffie, Darrell (1992), *Dynamic Asset Pricing Theory.* Princeton, NJ: Princeton University Press.

Duffie, Darrell, and Rui Kan (1993), "A yield-factor model of interest rates," unpublished manuscript, Stanford University, September.

Engel, Charles, and James Hamilton (1990), "Long swings in the dollar: Are they in the data and do markets know it?" *American Economic Review* 80, 689-713.

Evans, Martin, and Karen Lewis (1993), "Do long-term swings in the dollar affect estimates of the risk premium?" unpublished manuscript, Stern School of Business, New York University, June.

Fama, Eugene (1984), "Forward and spot exchange rates," *Journal of Monetary Economics* 14, 319-338.

Frankel, Jeffrey, and Charles Engel (1984), "Do asset demand functions optimize over the mean and variance of real returns? A six currency test," *Journal of International Economics* 17, 309-323.

Gray, Stephen (1993), "Regime-switching models: A new approach," unpublished manuscript, Stanford University, November.

Hodrick, Robert (1987), *The Empirical Evidence on the Efficiency of Forward and Futures Foreign Exchange Markets.* New York: Harwood Academic Publishers.

Karlin, Samuel, and Howard Taylor (1981), *A Second Course in Stochastic Processes.* San Diego: Academic Press.

Levich, Richard (1985), "Empirical studies of exchange rates: Price behavior, rate determination, and market efficiency," in R. Jones and P. Kenen, (eds.), *Handbook of International Economics, Volume II.* Amsterdam: North Holland, 979-1040.

Lewis, Karen (1989), "Changing beliefs and systematic rational forecast errors with evidence from foreign exchange," *American Economic Review* 79, 621-636.

Macklem, Tiff (1991), "Forward exchange rates and risk premiums in artificial economies," *Journal of International Money and Finance* 10, 365-391.

Mark, Nelson (1988), "Time varying betas and risk premia in the pricing of forward foreign exchange contracts," *Journal of Financial Economics* 22, 335-354.

Nielsen, Lars, and Jesús Saá-Requejo (1993), "Exchange rate and term structure dynamics and the pricing of derivative securities," unpublished manuscript, INSEAD, May.

Saá-Requejo, Jesús (1994), "The dynamics and the term structure of risk premia in foreign exchange markets," unpublished manuscript, INSEAD, May.

Stuart, Alan, and J.K. Ord (1987), *Kendall's Advanced Theory of Statistics, Volume 1.* New York: Oxford University Press.

Sun, Tong-sheng (1992), "Real and nominal interest rates: A discrete-time model and its continuous-time limit," *Review of Financial Studies* 5, 581-611.

A Proof of Proposition 2

Duffie and Kan (1993), translated into discrete time, show that the affine class is based on an n-dimensional vector of state variables z following

$$z_{t+1} - z_t = (I - \Phi)(\delta - z_t) + \Sigma V(z)^{1/2}\epsilon_{t+1}, \tag{22}$$

where $\{\epsilon_t\} \sim \mathrm{NID}(0, I)$ and $V(z)$ is a diagonal matrix with typical element

$$v_i(z) = \alpha_i + \beta_i' z.$$

This process requires, obviously, that the volatility functions v_i not turn negative (subject to the qualification that we are talking here about a discrete time approximation). The set of feasible z's (those for which volatility is positive) is thus

$$D = \{z : v_i(z) > 0 \text{ all } i\}.$$

Duffie and Kan (1993, Section 4) show that for z to remain in D, and thus for volatilities to remain positive, the process must satisfy

Condition A

(a) For each i and all z satisfying $v_i(z) = 0$ (the boundary of positive volatility), the drift is sufficiently positive: $\beta_i'(I - \Phi)(\delta - z) > \beta_i'\Sigma\Sigma'\beta_i/2$.

(b) For all i, and $j \neq i$, if the jth component of $\beta_i'\Sigma$ is nonzero, then $v_i(z) = v_j(z)$.

Given the state variables z, asset prices are generated by a pricing kernel of the form

$$-\log m_{t+1} = \mu + \theta' z_t + \lambda' V(z_t)^{1/2}\epsilon_{t+1}. \tag{23}$$

We consider a similar structure for the foreign kernel, and define a symmetric model as one in which $\log m^*$ has the same form and parameter values as $\log m$, but depends on a vector z^*. To the extent they depend on the same state variables, let the appropriate elements of z and z^* be equivalent.

We turn now to the forward premium anomaly. In this setting, expected depreciation is

$$q_t = \theta'(z_t - z_t^*),$$

the difference between the means of $\log m_{t+1}^*$ and $\log m_{t+1}$. Define $\tau = \sum_j \lambda_j^2 \alpha_j$ and $\gamma = \sum_j \lambda_j^2 \beta_j/2$. Then the short rate is $r_t = (\mu - \tau) + (\theta - \gamma)' z_t$ and, with a similar expression for the foreign short rate, the forward premium is

$$f_t - s_t = (\theta - \gamma)'(z_t - z_t^*).$$

The forward premium anomaly then implies

$$0 > Cov(q_t, f_t - s_t) = (\theta - \gamma)' Var(z - z^*)\theta. \tag{24}$$

The question is whether this is consistent with interest rates that are always positive.

Proposition 2 says that (24) cannot hold in a symmetric model with positive interest rates. We start by reviewing the set D of feasible states. Part (b) of Condition A says, essentially, that either (i) the ith element of $\beta_i'\Sigma$ is its only nonzero element or, if element $j \neq i$ is nonzero, (ii) $v_i(z) = v_j(z)$. Either way, there are only n nonzero elements of $[\beta_1, \ldots, \beta_n]'\Sigma$. In this case, we can choose the β_i's to be positive, since any negative values can be reversed by redefining z as $-z$ and adjusting other parameters accordingly. The feasible space D then takes the following form. In case (i) $v_i(z) > 0$ defines points above a hyperplane (in two dimensions, a line) parallel to the z_i axis. In case (ii), we define a half-space that intersects the "upper-right" part of the positive orthant (in two dimensions, points above a downward-sloping line). Either way, if we start at any point $z \in D$, we can increase any one of the elements of z without bound, holding fixed the other elements, without leaving D.

With this taken care of, we can return to the proposition. Since the β_i's are positive, we see from its definition that γ must be positive, too. For interest rates to be positive we need

$$r_t = (\mu - \tau) + (\theta - \gamma)' z_t > 0 \text{ for all } z \in D.$$

We've seen that this includes z's that increase without bound in one dimension at a time, so $\theta - \gamma$ must be strictly positive. Since γ is positive, this requires strictly positive θ. But if θ and $\theta - \gamma$ are both positive, and $Var(z - z^*)$ is positive semidefinite, inequality (24) is reversed. That is: positive interest rates and the forward premium anomaly are incompatible in this symmetric affine environment.

Figure 1
Cross Section Forward Premium Relation

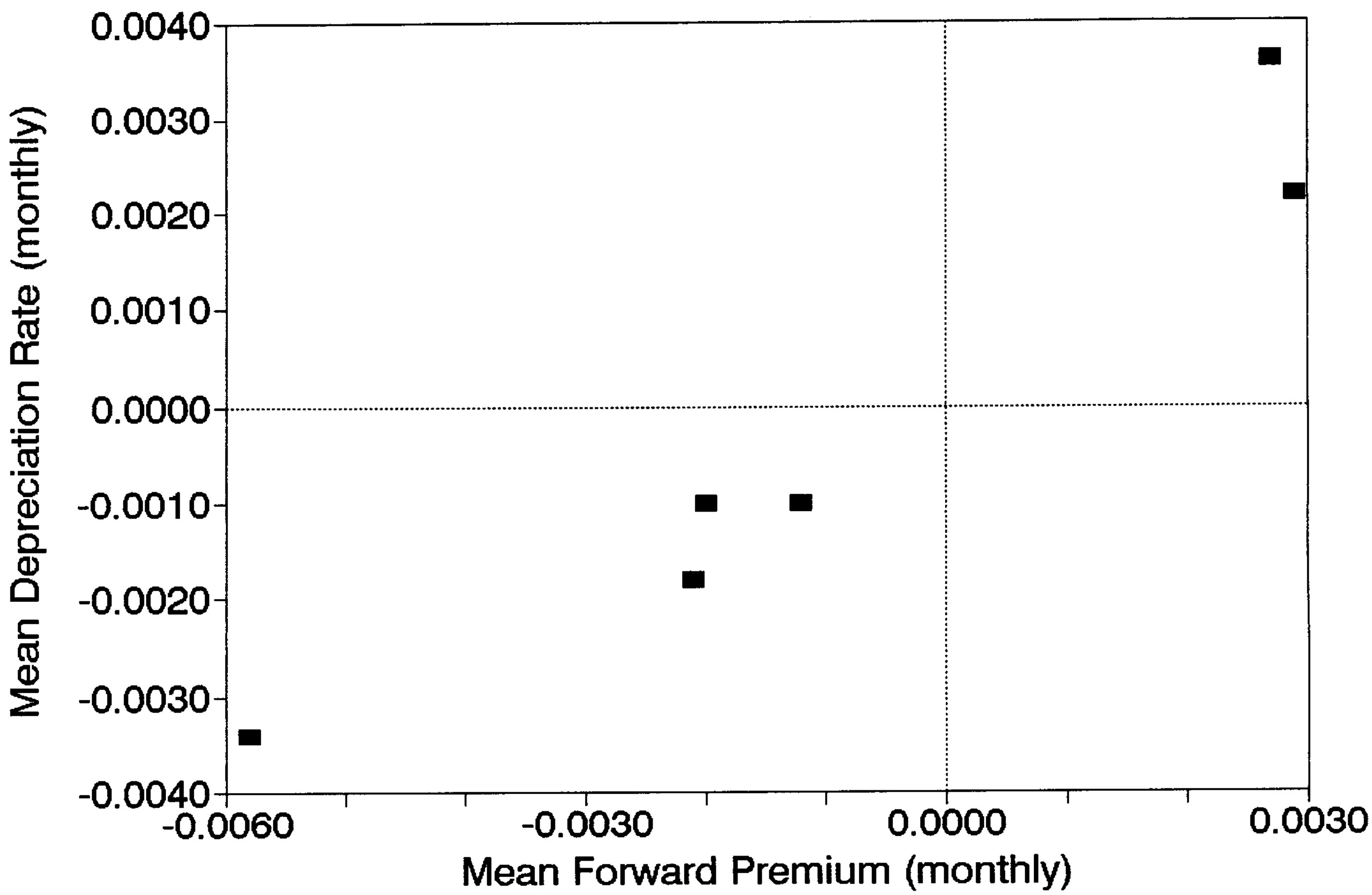

Table 1

Summary Statistics for Exchange Rates

Entries are sample moments of the depreciation rate, $s_{t+1} - s_t$, and the forward premium, $f_t - s_t$, where s and f are logarithms of spot and one-month forward exchange rates, respectively, measured in dollars per unit of foreign currency. Mean is the sample mean, St Dev the sample standard deviation, Skewness an estimate of the skewness parameter γ_1, Kurtosis an estimate of the kurtosis parameter γ_2, and Autocorr the first autocorrelation. Both γ_1 and γ_2 are zero for normal random variables, so nonzero values indicate departures from normality. The data are monthly, last Friday of the month, from the Harris Bank's *Weekly Review: International Money Markets and Foreign Exchange*, compiled by Richard Levich at New York University's Stern School of Business. Dates t run from July 1974 to April 1990 (190 observations).

Currency	Mean	Std Dev	Skewness	Kurtosis	Autocorr
		A. Depreciation Rate, $s_{t+1} - s_t$			
British Pound	-0.0018	0.0328	0.483	1.106	0.046
Canadian Dollar	-0.0010	0.0122	0.012	0.727	0.046
French Franc	-0.0010	0.0326	0.032	0.769	-0.042
German Mark	0.0022	0.0338	0.033	0.265	-0.054
Italian Lira	-0.0034	0.0309	0.033	1.027	-0.004
Japanese Yen	0.0036	0.0338	0.033	0.602	0.074
		B. Forward Premium, $f_t - s_t$			
British Pound	-0.0021	0.0028	-0.265	1.839	0.884
Canadian Dollar	-0.0012	0.0014	-0.014	0.570	0.813
French Franc	-0.0020	0.0033	-0.865	2.834	0.625
German Mark	0.0029	0.0018	0.606	0.667	0.838
Italian Lira	-0.0058	0.0049	-2.047	5.575	0.726
Japanese Yen	0.0027	0.0030	-0.081	0.465	0.866

Table 2

Forward Premium Regressions

The table reports statistics from regressions of the depreciation rate, $s_{t+1}-s_t$, on the forward premium, $f_t - s_t$:

$$s_{t+1} - s_t = a_1 + a_2(f_t - s_t) + \text{residual},$$

where s and f are logarithms of spot and forward exchange rates, respectively, measured as dollars per unit of foreign currency. The data are described in the notes to Table 1. Dates t run from July 1974 to April 1990. Numbers in parentheses are Newey-West standard errors and Std Er is the estimated standard deviation of the residual.

Currency	a_1	a_2	Std Er	R^2
British Pound	-0.0067	-2.306	0.0322	0.0344
	(0.0028)	(0.862)		
Canadian Dollar	-0.0027	-1.464	0.0120	0.0247
	(0.0009)	(0.581)		
French Franc	-0.0026	-0.806	0.0326	0.0015
	(0.0032)	(0.928)		
German Mark	0.0032	-3.542	0.0333	0.0287
	(0.0043)	(1.348)		
Italian Lira	-0.0061	-0.461	0.0309	0.0053
	(0.0044)	(0.403)		
Japanese Yen	0.0084	-1.813	0.0334	0.0201
	(0.0032)	(0.719)		

Table 3

Fama's Decomposition of the Forward Premium

Fama (1984) decomposes the forward premium into a risk premium p and the expected rate of depreciation q:

$$\begin{aligned} f_t - s_t &= (f_t - E_t s_{t+1}) + (E_t s_{t+1} - s_t) \\ &= p_t + q_t. \end{aligned}$$

Table entries are means (Mean) and standard deviations (Std Dev) of the two components, computed from fitted values of the forward premium regression of Table 2 and its complement,

$$s_{t+1} - f_t = -a_1 + (1 - a_2)(f_t - s_t) - \text{residual}.$$

Currency	Risk Premium p Mean	Risk Premium p Std Dev	Exp. Depreciation q Mean	Exp. Depreciation q Std Dev
British Pound	-0.0003	0.0094	-0.0018	0.0065
Canadian Dollar	-0.0002	0.0035	-0.0010	0.0021
French Franc	-0.0010	0.0060	-0.0010	0.0029
German Mark	0.0007	0.0080	0.0022	0.0062
Italian Lira	-0.0023	0.0071	-0.0034	0.0023
Japanese Yen	-0.0009	0.0083	0.0036	0.0054