

Affine Term Structure Models and the Forward Premium Anomaly

DAVID K. BACKUS, SILVERIO FORESI and CHRIS I. TELMER*

ABSTRACT

One of the most puzzling features of currency prices is the *forward premium anomaly*: the tendency for high interest rate currencies to appreciate. We characterize the anomaly in the context of affine models of the term structure of interest rates. In affine models, the anomaly requires either that state variables have asymmetric effects on state prices in different currencies or that nominal interest rates take on negative values with positive probability. We find the quantitative properties of either alternative to have important shortcomings.

PERHAPS THE MOST PUZZLING FEATURE of currency prices is the tendency for high interest rate currencies to appreciate when one might guess, instead, that investors would demand higher interest rates on 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 attributes the behavior of forward and spot exchange rates to a time-varying risk premium. Fama shows that the implied risk premium on a currency must (1) be negatively correlated with its expected rate of depreciation and (2) have 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. Attempts include 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 (Hansen and Hodrick (1983), Domowitz and Hakkio (1985), Cumby (1988)), and consumption-based asset pricing theories, including departures from time-

* Backus is with the Stern School of Business, New York University and the National Bureau of Economic Research; Foresi is with Goldman Sachs. Telmer is with the Graduate School of Industrial Administration, Carnegie Mellon University. In addition to numerous participants at seminars and conferences, we thank Ravi Bansal, Geert Bekaert, Wayne Ferson, Burton Hollifield, Andrew Karolyi, Kenneth Singleton, Amir Yaron, Stanley Zin, and especially the editor, René Stulz, and two referees of this journal for helpful comments and suggestions. Earlier versions of this paper circulated as “The forward premium anomaly: Three examples in search of a solution” (presented at the NBER Summer Institute, July 1994) and “Affine models of currency pricing” (NBER Working Paper 5623). Backus thanks the National Science Foundation for financial support.

additive preferences (Backus, Gregory, and Telmer (1993), Bansal et al. (1995), and Bekaert (1996)), from expected utility (Bekaert, Hodrick, and Marshall (1997)), and from frictionless trade in goods (Hollifield and Uppal (1997)).

Against this backdrop, we ask whether popular models of the term structure of interest rates are consistent with the anomaly, once the models are adapted to a multicurrency setting. Term structure models across currencies imply a term structure of forward exchange rates. The main empirical regularity associated with forward rates is, arguably, the forward premium anomaly. An important question, therefore, asks whether standard term structure models—models which have proven useful for understanding yields in a single currency—imply forward rates that are consistent with the anomaly. Papers that ask a similar question include Amin and Jarrow (1991), Nielsen and Saá-Requejo (1993), Saá-Requejo (1994), Frachot (1996), Ahn (1997), Bakshi and Chen (1997), and Bansal (1997).

Our approach considers a generalization of the models used in related work, an adaptation to currencies of Duffie and Kan's (1996) class of affine yield models. We formulate our models as discrete time processes for currency-specific pricing kernels—essentially, processes for prices of state-contingent claims—and translate Fama's (1984) conditions for risk premiums into restrictions on pricing kernels. We show that these restrictions have strong implications for the structure and parameter values of affine models, and then consider several specific examples. We find, based on both theory and estimation, that affine models have difficulty accounting for the anomaly. Such models must either allow for some probability of negative interest rates or for asymmetric effects of state prices on interest rates in different currencies. Our estimates suggest important quantitative drawbacks associated with either alternative.

We proceed as follows. In Section I we briefly review the properties of currency and interest rate data that our quantitative analysis requires. We then go on to describe the relations among pricing kernels, currency prices, and interest rates that are dictated by general arbitrage-free environments and our specific class of affine models. In Section II, we develop and estimate several affine models and show how different features and parameter values bear on the forward premium anomaly. Section III offers conclusions and suggestions for further research.

I. Theory

We use the notation r_t to denote the continuously compounded one-month eurocurrency interest rate, s_t to denote the logarithm of the spot exchange rate (expressed as U.S. dollars per unit of foreign currency), and f_t to denote the logarithm of the one-month forward exchange rate. Our quantitative assessment of affine yield models requires estimates of a number of moments of these variables. Accordingly, Table I reports selected sample moments as well as the well-known regression results that underlie the forward premium anomaly. The main implications are well known and have proven

Table I
Properties of Currency Prices and Interest Rates

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. The sample is July 1974 to November 1994 (245 observations). An asterisk (*) indicates a sample moment at least twice its Newey-West standard error. The letters s and f denote logarithms of spot and one-month forward exchange rates, respectively, measured in dollars per unit of foreign currency, and r denotes the continuously compounded one-month yield (not an annual percentage). Mean is the sample mean, Std Deviation the sample standard deviation, and Autocorr the first autocorrelation. Forward premium regressions are of the form

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

Numbers in parentheses are Newey-West standard errors (3 lags) and Std Er is the estimated standard deviation of the residual.

Panel A: Summary Statistics				
Currency	Mean	Std Deviation	Autocorrelation	
I. Depreciation rate, $s_{t+1} - s_t$				
British pound	-0.0017	0.0342*	0.084	
German mark	0.0021	0.0340*	-0.015	
Japanese yen	0.0044	0.0324*	0.067	
II. One-month interest rate, r_t				
American dollar	0.0069*	0.0030*	0.957*	
British pound	0.0093*	0.0027*	0.915*	
German mark	0.0053*	0.0020*	0.969*	
Japanese yen	0.0046*	0.0020*	0.914*	
III. Forward premium, $f_t - s_t = r_t - r_t^*$				
British pound	-0.0024*	0.0027*	0.900*	
German mark	0.0017*	0.0029*	0.953*	
Japanese yen	0.0021*	0.0029*	0.888*	
Panel B: Forward Premium Regressions				
Currency	a_1	a_2	Std Er	R^2
British pound	-0.0062 (0.0027)	-1.840 (0.847)	0.0339	0.0213
German mark	0.0033 (0.0025)	-0.743 (0.805)	0.0340	0.0041
Japanese yen	0.0080 (0.0024)	-1.711 (0.643)	0.0320	0.0230

robust to a number of challenges (see, e.g., surveys by Hodrick (1987), Canova and Marrinan (1995), Engel (1996), and Flood and Rose (1996)). Depreciation rates have means that are indistinguishable from zero, are highly volatile, and display little, if any, autocorrelation. In contrast, interest rates and interest rate differentials are less volatile—both in absolute terms and relative to their means—and are highly autocorrelated. Finally, and most

importantly for our purposes, forward premiums are negatively correlated with subsequent depreciation rates. In Table I, this is manifest in the estimates of the slope coefficients from the forward premium regressions, which are -1.84 , -0.74 and -1.71 for the pound, the deutschemark, and the yen, respectively. The magnitude of these estimates will prove instrumental in our assessment of the currency pricing implications of affine yield models.

To understand how these moments restrict theory, it is useful to begin by following Fama (1984) in decomposing the forward premium, $f_t - s_t$, into the expected rate of depreciation on the dollar, q_t , and the expected excess return, p_t , on taking a short forward position in the nondollar currency,

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

The variable $p_t = f_t - E_t s_{t+1}$ is often interpreted as a “risk premium,” language that we find useful and will continue to use. Note, however, that our approach neither requires that this interpretation be valid nor imposes sufficient structure so as to make it meaningful.

The cross-sectional evidence (Table I) suggests that risk premiums p are small on average, but the time-series evidence implies they are highly variable. Because the population regression coefficients, a_2 from Table I, are

$$a_2 = \frac{\text{cov}(q, p + q)}{\text{var}(p + q)} = \frac{\text{cov}(q, p) + \text{var}(q)}{\text{var}(p + q)}, \tag{2}$$

it is clear that a constant risk premium p generates $a_2 = 1$. To generate a negative value of a_2 , we need $\text{cov}(q, p) + \text{var}(q) < 0$. Fama (1984) translates this into two necessary conditions, each of which will impose strong restrictions on our theory.

Fama’s Conditions. A negative regression coefficient, $a_2 < 0$, requires

1. Negative covariance between p and q , and
2. Greater variance of p than q .

Much of what we do amounts to translating these conditions into restrictions on pricing kernels and assessing the extent to which affine pricing kernels satisfy the restrictions.

A. Pricing Kernels and Currency Prices

Prior to imposing the structure represented by affine models, it is useful to consider currency pricing in a more general context. We describe models in terms of *pricing kernels*: stochastic processes governing the prices of state-contingent claims. The existence of a pricing kernel (or, equivalently, of risk-neutral probabilities) is guaranteed in any economic environment that precludes arbitrage opportunities, including those that admit “irrational”

beliefs, differential information, and so on. This generality, of course, comes at the cost of a minimalist set of economic restrictions that limits what one can ultimately say about the fundamentals driving the pricing kernel.

A pricing kernel is essentially an intertemporal price. It represents the probability-weighted cost of receiving a state-contingent payoff sometime in the future. The price of currency, on the other hand, is a contemporaneous price (ignoring some minor institutional details). What drives our theory is the way in which these two prices are related; the ratio of the pricing kernel denominated in, say, pounds to the kernel denominated in dollars equals the depreciation rate of the dollar vis-à-vis the pound. If markets are complete, these pricing kernels are unique and so is the ratio of the two. Otherwise, as we show below, we can choose from the set of admissible kernels so that the relationship remains satisfied.

Consider assets denominated in either domestic currency (“dollars”) or foreign currency (“pounds”). The dollar value v_t of a claim to the stochastic cash flow of d_{t+1} dollars one period later satisfies

$$v_t = E_t(m_{t+1}d_{t+1}), \quad (3)$$

or

$$1 = E_t(m_{t+1}R_{t+1}), \quad (4)$$

where $R_{t+1} = d_{t+1}/v_t$ is the gross one-period return on the asset. We refer to m as the dollar pricing kernel. In a more fully articulated general equilibrium model, m might be an agent’s nominal intertemporal marginal rate of substitution and equation (4) a first-order condition. More generally, m is a positive random variable that satisfies the pricing relation (4) for returns R on all traded assets, and whose existence is both necessary and sufficient for an economy that does not admit riskless arbitrage opportunities. If such an economy has complete markets for state-contingent claims, m is the unique solution to equation (4). Otherwise there exist a large number of random variables m that satisfy the pricing relation for returns on all traded assets. These issues, and the relevant literature, are reviewed by Duffie (1992).

A pricing kernel m and the associated pricing relation (4) are the basis of modern theories of bond pricing: given a pricing kernel, we use equation (4) to compute prices and yields for bonds of all maturities. The price of a one-period bond, for example, is $b_t^1 = E_t m_{t+1}$, and the (one-period) short rate r_t is

$$r_t = -\log b_t^1 = -\log E_t m_{t+1}. \quad (5)$$

We return to this equation shortly.

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

$$1 = E_t(m_{t+1}^* R_{t+1}^*). \quad (6)$$

If we use the second approach, with $S = \exp(s)$ denoting the dollar spot price of one pound, dollar returns on this asset are $R_{t+1} = (S_{t+1}/S_t)R_{t+1}^*$ and

$$1 = E_t[m_{t+1}(S_{t+1}/S_t)R_{t+1}^*].$$

If the pound asset and currencies are both traded, the return must satisfy both conditions:

$$E_t(m_{t+1}^* R_{t+1}^*) = E_t[m_{t+1}(S_{t+1}/S_t)R_{t+1}^*].$$

This equality ties the rate of depreciation of the dollar to the random variables m and m^* that govern state prices in dollars and pounds. 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 equations (4) and (6) are not unique, but we can choose them to satisfy the same equation.

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 pound denominated assets. If these processes do not admit arbitrage opportunities, then there exist pricing kernels m and m^* for dollars and pounds that satisfy both*

$$m_{t+1}^*/m_{t+1} = S_{t+1}/S_t \quad (7)$$

and the pricing relations (4) and (6).

Proof: Consider dollar returns on the complete set of traded assets, including the dollar returns $(S_{t+1}/S_t)R_{t+1}^*$ on pound-denominated assets. If these returns do not admit arbitrage opportunities, then there exists a positive random variable m_{t+1} satisfying equation (4) 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 equation (6). Q.E.D.

The proposition tells us that of the three random variables— m_{t+1} , m_{t+1}^* , and S_{t+1}/S_t —one is effectively redundant and can be constructed from the other two. Most of the existing literature uses the domestic pricing kernel m

(or its equivalent expressed as state prices or risk-neutral probabilities) and the depreciation rate. We start instead with the two pricing kernels, which highlights the essential symmetry of the theory between the two currencies.

The essence of the proposition is that, given processes for S_{t+1}/S_t and m , we can always use equation (7) to choose a process m^* that satisfies equation (6). Unless markets are complete, however, this choice will not be unique. Consider, for instance, the process $\hat{m}_{t+1}^* = m_{t+1}^* \exp(\eta_{t+1})$, where m^* satisfies both equations (6) and (7) and η is some random variable. Complete markets imposes the restriction that, if \hat{m}^* is to satisfy equation (6), then $\eta = 0$. No arbitrage imposes the weaker set of restrictions that, for returns R and R^* on all traded portfolios,

$$E_t(m_{t+1}^* R_{t+1}^*) = E_t[m_{t+1}(S_{t+1}/S_t) R_{t+1}^*] = E_t(m_{t+1}^* e^{\eta_{t+1}} R_{t+1}^*) \quad (8)$$

$$E_t(m_{t+1} R_{t+1}) = E_t[m_{t+1}^*(S_t/S_{t+1}) R_{t+1}] = E_t(m_{t+1} e^{-\eta_{t+1}} R_{t+1}).$$

These conditions restrict η in meaningful ways. Nevertheless, the set of admissible deviations from m^* is potentially large. We show below, however, that once our attention is restricted to affine models, all elements of this set are observationally equivalent.

B. Forward Rates and Risk Premiums

We can now relate the risk premium defined by Fama to properties of the two pricing kernels. Consider a forward contract specifying at date t the exchange at $t + 1$ of one pound and $F_t = \exp(f_t)$ dollars, with the forward rate F_t set at date t as the notation suggests. This contract specifies a net dollar cash flow at date $t + 1$ of $F_t - S_{t+1}$. Because it involves no payments at date t , pricing relation (3) implies

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

Dividing by S_t and applying Proposition 1, we find

$$(F_t/S_t)E_t(m_{t+1}) = E_t(m_{t+1}S_{t+1}/S_t) = E_t(m_{t+1}^*).$$

Thus the forward premium is

$$f_t - s_t = \log E_t m_{t+1}^* - \log E_t m_{t+1}. \quad (9)$$

This equation and definitions of short rates (equation (5) for the domestic rate and an analogous relation for the foreign rate) give us $f_t - s_t = r_t - r_t^*$, the familiar covered interest rate parity condition.

Now consider the components of the forward premium. The expected rate of depreciation is, from equation (7),

$$q_t \equiv E_t s_{t+1} - s_t = E_t \log m_{t+1}^* - E_t \log m_{t+1}. \quad (10)$$

Thus we see that the first of Fama's components is the difference in conditional means of the logarithms of the pricing kernels. The risk premium is, from equations (1) and (9),

$$p_t = (\log E_t m_{t+1}^* - E_t \log m_{t+1}^*) - (\log E_t m_{t+1} - E_t \log m_{t+1}), \quad (11)$$

the difference between the "log of the expectation" and the "expectation of the log" of the pricing kernels m and m^* .

These derivations shed light on how the joint distribution of the pricing kernels affects interest rates and currency prices. Equation (10) tells us that only the first moments of the log pricing kernel affect the expected rate of depreciation, q . The effects of higher moments, according to equation (10), are manifest in p , the risk premium. To see this more clearly, we can expand the variable $\log E_t m_{t+1}$ in terms of the cumulants, κ_{jt} , of the conditional distribution of $\log m_{t+1}$:

$$\log E_t m_{t+1} = \sum_{j=1}^{\infty} \kappa_{jt} / j!. \quad (12)$$

Equation (12) is an expansion of the cumulant generating function (the logarithm of the moment generating function) evaluated at one; see, for example, Stuart and Ord (1987). Cumulants are closely related to moments, as we 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_{1t} denoting the conditional mean of $\log m_{t+1}$ and μ_{jt} , for $j > 1$, denoting the j th central conditional moment.

Given a representation for the foreign kernel analogous to equation (12), we can express the forward premium and its components in terms of cumulants:

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

Equation (10) tells us that the first term in this summation is the expected depreciation rate, q . The risk premium, p , is therefore,

$$p_t = \kappa_{-1,t}^* - \kappa_{-1,t}, \quad (13)$$

where

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

We refer generically to the sums $\kappa_{-1,t}$ and $\kappa_{-1,t}^*$ as “higher-order cumulants.”

We can now translate Fama’s necessary conditions on p and q into restrictions on pricing kernels.

Remark 1: 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 (1) negative correlation between differences in conditional means, $\mu_{1t}^* - \mu_{1t}$, and differences in higher-order cumulants, $\kappa_{-1,t}^* - \kappa_{-1,t}$, and (2) greater variation in the latter. A necessary and sufficient condition is a negative covariance between $q_t = \mu_{1t}^* - \mu_{1t}$ and $f_t - s_t = \mu_{1t}^* - \mu_{1t} + \kappa_{-1,t}^* - \kappa_{-1,t}$.

With additional structure we can be more specific. Many popular models of bond and currency prices, including the affine models we examine shortly, start with conditionally log-normal pricing kernels: $\log m_{t+1}$ and $\log m_{t+1}^*$ are conditionally normal with (say) means (μ_{1t}, μ_{1t}^*) and variances (μ_{2t}, μ_{2t}^*) . One-period bond prices become

$$E_t m_{t+1} = \exp(\mu_{1t} + \mu_{2t}/2)$$

$$E_t m_{t+1}^* = \exp(\mu_{1t}^* + \mu_{2t}^*/2)$$

and the expected rate of depreciation is

$$q_t = \mu_{1t}^* - \mu_{1t}.$$

Because cumulants above the first two are zero for the normal distribution, the risk premium is simply the difference between the conditional variances,

$$p_t = (\mu_{2t}^* - \mu_{2t})/2. \quad (14)$$

In this case, Fama’s conditions amount to the requirement of (1) negative correlation between differences in conditional means and conditional variances of the two pricing kernels and (2) greater variation in one-half the difference in the conditional variances. In a nutshell, what is necessary is a great deal of variation in conditional variances.

This characterization of the risk premium suggests an interpretation for the failure of GARCH-M models, which model the risk premium as a function of the conditional variance of the depreciation rate. Studies by Domowitz and Hakkio (1985), Bekaert and Hodrick (1993), and Bekaert (1995)

document strong evidence of time-varying conditional variances of depreciation rates, but little that connects the conditional variance to the risk premium p . One view of this failure is that GARCH-M models violate our sense of symmetry: An increase in the conditional variance of the depreciation rate increases risk on both sides of the market, and hence carries no presumption in favor of one currency or the other. Our framework indicates why. The conditional variance of the depreciation rate is

$$\text{var}_t(s_{t+1} - s_t) = \text{var}_t(\log m_{t+1}^* - \log m_t),$$

the conditional variance of the difference between the logarithms of the two kernels. The risk premium, on the other hand, is half the difference in the conditional variances (equation (14)) and possibly higher moments (equation (13)), which need not bear any specific relation to the conditional variance of the depreciation rate. GARCH-M models, to put it simply, focus on a different conditional variance.

C. The General Affine Model

Further progress requires more structure. We explore an adaptation to currencies of the Duffie and Kan (1996) class of affine yield models. The linearity of these models makes it relatively easy to explore the implications of different structures and parameter values for the forward premium anomaly.

This class of affine currency models starts with a vector z of state variables following the law of motion

$$z_{t+1} = (I - \Phi)\theta + \Phi z_t + V(z_t)^{1/2} \varepsilon_{t+1} \quad (15)$$

where $\{\varepsilon_t\} \sim \text{NID}(0, I)$, Φ is stable with positive diagonal elements, and V is diagonal with typical element

$$v_i(z) = \alpha_i + \beta_i^\top z.$$

Further conditions on the parameters are required to guarantee that the state z never leaves the region defined by nonnegative values of the volatility functions v_i ; see Appendix A or Duffie and Kan (1996) for details.

Given z , pricing kernels are

$$-\log m_{t+1} = \delta + \gamma^\top z_t + \lambda^\top V(z_t)^{1/2} \varepsilon_{t+1} \quad (16)$$

$$-\log m_{t+1}^* = \delta^* + \gamma^{*\top} z_t + \lambda^{*\top} V(z_t)^{1/2} \varepsilon_{t+1}, \quad (17)$$

which imply short rates

$$\begin{aligned} r_t &= (\delta - \omega) + (\gamma - \tau)z_t \\ r_t^* &= (\delta^* - \omega^*) + (\gamma^* - \tau^*)z_t, \end{aligned}$$

where $\omega = \sum_j \lambda_j^2 \alpha_j / 2$, $\omega^* = \sum_j \lambda_j^{*2} \alpha_j / 2$, $\tau = \sum_j \lambda_j^2 \beta_j / 2 \geq 0$, and $\tau^* = \sum_j \lambda_j^{*2} \beta_j / 2 \geq 0$. The depreciation rate is

$$s_{t+1} - s_t = (\delta - \delta^*) + (\gamma - \gamma^*)^\top z_t + (\lambda - \lambda^*)^\top V(z_t)^{1/2} \varepsilon_{t+1}. \quad (18)$$

Thus, log kernels and depreciation rates are conditionally normal with conditional means and variances that are linear in the state z .

These two equations imply a negative slope coefficient for the forward premium regression coefficients a_2 (see Table I) if

$$\text{cov}(s_{t+1} - s_t, f_t - s_t) = [(\gamma - \gamma^*) - (\tau - \tau^*)]^\top \text{var}(z)(\gamma - \gamma^*) < 0.$$

We see immediately that the anomaly hinges on some sort of asymmetry between the distributions of the two kernels, a feature that will be a consistent theme of the more tightly parameterized models of Section II.

The general affine structure requires parameter restrictions to say much more, something we turn to in the next section. One exception, however, is Proposition 1, for which the complete/incomplete markets distinction is no longer empirically relevant. The presumption of the affine model is that predictable and unpredictable movements in log bond prices and depreciation rates are spanned, respectively, by the state z and the innovations $V(z)^{1/2} \varepsilon$. Hence, any error in equation (7) is affine and Proposition 1 might be satisfied by an alternative affine choice of foreign pricing kernel,

$$-\log \hat{m}_{t+1}^* = (\delta^* + \delta') + (\gamma^* + \gamma')^\top z_t + (\lambda^* + \lambda')^\top V(z_t)^{1/2} \varepsilon_{t+1},$$

for some arbitrary nonzero choice of $(\delta', \gamma', \lambda')$. Equation (8) implies, however, that m^* and \hat{m}^* price all traded assets the same way. In this sense, the two kernels are observationally equivalent and there is no loss of generality in setting $\delta' = \gamma' = \lambda' = 0$.

II. Accounting for the Anomaly

Remark 1 suggests that it should be relatively easy to construct examples that reproduce the anomaly: We simply arrange for differences in first and second moments of pricing kernels to move in opposite directions. For example, consider a model like Engel and Hamilton's (1990), in which the conditional distributions of two pricing kernels alternate between two log-

normal regimes. If the difference in conditional means of the pricing kernels is higher in regime 1, and one-half the difference in conditional variances is higher in regime 2 and varies more than the difference in means, then the model will reproduce the anomaly.

Accounting for the anomaly in the context of affine term structure models poses a somewhat greater challenge. In addition to a complex set of parameter restrictions imposed by the affine structure, the model must also account for the properties of currency prices and interest rates in a more general sense. With this in mind, this section provides a more in-depth exploration of affine models, beginning with one of the better known examples, the Cox, Ingersoll, and Ross (1985) model. The Cox-Ingersoll-Ross model serves as a useful benchmark in that its essential features—the satisfaction of Fama’s condition (1) but not condition (2)—turn out to apply to a more general subset of models in the affine class. We formalize this in a proposition and then go on to discuss models designed to overcome the Cox-Ingersoll-Ross model’s shortcomings.

A. A Two-Currency Cox-Ingersoll-Ross Model

Our model is adapted from Sun’s (1992) discrete-time translation. We start with two state variables, indexed by $i = 1, 2$, that obey independent “square-root” processes

$$z_{it+1} = (1 - \varphi_i)\theta_i + \varphi_i z_{it} + \sigma_i z_{it}^{1/2} \varepsilon_{it+1}, \quad (19)$$

with $0 < \varphi_i < 1$, $\theta_i > 0$, and $\{\varepsilon_{it}\} \sim \text{NID}(0, 1)$. The unconditional mean of z_i is θ_i , the autocorrelation is φ_i , the conditional variance is $\sigma_i^2 z_{it}$, and the unconditional variance is $\sigma_i^2 \theta_i / (1 - \varphi_i^2)$. A variant of equation (19),

$$z_{it+1} - z_{it} = (1 - \varphi_i)(\theta_i - z_{it}) + \sigma_i z_{it}^{1/2} \varepsilon_{it+1},$$

is a direct analog of the continuous-time original (Cox, Ingersoll, and Ross (1985), equation (17)). A salient feature of equation (19) is the square-root term in the innovation, whose conditional variance falls to zero as z_i approaches zero. In continuous time, this feature guarantees that z_i remains positive. In discrete time, z_i can turn negative with a large enough negative realization of ε_i . This happens with positive probability, but the probability approaches zero as the time interval goes to zero (Sun (1992)). The *Feller condition*,

$$\frac{2(1 - \varphi_i)\theta_i}{\sigma_i^2} \geq 1, \quad (20)$$

controls the shape of the unconditional distribution of z_i .

In the standard one-factor Cox-Ingersoll-Ross model, z_1 (say) is the dollar short rate and the pricing kernel is

$$-\log m_{t+1} = (1 + \lambda_1^2/2)z_{1t} + \lambda_1 z_{1t}^{1/2} \varepsilon_{1t+1}, \quad (21)$$

a special case of equation (17). The coefficient of z_1 makes it the one-period rate of interest. The parameter λ_1 controls the covariance of the kernel with movements in interest rates and thus governs the risk of long bonds and the average slope of the yield curve.

This structure is an example of the conditionally log-normal pricing kernels described in Section I. Moreover, equation (21) builds in an inverse relation between the conditional mean and variance of the logarithm of the pricing kernel, as required by Fama's condition (1). The conditional mean and variance,

$$E_t \log m_{t+1} = -(1 + \lambda_1^2/2)z_{1t}$$

$$\text{var}_t \log m_{t+1} = \lambda_1^2 z_{1t},$$

are both linear in the state variable z_1 . The short rate is therefore

$$r_t = -\log E_t m_{t+1} = -(E_t \log m_{t+1} + \tfrac{1}{2} \text{var}_t \log m_{t+1}) = z_{1t}, \quad (22)$$

as claimed earlier.

A natural extension to two currencies is to posit an analogous pricing kernel for valuing foreign-currency cash flows. The pricing kernel m^* for (say) pounds is based on a second state variable z_2 and follows

$$-\log m_{t+1}^* = (1 + \lambda_2^2/2)z_{2t} + \lambda_2 z_{2t}^{1/2} \varepsilon_{2t+1}. \quad (23)$$

Then the pound short rate is $r_t^* = z_{2t}$ and the forward premium is

$$f_t - s_t = z_{1t} - z_{2t}.$$

If we impose the symmetry restriction $\lambda_1 = \lambda_2$, we can write the expected depreciation rate as $q_t = (1 + \lambda_1^2/2)(z_{1t} - z_{2t})$ and the risk premium as $p_t = -(\lambda_1^2/2)(z_{1t} - z_{2t})$. Thus the linearity of the conditional mean and variance translate into forward premium components that are linear functions of the differential $z_1 - z_2$. More important, this structure automatically generates the negative correlation between p and q of Fama's condition (1): Because equation (21) implies an inverse relation between the conditional mean and variance of $\log m_{t+1}$, and the two pricing kernels are independent, the difference in conditional means is inversely related to the difference in conditional variances.

This model cannot, however, satisfy Fama's condition (2) or reproduce the anomalous regression slope. If we regress the depreciation rate on the forward premium in this model, the slope is

$$\alpha_2 = 1 + \lambda_1^2/2.$$

The slope is not only positive, and therefore inconsistent with the anomaly, it exceeds one.

B. Independent Factor Models

The shortcomings of our Cox-Ingersoll-Ross example turn out to apply to a wider class of affine models. Specifically, if we restrict ourselves to models in which foreign and domestic interest rates depend on independent factors, so that the only connection between them is a common factor that affects both rates the same way, then we cannot simultaneously account for the anomaly and maintain strictly positive interest rates.

A more precise statement is as follows. Consider the general class of affine models outlined in Section I.C. Partition the state vector z into three independent components: z_0 , z_1 , and z_2 . The vector z_0 denotes the set of common factors that affect interest rates in both currencies in the same way, whereas z_1 and z_2 denote vectors of currency-specific factors. If we express the components of parameter vectors with similar subscripts, then what we mean by an "independent factor structure" amounts to the following restrictions on the parameters of Section I.C:

- Φ is block-diagonal with elements Φ_i for $i = 0, 1, 2$.
- The common factor affects both pricing kernels the same way: $\gamma_0 = \gamma_0^*$ and $\lambda_0 = \lambda_0^*$.
- The currency-specific factors affect the pricing kernel of only one currency: $\gamma_2 = \lambda_2 = 0$ and $\gamma_1^* = \lambda_1^* = 0$.

Ahn (1998) and Amin and Jarrow (1991) consider similar structures.

This independent factor structure exhibits the same tension between the forward premium anomaly and positive interest rates as the Cox-Ingersoll-Ross example:

PROPOSITION 2: *Consider the affine class of currency models described by equations (15)–(18) with the independent factor structure described above. If such a model implies positive bond yields for all admissible values of the state variables, then it cannot generate a negative value of the slope parameter α_2 from forward premium regressions.*

A proof is given in Appendix A, but the intuition is straightforward. The affine models permitted in Proposition 2 are based on state variables that are unbounded from above. Such state variables have two effects on the short rate, one operating through the mean of the pricing kernel, the other

through the variance. An increase in the conditional mean tends to raise the short rate, whereas an increase in the conditional variance lowers it. If the mean effect is larger, as it is in the Cox-Ingersoll-Ross model, then the short rate is unbounded above. The anomaly requires instead that the effect of the variance must be larger, and thus that increases in variance be associated with decreases in the short rate. But because the conditional variance is unbounded above, the short rate will be negative for large enough values of the state variable.

Proposition 2 implies that if we are insistent about ruling out negative interest rates, then we must abandon independent factors, something we turn to shortly. Alternatively, it suggests that by allowing for some probability that interest rates are negative, we might overcome the roadblock encountered by the Cox-Ingersoll-Ross model. The following example is developed with this in mind, leaving the quantitative issue of the probability of negative interest rates to the latter part of this section.

Consider, then, the same two-currency world as above, but where z_0 , z_1 , and z_2 are scalar-valued variables, each following the law of motion in equation (19). Pricing kernels in the domestic and foreign currency are

$$\begin{aligned} -\log m_{t+1} &= (1 + \lambda_0^2/2)z_{0t} + (-1 + \lambda_1^2/2)z_{1t} + \lambda_0 z_{0t}^{1/2} \varepsilon_{0t+1} + \lambda_1 z_{1t}^{1/2} \varepsilon_{1t+1} \\ -\log m_{t+1}^* &= (1 + \lambda_0^2/2)z_{0t} + (-1 + \lambda_2^2/2)z_{2t} + \lambda_0 z_{0t}^{1/2} \varepsilon_{0t+1} + \lambda_2 z_{2t}^{1/2} \varepsilon_{2t+1}, \end{aligned}$$

with $\{\varepsilon_{it}\}$ independent standard normal random variables. The depreciation rate is therefore

$$s_{t+1} - s_t = (-1 + \lambda_1^2/2)(z_{1t} - z_{2t}) + \lambda_1(z_{1t}^{1/2} \varepsilon_{1t+1} - z_{2t}^{1/2} \varepsilon_{2t+1}), \quad (24)$$

where we have again imposed the symmetry condition, $\lambda_1 = \lambda_2$, for analytic convenience.

The common factor, z_0 , allows us to account for a nonzero correlation of interest rates across currencies, something which will be important in our quantitative implementation. However, because it affects both pricing kernels in the same way, it has no effect on currency prices, interest rate differentials, or the slope parameter a_2 that characterizes the forward premium anomaly. The currency specific factors, in contrast, bear directly on the anomaly. Short-term interest rates are $r_t = z_{0t} - z_{1t}$ and $r_t^* = z_{0t} - z_{2t}$, which implies a forward premium of $-(z_{1t} - z_{2t})$. The expected rate of depreciation is $(-1 + \lambda_1^2/2)(z_{1t} - z_{2t})$ and the slope coefficient from the forward premium regression is

$$a_2 = 1 - \lambda_1^2/2, \quad (25)$$

which is always less than one, and negative for large enough values of λ_1 .

We thus have a model that can account for the forward premium anomaly. The cost is that we can no longer maintain strictly positive interest rates. Whether this is a serious difficulty for the model is largely an empirical issue that we address shortly. If a small probability of negative interest rates leads to an affine model that is realistic in other respects, we might regard this a small cost paid for the convenience of linearity. Duffie and Singleton (1997) and Pearson and Sun (1994) make similar arguments in extending the Cox-Ingersoll-Ross model of bond pricing to more general affine structures.

C. Interdependent Factor Models

Proposition 2 and the subsequent example indicate that one approach to explaining the anomaly is to abandon the requirement of strictly positive interest rates. Another is to posit interdependence between factors: Domestic and foreign interest rates depend in different ways on the same state variables. We consider a simple two-factor structure, the main reason being its desirable quantitative properties (e.g., it can account for an imperfect correlation between foreign and domestic interest rates). Multiple factors, however, turn out to be unnecessary if the only objective is accounting for the forward premium anomaly. This is demonstrated in both a previous version of our paper and the related, continuous time examples of Ahn (1997) and Frachot (1996).

Consider a model based on two state variables, z_1 and z_2 , obeying identical independent square-root processes equation (19), and pricing kernels

$$-\log m_{t+1} = (1 + \lambda_1^2/2)z_{1t} + (\gamma^* + \lambda_2^2/2)z_{2t} + \lambda_1 z_{1t}^{1/2} \varepsilon_{1t+1} + \lambda_2 z_{2t}^{1/2} \varepsilon_{2t+1}$$

$$-\log m_{t+1}^* = (\gamma^* + \lambda_2^2/2)z_{1t} + (1 + \lambda_1^2/2)z_{2t} + \lambda_2 z_{1t}^{1/2} \varepsilon_{1t+1} + \lambda_1 z_{2t}^{1/2} \varepsilon_{2t+1}.$$

Similar models are considered by Bakshi and Chen (1997), Nielsen and Saá-Requejo (1993) and Saá-Requejo (1994). Our version is symmetric in the sense that the unconditional distributions of the two pricing kernels are the same, but interdependent in the sense that the state variables z_1 and z_2 potentially affect the two kernels in different ways.

Consider the implications for the anomaly. Short rates in this model are

$$r_t = z_{1t} + \gamma^* z_{2t}$$

$$r_t^* = \gamma^* z_{1t} + z_{2t}.$$

The forward premium,

$$f_t - s_t = r_t - r_t^* = (1 - \gamma^*)(z_{1t} - z_{2t}),$$

and depreciation rate,

$$s_{t+1} - s_t = [1 - \gamma^* + (\lambda_1^2 - \lambda_2^2)/2](z_{1t} - z_{2t}) + (\lambda_1 - \lambda_2)(z_{1t}^{1/2} \varepsilon_{1t+1} - z_{2t}^{1/2} \varepsilon_{2t+1}),$$

imply a regression slope of

$$a_2 = 1 + \frac{\lambda_1^2 - \lambda_2^2}{2(1 - \gamma^*)}.$$

Clearly, an appropriate choice of parameter values allows us to generate a negative value. The critical feature, again, is that the state variables affect the two kernels differently.

The form that interdependence takes here is particularly striking. Suppose (without loss of generality) that $\gamma^* < 1$. From the short rate equations, we might say that z_1 is the “dollar factor,” because it has a greater effect on the dollar short rate than z_2 . For similar reasons we might refer to z_2 as the “pound factor.” But the anomaly (in fact, any value of a_2 less than one) implies $\lambda_2^2 > \lambda_1^2$, implying that innovations in the pound factor have greater influence on the dollar kernel than do innovations in the dollar factor. It is as if (to use a concrete example) U.S. money growth had a larger influence than British monetary policy on dollar interest rates, but a smaller influence on the dollar pricing kernel.

D. Estimation

We now have two examples of two-currency, affine term structure models in which interest rates and currency prices can behave in a manner consistent with the forward premium anomaly. The mechanics with which they achieve this, however, are starkly different. One model envisions international interest rates as arising from a common-idiosyncratic-factor structure, with the forward premium being driven by differences in idiosyncratic factors. The other envisions international interest rates as being driven by a set of global factors, with different currencies reacting to the factors in different ways. Discriminating between the models, therefore, seems important both in terms of the forward premium anomaly and our view of what constitutes a useful international factor structure.

Our approach is to estimate the parameters of each model by GMM using one moment condition for each parameter. The spirit is similar to Constantinides’ (1992) study of interest rates, but the formalism of GMM provides us with standard errors for the estimated parameters. In each case, we use data on the dollar–pound exchange rate, the dollar short rate, and the dollar–pound forward premium. All of the relevant sample moments are reported in Table I.

Table II
Estimates of Independent and Interdependent Models

Entries are exactly identified GMM estimates of the parameters of the models outlined in Sections II.B and II.C, respectively. Estimates are based on monthly observations on the dollar–pound exchange rate, July 1974 to November 1994. The independent factor model is estimated to reproduce these moments: the mean, variance, and autocorrelation of the dollar short rate; the variance and autocorrelation of the dollar–pound forward premium; the variance of the dollar–pound depreciation rate; and the regression slope a_2 of Table I. The interdependent model is similar. We drop the condition based on the autocorrelation of the short rate, because the (symmetric) model cannot reproduce different values for the autocorrelation of the short rate and the forward premium. Standard errors are computed by the Newey-West method (12 lags).

Parameter	Estimate	Standard Error
Panel A: Independent Factor Model		
θ_0	0.007	6.567×10^{-4}
σ_0	0.003	0.009
φ_0	0.992	0.041
θ_1	1.004×10^{-4}	3.614×10^{-5}
σ_1	0.081	0.022
φ_1	0.919	0.032
$ \lambda_1 $	2.383	0.384
Panel B: Interdependent Factor Model		
θ_1	0.005	4.950×10^{-4}
σ_1	0.017	0.004
φ_1	0.919	0.037
γ^*	0.331	0.080
λ_1	-5.886	2.184
λ_2	-5.541	2.202

In Table II, we report estimates of parameter values and standard errors for the independent factor model (presented in Section II.B) and the two-factor interdependent model of Section II.C. For the former we choose parameters to reproduce the following moments: the mean, variance, and autocorrelation of the dollar short rate; the variance and autocorrelation of the dollar–pound forward premium; the variance of the dollar–pound depreciation rate; and the regression slope a_2 of Table I that we use to characterize the anomaly.

These estimates allow us to assess the quantitative impact of Proposition 2. Because all three state variables range between zero and infinity, short rates ($r_t = z_{0t} - z_{1t}$ and $r_t^* = z_{0t} - z_{2t}$) are negative with positive probability, as required by the proposition. We note in Appendix B that the state variables have (approximately) gamma distributions. With estimated parameters, we find that the probability of negative r is less than 10^{-5} . The difficulty, instead, is that the model exhibits extreme distributional properties for the currency-specific factors and the forward premium. The problem

can be traced directly to the anomaly. The regression coefficient dictates the large value of λ_1^2 that we see in Table II; see equation (25). But the variance of the depreciation rate dictates a small value of $\lambda_1^2 \theta_1$:

$$\text{var}(s_{t+1} - s_t) \geq \lambda_1^2 \theta_1;$$

see equation (24). Because the monthly standard deviation of the depreciation rate is about 3 percent (see Table I), we must choose an extremely small value for θ_1 . This value, in turn, implies that we violate the Feller condition equation (20) by more than two orders of magnitude:

$$\frac{2(1 - \varphi_1)\theta_1}{\sigma_1^2} = 0.003 \neq 1.$$

The Feller ratio governs the distributional properties of z_1 . It also implies extreme values for its skewness and kurtosis; see Appendix B. None of these features is peculiar to the dollar–pound rate; parameter estimates are similar for the other currencies described in Table I.

Table II also lists estimates for the interdependent factor model of Section II.C. There is one difference in the moments used to estimate parameter values: We drop the autocorrelation of the short rate, because the model cannot reproduce different values for the autocorrelation of the short rate and the forward premium. The resulting estimates do not have the difficulty noted for those of the independent factor example. The Feller ratio,

$$\frac{2(1 - \varphi_1)\theta_1}{\sigma_1^2} = 3.455 > 1,$$

easily satisfies the Feller condition. As a result, the unconditional distributions of interest rates do not exhibit the extreme behavior we noted for the case of independent factors. Where the model runs into difficulty, however, is the magnitude of the “price of risk” coefficients, λ_1 and λ_2 , which need to be large (in absolute value) to simultaneously account for Fama’s second condition and the unconditional variance of the depreciation rate. The implications for other risk premiums are strongly counterfactual. The implied yield curve, for instance, is hump shaped with long yields reaching as high as 80 percent per annum. In a nutshell, the price of risk required to account for the anomaly seems far in excess of that which generates realistic risk premiums on other fixed-income payoffs.

III. Final Remarks

We have studied the relation between interest rates and currency prices in theory and data. The relation shows up in theory as a restriction on the joint behavior of currencies and state prices in arbitrage-free environments. In

the data, it appears in the form of the forward premium anomaly, in which interest rate differentials help to predict future movements in currency prices. We find that the anomaly imposes further conditions on affine models: Either interest rates must be negative with positive probability or the effects of one or more factors on pricing kernels must differ across currencies (what we term interdependence). Our estimates indicate that each of these alternatives has serious shortcomings when asked to provide a quantitative account of both the anomaly and the salient properties of interest rates and exchange rates.

We are left with two outstanding issues. The first is whether affine models with a small number of state variables are capable of approximating the properties of currency prices and a more comprehensive set of fixed income securities. Ahn (1997) and Saá-Requejo (1994) have made some progress along these lines, extending the analysis to yields on bonds with longer maturities. The second is the economic foundations of pricing kernels that reproduce the anomaly. We have followed a “reverse-engineering” strategy in which pricing kernels are simply stochastic processes that account for observed asset prices, but one might reasonably ask what kinds of behavior by private agents and policy makers might lead to such pricing kernels. Several possibilities are outlined by Alvarez, Atkeson, and Kehoe (1999), Bakshi and Chen (1997), Stulz (1987), and Yaron (1995), who develop dynamic general equilibrium models in which interest rates and currency prices reflect preferences, endowments, and monetary policies. Perhaps further work will connect pricing kernels in these models to properties of interest rates, currency prices, output growth, and monetary aggregates.

Appendix A. Proof of Proposition 2

We begin with some preliminary results on affine yield models. The process for z requires that the volatility functions from Section I.C, $v_i = \alpha_i + \beta_i^\top z$, be positive. We define the set D of *admissible* states as those values of z for which volatility is positive:

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

Duffie and Kan (1996, Section 4) show that z remains in D if the process satisfies the following condition.

Condition A: For each i :

1. for all $z \in D$ satisfying $v_i(z) = 0$ (the boundary of positive volatility), the drift is sufficiently positive: $\beta_i^\top (I - \Phi)(\theta - z) > \beta_i^\top \beta_i/2$; and
2. if the j th component of β_i is nonzero for any $j \neq i$ then $v_i(z)$ and $v_j(z)$ are proportional to each other (their ratio is a positive constant).

We refer to models characterized by equations (15) and (16) and satisfying Condition A as the *Duffie-Kan class of affine models*.

In the affine models of Section I.C, bond prices are log-linear functions of the state variables z . If b_t^n denotes the price at date t of a claim to one dollar in all states at date $t + n$, then

$$-\log b_t^n = A(n) + B(n)^\top z_t$$

for some parameters $\{A(n), B(n)\}$. Because bond yields are $y_t^n = -n^{-1} \log b_t^n$, they are linear in z :

$$y_t^n = A'(n) + B'(n)z_t,$$

where $A'(n) = n^{-1}A(n)$ and $B'(n) = n^{-1}B(n)$. We use the pricing relation (3) to generate parameters recursively:

$$\begin{aligned} A(n+1) &= A(n) + \delta + B(n)^\top (I - \Phi)\theta - \frac{1}{2} \sum_{j=1}^k (\lambda_j + B(n)_j)^2 \alpha_j \\ B(n+1)^\top &= (\gamma^\top + B(n)^\top \Phi) - \frac{1}{2} \sum_{j=1}^k (\lambda_j + B(n)_j)^2 \beta_j^\top, \end{aligned}$$

starting with $A(0) = 0$ and $B(0) = 0$. We say that a model is *invertible* if there exist k maturities for which the matrix

$$B = [B'(n_1) \cdots B'(n_k)]$$

is nonsingular. The assumption of invertibility is not restrictive: If a model is not invertible, we can construct an equivalent invertible model with a smaller state vector. Analogously, define the vector $A^\top = [A'(n_1), \dots, A'(n_k)]$.

Consider now a subclass of affine models in which bond yields are always positive.

LEMMA 1: *Consider the Duffie-Kan class of affine models. If the model is invertible and bond yields are positive for all admissible states z , then β is diagonal with strictly positive elements.*

In words: The volatility functions have the univariate square root form

$$v_i(z) = \alpha_i + \beta_{ii} z_i,$$

with strictly positive β_{ii} . As a consequence, β has full rank. This rules out both pure Gaussian factors like $\beta_i^\top = (0, 0, \dots, 0)$ and multivariate factors like $\beta_i^\top = (1, 1, \dots, 1)$.

Proof: Suppose, in contradiction to the lemma, that β has less than full rank. Then there exists a nonzero vector h satisfying $\beta^\top h = 0$. For any admissible z , $z' = z + \rho h$ is also admissible for any real ρ because it generates the same values for the volatility functions. Now consider bond yields. For yields to be positive, we need bond prices to be less than one. If y denotes a vector of yields for a set of maturities for which B is invertible, then we need

$$y = A + B^\top z \geq 0$$

for all admissible z . Because $z' = z + \rho h$ is also admissible, we have

$$y = A + B^\top z + \rho B^\top h.$$

By assumption, B is invertible, so $B^\top h \neq 0$. Thus we can choose ρ to make yields as negative as we like, thereby violating the premise of the lemma. We conclude that β has full rank. Condition A(2) then tells us that β must be diagonal. Q.E.D.

A second result is that in this environment (univariate volatility functions), $\text{var}(z)$ has nonnegative elements.

LEMMA 2: *Consider the Duffie-Kan class of affine models in which β is diagonal with strictly positive elements β_{ii} . Then $\Omega = \text{var}(z)$ has all positive elements.*

Proof: The proof hinges on Condition A(1), Duffie and Kan's multivariate analog of the Feller condition. Because β is diagonal with positive elements, the condition implies that for each i

$$\sum_{j=1}^k \kappa_{ij}(\theta_j - z_j) \geq \beta_{ii}/2 > 0.$$

for all admissible z satisfying $v_i(z) = 0$, where $K = I - \Phi$ has elements κ_{ij} . The new ingredient relative to the univariate Feller condition is the effect of variables $z_j, j \neq i$, on the drift of z_i . The structure placed on β means that the set of admissible z s includes values of z_j that are arbitrarily large. The condition therefore implies $\kappa_{ij} \leq 0$ for all $j \neq i$. The admissible set also includes $z_j = \theta_j$, so $\kappa_{ii} > 0$. Because, by assumption, $\Phi = I - K$ has positive diagonal elements, $0 < \kappa_{ii} < 1$. This additional step is an artifact of our discrete time approximation: If κ_{ii} were greater than one, we would simply choose a smaller time interval.

We have established that the elements of Φ are nonnegative. We now show that the unconditional variance of z , which we denote by the matrix Ω , has no negative elements. Because z is a first-order autoregression with stable Φ , its variance is the solution to

$$\Omega = \Phi \Omega \Phi^\top + V(\theta),$$

where $V(\theta)$ is a diagonal matrix with positive elements $v_i(\theta_i) = \alpha_i + \beta_{ii}\theta_i$. Because Φ is stable, we can compute Ω iteratively using

$$\Omega_{j+1} = \Phi\Omega_j\Phi^\top + V(\theta),$$

starting with $\Omega_0 = 0$. We see that at each stage, the elements of Ω_{j+1} are sums of products of nonnegative numbers, so we conclude that the elements of Ω are nonnegative. Q.E.D.

We come at last to a proof of Proposition 2. The proposition is based on the model described in Section II.B. It has three independent state variables or factors: a common factor z_0 and currency-specific factors z_1 and z_2 . The common factor has, by construction, no influence on currency prices or the forward premium. It therefore has no influence on the anomaly, and we can disregard it.

With this simplification, interest rates in the two currencies are

$$\begin{aligned} r_t &= (\delta - \omega) + (\gamma - \tau)^\top z_{1t} \\ r_t^* &= (\delta^* - \omega^*) + (\gamma^* - \tau^*)^\top z_{2t}, \end{aligned}$$

where $\omega = \sum_j \lambda_j^2 \alpha_j^1$, $\omega^* = \sum_j \lambda_j^{*2} \alpha_j^2$, $\tau = \sum_j \lambda_j^2 \beta_j^1/2$, and $\tau^* = \sum_j \lambda_j^{*2} \beta_j^2/2$. The forward premium is therefore

$$f_t - s_t = (\gamma - \tau)^\top z_{1t} - (\gamma^* - \tau^*)^\top z_{2t}$$

and the depreciation rate is

$$\begin{aligned} s_{t+1} - s_t &= (\delta - \delta^*) + \gamma^\top z_{1t} - \gamma^{*\top} z_{2t} + \lambda^\top V^1(z_{1t})^{1/2} \varepsilon_{1t+1} \\ &\quad - \lambda^{*\top} V^2(z_{2t})^{1/2} \varepsilon_{2t+1}. \end{aligned}$$

The anomaly therefore requires

$$\text{cov}(s_{t+1} - s_t, f_t - s_t) = (\gamma - \tau)^\top \text{var}(z_1)\gamma + (\gamma^* - \tau^*)^\top \text{var}(z_2)\gamma^* < 0. \quad (\text{A1})$$

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

The condition that interest rates are positive for all admissible states places restrictions on the parameters. For the “dollar” short rate r we need the elements of $\gamma - \tau$, and hence of γ , to be nonnegative, because $\tau > 0$ and each element of z_1 is unbounded above. By Lemma 2, $\text{var}(z_1)$ has nonnegative elements, so the bilinear form

$$(\gamma - \tau)^\top \text{var}(z_1)\gamma$$

is nonnegative. Identical reasoning applies to the second term in equation (A1). We conclude that the model cannot reproduce the anomaly with strictly positive interest rates. Q.E.D.

Appendix B. Distribution of State Variables

We clarify the role played by the Feller condition in determining higher moments of state variables following square-root processes, an issue that arises in estimates of the independent factor model (Section II.D).

Consider a state variable z following the square-root process equation (19). The unconditional distribution of z is Gamma with density

$$f(z) = [b^a \Gamma(a)]^{-1} z^{a-1} e^{-z/b}.$$

and parameters $a, b > 0$. This statement is exact in continuous time, approximate in discrete time. The mean and variance of a Gamma random variable are $ab = \theta$ and $ab^2 = \theta\sigma^2/(1 - \varphi^2)$, which defines the parameters as

$$a = (1 - \varphi^2)\theta/\sigma^2$$

$$b = \sigma^2/(1 - \varphi^2).$$

In the continuous-time limit, $1 - \varphi^2 \rightarrow 2\kappa = 2(1 - \varphi)$, so $a \rightarrow 2(1 - \varphi)\theta/\sigma^2$, the ratio in the Feller condition, inequality (20). This ratio governs the distribution's higher moments:

$$\gamma_1 = \frac{E(z - \theta)^3}{[\text{var}(z)]^{3/2}} = \frac{2}{a^{1/2}} \text{ (skewness)}$$

$$\gamma_2 = \frac{E(z - \theta)^4}{[\text{var}(z)]^2} = \frac{6}{a} \text{ (kurtosis)}$$

Parameter values implying $a > 1$ satisfy the Feller condition. Smaller values violate the condition and generate large values of the skewness and kurtosis measures, γ_1 and γ_2 .

REFERENCES

- Ahn, Dong-Hyun, 1997, Common factors and local factors: Implications for term structures and exchange rates, Unpublished manuscript, University of North Carolina.
- Ahn, Dong-Hyun, 1998, Generalized squared-autoregressive-independent-variable term structure model, Unpublished manuscript, University of North Carolina.
- Amin, Kaushik, and Robert Jarrow, 1991, Pricing foreign currency options under stochastic interest rates, *Journal of International Money and Finance* 10, 310–329.
- Alvarez, Fernando, Andrew Atkeson, and Patrick Kehoe, 1999, Volatile exchange rates and the forward premium anomaly: A segmented asset market view, Unpublished manuscript, Federal Reserve Bank of Minneapolis.

- Backus, David, Allan Gregory, and Chris Telmer, 1993, Accounting for forward rates in markets for foreign currency, *Journal of Finance* 48, 1887–1908.
- Bakshi, Gurdip, and Zhiwu Chen, 1997, Equilibrium valuation of foreign exchange claims, *Journal of Finance* 52, 799–826.
- Bansal, Ravi, 1997, An exploration of the forward premium puzzle in currency markets, *Review of Financial Studies* 10, 369–403.
- Bansal, Ravi, A. Ronald Gallant, Robert Hussey, and George Tauchen, 1995, Nonparametric estimation of structural models for high frequency currency market data, *Journal of Econometrics* 66, 251–287.
- Bekaert, Geert, 1995, The time variation of expected returns and volatility in foreign exchange markets, *Journal of Business and Economic Statistics* 13, 397–408.
- Bekaert, Geert, 1996, The time-variation of risk and return in foreign exchange markets: A general equilibrium perspective, *Review of Financial Studies* 9, 427–470.
- 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, 1997, The implications of first-order risk aversion for asset market risk premiums, *Journal of Monetary Economics* 40, 3–39.
- Canova, Fabio, and Jane Marrinan, 1995, Predicting excess returns in financial markets, *European Economic Review* 39, 35–69.
- Constantinides, George, 1992, A theory of the nominal term structure of interest rates, *Review of Financial Studies* 5, 531–552.
- 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 University Press, Princeton, NJ).
- Duffie, Darrell, and Rui Kan, 1996, A yield-factor model of interest rates, *Mathematical Finance* 6, 379–406.
- Duffie, Darrell, and Kenneth Singleton, 1997, An econometric model of the term structure of interest rate swap yields, *Journal of Finance* 52, 1287–1321.
- Engel, Charles, 1996, The forward discount anomaly and the risk premium: A survey of recent evidence, *Journal of Empirical Finance* 3, 123–191.
- 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.
- Fama, Eugene, 1984, Forward and spot exchange rates, *Journal of Monetary Economics* 14, 319–338.
- Flood, Robert, and Andrew Rose, 1996, Fixes: Of the forward discount puzzle, *Review of Economics and Statistics* 78, 748–752.
- Frachot, Antoine, 1996, A reexamination of the uncovered interest rate parity hypothesis, *Journal of International Money and Finance* 15, 419–437.
- 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.
- Hansen, Lars, and Robert Hodrick, 1983, Risk averse speculation in the forward foreign exchange market, in Jeffrey A. Frankel, ed.: *Exchange Rates and International Macroeconomics* (University of Chicago Press, Chicago, IL).
- Hodrick, Robert, 1987, *The Empirical Evidence on the Efficiency of Forward and Futures Foreign Exchange Markets* (Harwood Academic Publishers, New York).
- Hollifield, Burton, and Raman Uppal, 1997, An examination of uncovered interest rate parity in segmented international commodity markets, *Journal of Finance* 52, 2145–2170.

- 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.
- Pearson, Neil, and Tong-Sheng Sun, 1994, Exploiting the conditional density in estimating the term structure: An application to the Cox, Ingersoll, and Ross model, *Journal of Finance* 54, 1279–1304.
- Saá-Requejo, Jesús, 1994, The dynamics and the term structure of risk premia in foreign exchange markets, Unpublished manuscript, INSEAD.
- Stuart, Alan, and J. K. Ord, 1987, *Kendall's Advanced Theory of Statistics, Volume 1* (Oxford University Press, New York).
- Stulz, René, 1987, An equilibrium model of exchange rate determination and asset pricing with nontraded goods and imperfect competition, *Journal of Political Economy* 95, 1024–1040.
- 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.
- Yaron, Amir, 1995, Liquidity shocks and international asset pricing: A theoretical and empirical investigation, Unpublished manuscript, Carnegie Mellon University.