

NBER WORKING PAPER SERIES

“PESO PROBLEM” EXPLANATIONS FOR
TERM STRUCTURE ANOMALIES

Geert Bekaert
Robert J. Hodrick
David A. Marshall

Working Paper 6147
<http://www.nber.org/papers/w6147>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
August 1997

We are grateful for the comments of Christian Gilles, Karen Lewis, Chris Telmer, Paula Tkac and of participants in the many seminars at which this article was presented. Geert Bekaert gratefully acknowledges financial support from a NSF grant and the Financial Research Initiative at Stanford University. Andrew Ang provided excellent research assistance. This paper is part of NBER's research program in Asset Pricing. Any opinions expressed are those of the authors and not those of the National Bureau of Economic Research.

© 1997 by Geert Bekaert, Robert J. Hodrick and David A. Marshall. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

“Peso Problem” Explanations for Term
Structure Anomalies
Geert Bekaert, Robert J. Hodrick
and David A. Marshall
NBER Working Paper No. 6147
August 1997
JEL Nos. G1, F3, E4, C5
Asset Pricing

ABSTRACT

We examine the empirical evidence on the expectations hypothesis of the term structure of interest rates in the United States, the United Kingdom, and Germany using the Campbell-Shiller (1991) regressions and a vector-autoregressive methodology. We argue that anomalies in the U.S. term structure, documented by Campbell and Shiller (1991), may be due to a generalized peso problem in which a high-interest rate regime occurred less frequently in the sample of U.S. data than was rationally anticipated. We formalize this idea as a regime-switching model of short-term interest rates estimated with data from seven countries. Technically, this model extends recent research on regime-switching models with state-dependent transitions to a cross-sectional setting. Use of the small sample distributions generated by the regime-switching model for inference considerably weakens the evidence against the expectations hypothesis, but it remains somewhat implausible that our data-generating process produced the U.S. data. However, a model that combines moderate time-variation in term premiums with peso-problem effects is largely consistent with term structure data from the U.S., U.K., and Germany.

Geert Bekaert
Graduate School of Business
Stanford University
Stanford, CA 94305
and NBER
fbekaert@gsb-lira.stanford.edu

Robert J. Hodrick
Graduate School of Business
Columbia University
418 Uris Hall
New York, NY 10027
and NBER
robert.hodrick@columbia.edu

David A. Marshall
Federal Reserve Bank of Chicago
230 South La Salle
Chicago, IL 60604
dmarshall@frbchi.org

When researchers test the expectations hypothesis of the term structure with U.S. data, an interesting paradox emerges. Briefly, the change in the long-term interest rate does not behave as predicted by the expectations hypothesis, whereas future short rates do change in the direction predicted by the expectations hypothesis. Even so, at the short end of the term structure, future short rates do not move enough and the theory is still rejected (Campbell and Shiller (1991)). General equilibrium attempts to explain these observations with time-varying risk premiums generally fail.¹ The goal of our project is to see whether these problems may be driven, not by a failure of the economic theory, but by a failure of the asymptotic distribution theory used to test these models.

There is good reason to be suspicious of the asymptotic distribution theory underlying previous research. In Bekaert, Hodrick and Marshall (1997a), we document extreme small sample biases and deviations from asymptotic distribution theory in standard tests of the expectations hypothesis. Given the extreme persistence of short interest rates, the samples that researchers have available to study the expectations hypothesis are too small for asymptotic distribution theory to allow correct inference. One result in Bekaert, Hodrick and Marshall (1997a) is that the small sample distributions of different tests of the expectations hypothesis imply a more uniform rejection of the null in U.S. data than would be concluded from the asymptotic distributions. Nevertheless, the evidence against the expectations hypothesis is much weaker in other countries (see Hardouvelis (1994), Jorion and Mishkin (1991) and Gerlach and Smets (1997)) so that the U.S. data still appear puzzling.

In this paper, we attempt to explain the anomalous patterns in the U.S. term structure by focusing on a particular issue in small sample inference known as the "peso problem". Peso problems

¹ Recent examples of general equilibrium macroeconomic models of the term structure include Backus, Gregory and Zin (1989), den Haan (1995), and Bekaert, Hodrick and Marshall (1997b). Backus, Foresi, Mozumdar and Wu (1997), Fisher and Gilles (1996) and Roberds and Whiteman (1996) develop the implications of the Affine Class of general equilibrium financial models for the Campbell-Shiller (1991) results.

are often defined narrowly to arise when the distribution of the data generating process includes a low probability, usually catastrophic, state that generates extreme disutility to economic agents.² Because this state has low probability, it is unlikely to be observed in a given small sample of data. Because the state is catastrophic, the possibility that this state may occur substantially affects agents' decisions, which in turn determines equilibrium prices and rates of return. Here, we define a peso problem more broadly as arising whenever the ex post frequencies of states within the data sample differ substantially from their ex ante probabilities, and where these deviations distort econometric inference. When a peso problem is present, the sample moments calculated from the available data do not coincide with the population moments that agents actually use when making their decisions.

We formalize this peso problem intuition in a regime-switching model where changes in policy regimes affect the distribution of the short-term interest rate. In particular, returns to investors in bond markets depend crucially on whether the monetary policy regime implies low, moderate, or high inflation. Today's bond prices will reflect investors' beliefs about the probabilities of these alternative regimes in the future. A peso problem could arise if agents repeatedly believe that the probability of a regime switch is higher than average, but these switches fail to occur ex post. In principle, such a peso problem could explain patterns found in U.S. term structure data, in which increases (decreases) in the term spread fail to be followed, on average, by increases (decreases) in future long-term yields. As we demonstrate in section 2, below, the empirical evidence suggests a peso problem of this sort. The evidence against the expectations hypothesis is the weakest in the UK, which experienced the largest frequency of inflationary episodes during the sample period. Furthermore, the evidence against the expectations hypothesis is also considerably weaker when term spreads are not exceptionally large (in absolute value). Unusually large term spreads may coincide with periods

² For example, this is the type of peso problem considered in Rietz's (1988) explanation of the equity premium puzzle.

in which agents anticipate a switch to a high inflation regime. This suggests that expectations of regime shifts may play a role in explaining observed term structure patterns.

Expectations of regime changes have a long history as potential explanations of puzzling economic phenomena. While theoretical rational expectations models in the 1970's often analyzed how a financial variable, such as an interest rate or an exchange rate, would jump and evolve in anticipation of some future change in government policy, it was not until the 1980's that formal econometric models were developed incorporating these ideas. Flood and Garber (1980) is an early empirical contribution which used the term "process switching" to describe the evolution of monetary policy regimes and the possibility of monetary reform that agents must assess during a hyperinflation. In the 1970's it was also recognized that failure to model expectations of changes in regimes could seriously compromise time series econometric analysis of financial data. Krasker (1980) was the first to publish an econometric analysis of this issue as it related to exchange rates. Of course, the literature on the term structure of interest rates has also recognized that changes in regimes are important phenomena.³

There is an obvious difficulty in empirically implementing the peso-problem intuition: Some such explanation can always be constructed to explain observed return anomalies, but, by definition, the proposed explanation cannot be tested using the small sample of data available to the econometrician. We attempt to resolve this difficulty by estimating our regime-switching model using data from different countries simultaneously. In effect, we overcome the problem of the small sample from U.S. observations by adding data from other countries. Of course, doing so requires the strong assumption that the data from other countries are drawn from the same unconditional distribution as the U.S. data. We regard this as a reasonable starting point. The developed countries of the

³Hamilton (1988), Lewis (1991), Evans and Lewis (1994), Naik and Lee (1994), Sola and Driffill (1994) and Gray (1996) all consider how changes in regimes affect the empirical analysis of the term structure of interest rates. Evans (1995) surveys the evolution of the literature on peso problems.

world face a similar set of technological shocks and have similar political economies. Countries' rates of inflation and real interest rates therefore vary over time for similar reasons, but the inflation experiences and real interest rates observed in any given country during a limited time period represent potential realizations that could occur in any of the other countries. Rational investors will use information from other countries' inflation and real interest rate histories in determining the probability distribution of future nominal interest rates in their own country.⁴

Our regime-switching model of the short-term interest rate extends the univariate model in Gray (1996) to a cross-sectional setting of multiple countries with multiple regimes. We estimate the model using data from seven countries. We then incorporate the estimated short rate process into two term structure models. The first imposes the expectations hypothesis: that is, term premiums are assumed to be constant over time. In the second, we consider the possibility that regime switches may induce time-variation in term premiums. We do so by assuming that long-term interest rates are generated by a pricing kernel similar to that found in the affine class of general equilibrium models (see Duffie and Kan (1996), Backus (1993)). For each of these models, we compute the small sample distributions of several widely-used tests of the expectations hypothesis, and we use these small sample distributions to re-evaluate these test statistics. We also ask whether samples that are characterized by peso-problems resemble the U.S. data.

In the model with constant term premiums, we find that the evidence against the expectations hypothesis is substantially weaker when the small sample distribution is used. Interestingly, these results emerge in spite of the extreme positive small sample bias in many of these statistics, a bias that tends to work against the expectations hypothesis. The reason is that, in addition to this positive bias,

⁴ While inflation expectations are an important interest rate determinant that may be subject to regime switches, many empirical studies also document substantial variability in real interest rates. This may explain why regime switches in inflation do not coincide with regime switches in nominal interest rates (Evans and Lewis (1995)).

the small sample distributions display extreme dispersion. Furthermore, there is evidence that the patterns found in U.S. data are associated with "peso problem samples" in which the high-rate regime occurs infrequently. In the model with time-varying term premiums, these effects are magnified. Even though the amount of time-variation in term premiums is small and the population values of the test statistics are very close to those implied by the expectations hypothesis, the effect on the variance of the small sample distributions is striking. With only one exception, the statistics do not reject this model when the small sample inference is used. As in the previous model, the unusual patterns found in U.S. data are associated with simulated data samples in which peso-problem effects are present. We conclude that a combination of peso problems and moderate time-variation in term premiums can account for many of the anomalies found in the U.S. term structure.

The structure of the paper is the following. Section 1 briefly reviews the expectations hypothesis, and discusses why it is a useful starting point for our investigations. Section 2 sets forth evidence on the expectations hypothesis using data from the United States, the United Kingdom, and Germany. Section 3 develops our regime-switching model that formalizes the peso problem, and section 4 presents estimates of the model's parameters. In section 5, we present the small sample distributions of term-structure test statistics implied by our estimates when the expectations hypothesis is imposed. We then evaluate whether our peso-problem intuition can explain the empirical behavior of these statistics. In section 6, we perform a similar analysis for our model with time-varying term premiums. Section 7 concludes and outlines some directions for future research.

1. The Expectations hypothesis of the Term Structure

In economies that do not admit arbitrage opportunities, the term structure of interest rates follows the relation

$$\exp[-r(t,n)n] = E_t \left[\prod_{i=1}^n M_{t+i} \right], \quad (1)$$

where $r(t,n)$ denotes the continuously-compounded return per period on a zero-coupon bond purchased at date t and paying \$1 in $t+n$, and M_t denotes the pricing kernel for nominal assets purchased at date $t-1$ that pay off at date t . We follow Campbell and Shiller (1991) in defining the expectations hypothesis of the term structure of interest rates as the hypothesis that continuously-compounded long interest rates (the yields on long-term, zero-coupon bonds) are weighted averages of expected future values of continuously-compounded short interest rates, possibly with an additive time-invariant term premium. Formally,

$$r(t,n) = \frac{1}{n} \sum_{i=0}^{n-1} E_t r(t+i,1) + c(n). \quad (2)$$

Equation (2) can be derived from the basic asset-pricing equation (1) under particular distributional assumptions. Let $m_t \equiv \log(M_t)$, let

$$m_{t+n}^n \equiv \log \left[\prod_{i=1}^n M_{t+i} \right] = \sum_{i=1}^n m_{t+i} \quad (3)$$

and let $v_t^j(j) \equiv E_t(m_{t+n}^n - E_t m_{t+n}^n)^j$, the j^{th} conditional central moment of m_{t+n}^n conditional on date t information. If all conditional moments of m_{t+n}^n exist, then Taylor's theorem implies⁵

$$r(t,n) = \frac{1}{n} \sum_{i=1}^n E_t r(t+i-1,1) + c(t,n), \quad (4)$$

where the term premium $c(t,n)$ is defined by

⁵ A derivation of equations (4) and (5) can be found in the Appendix.

$$c(t,n) \equiv \frac{1}{n} \left[E_t \sum_{i=1}^n \log \left(1 + \sum_{j=2}^{\infty} \frac{v_{t+i-t}(j)}{j!} \right) - \log \left(1 + \sum_{j=2}^{\infty} \frac{v_t(j)}{j!} \right) \right]. \quad (5)$$

By comparing equation (4) with equation (2), we see that the expectations hypothesis holds in any economy where m_{t+n}^n is conditionally homoskedastic in the strong sense that *all* higher conditional moments are time-invariant. In particular, the expectations hypothesis is consistent with any amount of risk aversion, provided this strong-sense conditional homoskedasticity condition holds. For example, in the familiar consumption-based pricing model with time-additive utility and constant relative risk aversion preferences, $m_{t+n}^n = -\gamma \log(C_{t+n}/C_t) - \log(P_{t+n}/P_t) + \text{constant}$, where C_t denotes real consumption at date t , P_t denotes the money price of consumption at date t , and γ is the coefficient of relative risk aversion. If log consumption growth and log inflation are conditionally joint normal and homoskedastic, the expectations hypothesis holds regardless of the degree of risk aversion, as measured by γ .

It should be noted that, as a matter of theory, conditional heteroskedasticity of $r(t,1)$ is not incompatible with conditional homoskedasticity of m_{t+n}^n since the functional relationship between the forecast error of m_{t+n} and the corresponding forecast error of $r(t,1)$ has been left unspecified. Therefore, the well-documented conditional heteroskedasticity in interest rates need not, in and of itself, rule out the expectations hypothesis. Having said this, it is clearly difficult to rationalize the expectations hypothesis within the context of commonly-used models of the pricing kernel. For example, if the distribution of m_{t+n}^n conditional on date t information is normal, equation (5) simplifies to⁶

⁶ A derivation of equation (6) is in the appendix.

$$c(t,n) = \frac{1}{n} \sum_{i=1}^n \left[\frac{1}{2} E_t(\text{var}_{t+i-1}(m_{t+i})) - \frac{1}{2} \text{var}_t(m_{t+i}) - \sum_{j=i+1}^n \text{cov}_t(m_{t+i}, m_{t+j}) \right]. \quad (6)$$

Conditional heteroskedasticity in $E_t m_{t+i}$ is likely to imply time-variation in $\text{cov}_t(m_{t+i}, m_{t+j})$, for $i, j > 1$. Thus, we would not argue that m_{t+n}^a is, in fact, conditionally homoskedastic. We do believe, however, that it is difficult to construct a reasonable economic model in which m_{t+n}^a displays *sufficient* conditional heteroskedasticity to generate the term structure patterns documented in section 2, below.⁷ We therefore think it wise to consider whether small sample econometric problems can account for these patterns. The expectations hypothesis is a natural starting point, since it imposes the extreme assumption that *no* portion of the expectation hypothesis's failure can be attributed to time variation in term premiums. We impose this assumption in section 5, below. In section 6, this extreme assumption is relaxed.

2. Evidence on the Expectations hypothesis of the Term Structure

2.1 The Campbell-Shiller Regressions

Campbell and Shiller (1991) propose the following tests of equation (2) that involve current term spreads, which are the differences between the yields on long-term bonds and the yields on bonds with shorter maturities, $s(t,n) \equiv r(t,n) - r(t,m)$, where $n > m$ and $k \equiv n/m$ is an integer. First, equation (2) implies that a maturity specific multiple of the term spread predicts the m -period change in the longer term bond yield. In particular, the slope coefficient α_1 should equal unity in the following regression equation:

⁷ Backus and Zin (1994) make a similar point.

$$r(t+m, n-m) - r(t, n) = \alpha_0 + \alpha_1 \left[\frac{m}{n-m} \right] [r(t, n) - r(t, m)] + u(t+m). \quad (7)$$

In equation (7), the error term $u(t+m)$ is a moving average process of order $m-1$. Second, equation (2) implies that the current term spread between the n -period yield and the m -period yield forecasts a weighted average of future m -period interest rates minus the current m -period rate. In particular, the slope coefficient δ_1 should equal unity in the following regression equation:

$$\frac{1}{k} \left[\sum_{i=0}^{k-1} r(t+i, m, m) \right] - r(t, m) = \delta_0 + \delta_1 [r(t, n) - r(t, m)] + \nu(t+n-m). \quad (8)$$

In equation (8), the error term $\nu(t+n-m)$ is a moving average process of order $n-m-1$.⁸

Table 1 displays results from regression tests of equations (7) and (8) using data from government bonds denominated in the three different currencies, the U.S. dollar (USD), the British pound (GBP), and the Deutsche mark (DEM).⁹ The regressions based on equation (7) use the approximation $r(t+3, n) = r(t+3, n-3)$ because data on zero discount bonds with maturities of $n-3$ months are not available.¹⁰

Two things are noteworthy in Table 1. First, there is strong evidence against the expectations hypothesis using the USD, especially from equation (7). Second, the evidence against the expectations hypothesis is much weaker using GBP and DEM data than it is using USD data. If one

⁸ A third way to investigate the implications of the term structure of interest rates, which is closely related to the second Campbell-Shiller (1991) specification, is to examine the forward interest rates implicit in the term structure as predictors of future spot interest rates. (see Fama (1984), Fama and Bliss (1987), Stambaugh (1988), and Backus, Foresi, Mozumdar and Wu (1997).)

⁹ See the Data Appendix for a description of the data.

¹⁰ Bekaert, Hodrick and Marshall (1997a) find severe biases in the small sample distributions of the regression test statistics for equations (7) and (8). The biases are positive, and the use of the approximation $r(t+3, n) = r(t+3, n-3)$ produces even larger positive biases. Since all three currencies would be subject to the same biases, we do not attempt to bias-adjust the results.

uses the asymptotic distributions of the OLS slope estimators, one concludes that the GBP regressions show only slight evidence against the hypothesis that the slope coefficients equal unity for both regressions. The regressions using DEM data reject the hypothesis of a unit slope coefficient for all three horizons in equation (7), but the point estimates are far less negative than in the USD regressions. The results for equation (8) using DEM data show evidence against the expectations hypothesis only for the 12-month bond.¹¹

2.2 VAR Statistics

Campbell and Shiller (1991) propose alternative tests of the expectations hypothesis based on vector autoregressions (VARs). In our implementation of their approach, the first variable in the VAR is the m -period rate, and the other variables are spreads between longer-period rates and the m -period rate. Let the "theoretical spread" (denoted $s'(t,n)$) be the term spread that satisfies the

expectations hypothesis. From equation (8), $s'(t,n) \equiv (1/k) \sum_{i=0}^{k-1} E_t[r(t+i \cdot m, m)] - r(t, m)$ If the

likelihood function of $s'(t,n)$ is constructed using the fitted values of the VAR, one obtains the

hypothesis are the correlation between the theoretical spread, $s'(t,n)$, and the actual spread, $s(t,n)$, and the ratio of the standard deviation of $s'(t,n)$ to the standard deviation of $s(t,n)$. Both statistics are functions of the coefficients of the VAR and the covariance matrix of the VAR innovations. Under the expectations hypothesis, both should equal one. We also compute the slope coefficients implied by the VARs for equations (7) and (8), which are given by

$$\frac{(e_n' + e_1')(A^3 - I)C_0 e_n(n-3)/3}{e_n' C_0 e_n} \quad (10)$$

and

$$\frac{e_1' [(1/k)(I - A^3)(I - A^3)^{-1} - I]C_0 e_n}{e_n' C_0 e_n} \quad (11)$$

(In equations (10) and (11), e_n is an indicator vector that picks out $s(t,n)$, and C_0 is the unconditional covariance matrix of the $z(t)$ vector, which is calculated using $\text{vec}(C_0) = (I - A \otimes A)^{-1} \text{vec}(\Omega)$ where Ω is the innovation covariance matrix of $z(t)$.)

For each country, we construct four-variable VARs incorporating the 3-month rates and the spreads between three long rates (12-months, 36-months, and 60-months) and the 3-month rate. The VAR methodology requires choice of the appropriate lag length. Table 2 reports three diagnostic statistics for VARs of first through fourth order. For the USD, sequential likelihood ratio tests and the Akaike (1973) criterion indicate that a fourth-order specification would be preferred, but the more conservative Schwarz (1978) criterion chooses a second-order specification. For the GBP, the sequential likelihood ratio tests indicate a third-order specification, the Akaike criterion indicates a second-order, and the Schwarz criterion indicates a first-order. For the DEM, the sequential likelihood ratio tests indicate a third-order specification, while the Akaike and Schwarz criteria indicate a second-order specification is preferred. Given these conflicting diagnostic tests and our desire to estimate a common order for each currency, we choose a second-order VAR.

Table 3 reports the VAR statistics discussed above for the three currencies, as well as a joint estimation that will be discussed below. The correlation statistics generally increase with the horizon. This is consistent with the results in Campbell and Shiller (1991) and Bekaert, Hodrick, and Marshall (1997a). In contrast to the findings in those papers, the standard deviation ratios are not all significantly less than one for our sample period. In fact, the point estimates for the USD and the GBP at $n=60$ are larger than one. Overall, these two statistics do not present a uniform case against the expectations hypothesis. The implied regression coefficients associated with equation (7) (as given in equation (10)) are very close to their OLS counterparts in Table 1. For equation (8), the implied slope coefficient (as given in equation (11)) for the USD differ from their OLS counterparts by no more than 0.082, but the implied slope coefficients for the GBP and the DEM are considerably smaller than their OLS counterparts for the longer horizons ($n = 36$ and 60).

2.3 Alternative Interpretations of the Evidence

The estimated coefficients for equations (7) and (8) for the USD in Table 1 are similar to those in Campbell and Shiller (1991). Campbell and Shiller (1991, p. 505) note that the two sets of regressions produce an apparent paradox:

[T]he slope of the term structure almost always gives a forecast in the wrong direction for the short-term change in the yield on the longer bond, but gives a forecast in the right direction for long-term changes in short rates.

Campbell and Shiller's (1991 p. 523) suggested interpretation of the data is an over-reaction model of the yield spread:

The long rate differs from the short rate in the direction implied by the expectations hypothesis; however, the spread between the two rates is larger than can be justified by rational expectations of future short rate changes.

An alternative interpretation of the data, which we investigate here, is that these results are

driven by small sample anomalies due to peso problems in the data analysis.¹² Consider an intuitive explanation of the peso problem. Suppose that short-term interest rates can evolve in three different regimes, with the mean and volatility of interest rates increasing together as we move across regimes. Further, suppose that any shock that increases (decreases) the short-term rate also increases the probability of switching to a higher-rate (lower-rate) regime. Then, as short-term rates rise, the term spread may rise as agents rationally forecast transitions into a higher-rate regime. However, if, in a particular sample, the higher rate regimes are observed less frequently than their unconditional probabilities, this increase in the spread will appear unjustified *ex post*. In such a sample, regression (7) will fail to deliver an estimated slope of unity and could produce negative coefficients if increases in the spread are subsequently followed by surprising transitions to lower rate regimes. The slope coefficient estimates in regression (8) will also be less than unity, but the estimation error here is likely to be smaller than in regression (7), since the short rates immediately following the shock will tend to be higher than their unconditional value even if rates stay within a regime because of the high serial correlation of short rates.

Suppose that this small sample explanation of the USD evidence is true and that other countries follow the same regime switching model. Due to sampling variation, the data from these other countries need not resemble the USD experience. The differences in estimated coefficients in Tables 1 and 3 for the different currencies could be due to different countries experiencing different realizations of the same population distribution but in relatively small samples. One would expect there to be less evidence against the expectations hypothesis in countries with a sample that is more representative of the population distribution. High and volatile interest rates were more common over the sample period in the U.K. than in the U.S. and Germany. Strikingly, there is very

¹² The peso problem intuition is but one reason why rational expectations econometrics may be inapplicable. An alternative explanation is that agents take time to learn about the underlying structure of the economy. See, for example, Lewis (1989).

weak evidence against the Expectations Hypothesis in the U.K.

Furthermore, if this peso explanation is true, the data for USD and to a lesser extent DEM may contain observations in which spreads increase dramatically (because a shift to a high-rate regime is anticipated) but the shift does not actually occur.¹³ These observations may have a disproportionate effect on the slope coefficients of regression (7). Table 4 investigates this possibility by allowing for different coefficients on "normal" and "large" spreads in equation (7). For Germany, the coefficients on large spreads are significantly different from one for all maturities, whereas the "normal" coefficients are not statistically significantly different from one. For the U.S., both slope coefficients are significantly different from one for all maturities. Not surprisingly, for the U.K. most coefficients are close to one. Note that for all nine regressions the slope coefficient for the large spreads is lower than for normal spreads.

The possibility that the three countries are following a common process suggests that the VARs of Table 3 should be estimated jointly to find a common set of parameters. We perform this estimation using Hansen's (1982) Generalized Method of Moments (GMM). (Details can be found in the Appendix.) The results of this joint estimation are presented in the last column of Table 3. Perhaps not surprisingly, the results are close to a weighted average of the results for the three individual currencies. This exercise yields little evidence against the expectations hypothesis for the longer maturities. The chief evidence against the expectations hypothesis is now mainly at the short-end of the term structure. When asymptotic distribution theory is used, the implied slope coefficients for equations (7) and (8) and the standard deviation ratio at $n=12$ are significantly at variance with the predictions of the expectations hypothesis. However, the peso problem intuition

¹³ Large negative spreads rarely occur for these countries. In some of the European countries, participating in the European Monetary System, negative spreads have accompanied speculative attacks on the currency and typically correctly predict that future short and long rates will fall back to normal levels in the future. Gerlach and Smets (1997) actually find that the Expectations Hypothesis for short horizons typically holds for such countries.

is fundamentally a problem of small sample inference, so it is not clear whether asymptotic distribution theory is appropriate. To investigate this question, we must formalize the peso-problem intuition. We do so in the next section.

3. A Regime-Switching Model of Interest Rates

In our model, the short interest rate is determined by a regime-switching model in which the regimes follow a Markov process. In each regime the 3-month short interest rate follows a first-order, autoregressive process that is conditionally heteroskedastic. The parameters that determine the mean, the serial correlation, and the conditional variance of the innovations within a regime are all regime dependent. We formalize this structure as follows. Suppose there are K possible regimes. Let s_t be an indicator variable identifying the regime at date t , so $s_t = i$, if regime i prevails at date t . For convenience, we use the simpler notation r_t for the 3-month interest rate $r(t,3)$. This short interest rate follows the law of motion

$$r_{t+1} = \mu_i + \beta_i r_t + h_i(r_t) \epsilon_{t+1} \text{ if } s_{t+1} = i \quad (12)$$

where $\{\epsilon_t\}$ is a sequence of independent standard normal random variables, and the time-varying conditional standard deviation $h_i(r_t)$ is given by

$$h_i(r_t) = \sigma_i r_t^\gamma, \text{ for } i = 1, \dots, K. \quad (13)$$

Notice that the realization of r_{t+1} is affected by two random shocks not known at date t : ϵ_{t+1} and s_{t+1} .

We now specify how the interest rate process shifts among the K possible regimes. It is useful to think of the regimes as corresponding to underlying macroeconomic phenomena like expectations of inflation and the real interest rate, which are highly serially correlated. We will identify higher numbered regimes with higher mean levels of interest rates. Then, it seems reasonable to argue that if $s_t = i$, s_{t+1} will be drawn from the set $\{i-1, i, i+1\}$. With this specification, the regimes characterizing the short interest rates can move gradually up and down, but they cannot jump from

very low values to very high values or vice versa without intermediate transitions.¹⁴

We assume that regime transition probabilities depend on the current state of the economy.

We parameterize these transition probabilities as follows:

$$\text{Prob}(s_{t+1} = i | s_t = i, r_t) = \frac{\exp(a_{ii} + b_{ii}r_t)}{1 + \exp(a_{ii} + b_{ii}r_t)}, \quad i = 1, K \quad (14)$$

$$\text{Prob}(s_{t+1} = i | s_t = i, r_t) = \frac{\exp(a_{ii} + b_{ii}r_t)}{1 + \exp(a_{ii} + b_{ii}r_t) + \exp(a_{ij} + b_{ij}r_t)}, \quad i = 2, \dots, K-1; j = i+1 \quad (15)$$

$$\text{Prob}(s_{t+1} = i+1 | s_t = i, r_t) = \frac{\exp(a_{ij} + b_{ij}r_t)}{1 + \exp(a_{ii} + b_{ii}r_t) + \exp(a_{ij} + b_{ij}r_t)}, \quad i = 2, \dots, K-1; j = i+1. \quad (16)$$

where $\{a_{ij}, b_{ij} \mid i = 1, \dots, K\}$ and $\{a_{ij}, b_{ij} \mid i = 2, \dots, K-1; j = i+1\}$ are parameters of the model. Under equations (14)-(16), the conditional distribution of r_{t+1} given r_t and s_t is a mixture of normals with state-dependent mixing probabilities. Gray (1996) examines a similar model with two regimes and finds that it fits the U.S. short rate data better than alternative models. In our model the probabilities that the high-interest rate regimes occur next period depend both on the current short interest rate and on the current regime. Thus, under the expectations hypothesis of the term structure, long rates are determined by two state variables: the current short rate and the current regime.

4. Estimates of the Regime-Switching Model

We use data from seven different countries to estimate the parameters of the regime switching model. Our hypothesis is that the data from different countries represent different draws

¹⁴ Abrupt changes in monetary regimes can occur in the case of hyperinflations. In particular, monetary reforms often occur suddenly, and the inflation rate can drop dramatically without moving through intermediate regimes of moderate intermediate inflation. Hyperinflations do not occur in any of the countries we consider during our data sample, so we abstract from this possibility in constructing our model.

from the same regime-switching process. Consequently, the parameters of the model are assumed to be the same for all countries, and our reason why different countries' data records may look different in the particular samples is that some countries spend more time in some regimes than in others.

We estimate a three-regime model using data on short interest rates from Australia, Germany, Italy, Japan, Sweden, the United Kingdom, and the United States. We choose these countries in order to get a fairly wide range of interest rate patterns while focusing on large industrial economies. We omit France because its interest rates are closely linked to German rates through the European Monetary System. Similarly, we omit Canada because its rates are highly correlated with those in the U.S. The realizations across countries are treated as independent observations. This independence assumption can only be relaxed at considerable computational cost. The estimated pair-wise correlations among our seven interest rate series range from -0.021 (Japan and Sweden) up to 0.663 (Sweden and Italy). However, these estimates need not constitute evidence of strong cross-sectional correlation. Because short rates are highly persistent stochastic processes with positive drift, substantial positive correlation may arise spuriously. We conducted a Monte Carlo experiment, drawing 2,000 pairs of independent series from our estimated model, each of sample size 297 months (the length of our sample). The 95% quantile cut-off for the estimated pair-wise correlation coefficient was 0.675, and the 99% quantile cut-off was 0.795.

The regime-switching model has twenty parameters that must be estimated: $\{\mu_i, \beta_i, \sigma_i, \gamma_i, a_{ii}, b_{ii}, a_{23}, b_{23}\}$, $i = 1, 2, 3$. We estimate the model using a cross-sectional extension of Gray's (1995) recursive maximum-likelihood procedure. The regime variable for this model is seven - dimensional: $S_t = [s_t^1, \dots, s_t^7]'$, where the superscripts indicate the seven different countries. Since the regime variable is unobserved, constructing the likelihood function is a non-trivial exercise and the Appendix contains a self-contained, detailed description. Briefly, the joint likelihood of the data is constructed

by multiplying the likelihood of observing a particular seven-dimensional interest rate vector conditional on past information and the current regime by the conditional probability of observing that particular regime vector and summing over all possible regime realizations S_t .

Table 5 provides the parameter estimates and robust standard errors from the maximum likelihood estimation. These parameter estimates capture a number of appealing features. If we abstract from regime switching, the law of motion for regime i implies an unconditional mean short rate of $\mu_i/(1-\beta_i)$. As discussed in section 3, above, we constrain $\mu_i/(1-\beta_i)$ to be increasing in i . (The values of these regime-specific means for regimes 1, 2, and 3 are 4.11%, 4.16%, and 14.01%, respectively.) However, we did not constrain the regimes to be increasing as well in mean reversion and in variance. None the less, this is clearly the case. Regime 1 is a near random walk ($\beta_1 = .9991$), while the regime-specific autocorrelation for regime 3 (β_3) is around 0.8. This is consistent with Conley, Hansen, Luttmer, and Scheinkman's (1994) estimates of a diffusion model for the U.S. short interest rate. These authors find that the short interest rate exhibits nearly zero mean reversion except when the short rate is either very high or very low, in which cases they find substantial mean reversion. In our model, the low mean regime is associated with extreme, near random walk persistence which may be indicative of the interest rate smoothing efforts of the monetary authorities.¹⁵ Nevertheless, our estimated interest rate process is stationary. Our estimate of b_{23} is positive, so persistent high interest rates eventually lead to a switch into the high mean regime, which exhibits substantial mean reversion. (In fact, in a simulation of 100,000 draws of our estimated regime-switching model, the maximum interest rate observed is only 23.5%. In an analogous experiment with a unit-root process, the maximum interest rate is typically above 200%.) It is not surprising that the high-inflation regime is associated with greater mean reversion in the interest-rate

¹⁵ The high persistence in short rates has been noted by other authors. Rudebusch (1995) models the weekly Fed funds rate as a unit-root process. Mankiw and Miron (1986) and McCallum (1994) argue that this high persistence is due to interest-rate smoothing by monetary authorities.

process, since the monetary authority must now eschew smoothing of interest rates in favor of fighting inflation.

The conditional volatility of interest rates also is increasing in the regime. There are two parameters that control the conditional standard deviation of interest rate innovations within a regime, σ_i and γ_i . While the estimates of γ_i decrease as the regime number increases, which would reduce the importance of conditional volatility for a fixed value of r_t , the mean value of r_t itself is increasing in the regime. Furthermore, the estimates of σ_i are higher for higher regimes. Taking all of these effects together, the net result is for the conditional standard deviation to be strongly increasing as one moves to higher-rate regimes. In particular, when we compute the mean value of $h_i(r_t) \equiv \sigma_i r_t^{\gamma_i}$ conditional on regime i prevailing at date t , we find that $h_1(r_t) = 0.08$, $h_2(r_t) = 0.42$, and $h_3(r_t) = 1.59$. Notice also that γ_1 is close to 0.5, the "square root" process assumed by Cox, Ingersoll, and Ross (1985). Ait-Sahalia (1996) non-parametrically estimates the diffusion function in a continuous-time model of U.S. interest rates, and finds that, for low-to-moderate interest rates, the square root specification is a close approximation to the estimated non-parametric diffusion function. Hence, we confirm that result, but we find that the square root specification deteriorates at high interest rates when we allow the σ_i to increase.

One somewhat disappointing aspect of the estimation involves the statistical significance of the dependence of the transition probabilities on the levels of the interest rates. Here, only b_{33} is significantly different from zero. This parameter, though, indicates that if the current regime is regime 3, the probability of remaining in this regime is increasing in r_t . Also, the positive point estimate of b_{23} indicates that if the economy is currently in regime 2, the probability of switching into regime 3, the higher mean, higher volatility regime is increasing in r_t . The positive point estimate of b_{22} also indicates that if the economy is currently in regime 2, the probability of staying in this regime

is increasing in r_t . These latter two features imply that the probability of switching from regime 2 to regime 1 is decreasing in the interest rate.

Table 6 provides some diagnostics on the three regime model. The Table reports minimum, mean and maximum values of the conditional standard deviations of interest rate innovations within each regime and for each of the seven countries. The values across countries are reasonably uniform as would be expected if the same model is appropriate.

The unconditional probabilities of the three regimes are 29.5% for regime 1, 58.9% for regime 2 and 11.6% for regime 3. (These probabilities were determined by simulating the model with 200,000 time periods.) One sense in which there would be a peso problem in a particular country is if the estimated fraction of time spent in the various regimes differs from these unconditional probabilities. Table 6 reports our estimates of the fraction of time spent in each regime for each of the seven countries.¹⁶ Our parameter estimates indicate that Germany and the United States spent too much time in regime 2 and too little time in regime 3, compared to the population probabilities of these regimes. This fact suggests that peso problems may be able to explain the differences across the countries that were documented above, although the peso problem does not appear severe. Of course, the number of switches between regimes in a given time interval is another dimension of the model that could indicate peso problems. For each country, Table 6 reports the number of switches between regimes 1 and 2, and between 2 and 3. Here the U.S. looks somewhat different than the U.K. and Germany. Ultimately, only simulations of the expectations hypothesis regression evidence using the estimated regime switching model as the data generating process can determine whether peso problems can account for the differences across countries.

¹⁶ The regime is, of course, unobservable. The recursive maximum likelihood algorithm delivers an estimate of the probability (given current information) that a given country is in a particular regime at each data point. We allocate each data point to the regime that is estimated to be most likely. Regime classification in the context of switching regressions is discussed in Lee and Porter (1984).

More generally, we find that the patterns of regime switches through time vary greatly among the seven countries. Italy, Sweden, and the U.K. move in and out of the high-rate regime intermittently throughout the sample period. In contrast, Australia enters this regime around 1980, and remains there consistently throughout the decade. Not surprisingly, Germany is in the high-rate regime very infrequently, Japan's experience with this regime largely coincide with the periods following the two big oil shocks, and the U.S. is in the high-rate regime only during the 1979-82 period of monetary targeting. These results indicate that the regime switches are not highly correlated across countries, so our simplifying assumption of cross-country independence does not do substantial violence to the data.

5. Monte Carlo Explorations of the Peso Problem under the Expectations Hypothesis

5.1 Monte Carlo Methodology

Our characterization of the peso problem implies that it is fundamentally just a small sample problem. There are good reasons to believe that our regime-switching model is particularly vulnerable to small sample problems. First, the substantial differences between the means in the regimes, the high persistence within each regimes, and the non-trivial conditional heteroskedasticity may imply skewed and highly leptokurtic short rate distributions in small samples.¹⁷ Second, two of the three regimes display near-unit-root behavior for the short rate, implying that the process appears nearly non-stationary within regimes.¹⁸ Both considerations suggest that an extremely long time series is necessary for the empirical distributions to replicate the population distributions.

Third, Bekaert, Hodrick and Marshall (1997a) demonstrate that the persistence in short rates

¹⁷ Bollerslev, Chou and Kroner (1992) survey the literature on GARCH models and note the small sample problems when there is substantial conditional heteroskedasticity.

¹⁸ Elliott and Stock (1994) discuss inference when there is near unit root behavior in a regressor. The high persistence of the short rate and the Markov regime structure for the short rate imply that the spread in our model is highly serially correlated as it is in the actual data.

induces extreme bias and dispersion in the small sample distributions of the Campbell-Shiller (1991) tests. Briefly, the well-known downward bias in estimating autocorrelations (Kendall (1954)) translates into a large upward bias in the slope coefficients of standard expectation hypothesis tests. On the one hand, the upward bias makes it *ex ante* more difficult to explain the negative slope coefficients found in estimating equation (7). On the other hand, the increased dispersion may be large enough so that the left tail of the small sample distribution encompasses the observed negative slope coefficients.

To evaluate whether small sample problems can explain the data, we must compute the long-maturity yields implied by the expectations hypothesis in the context of our regime-switching model. To do so, we must compute expected future short rates. Since the model is highly non-linear, we compute the expected future short rates using a Markov chain approximation to the estimated regime switching model. There are two state variables: the short interest rate and the regime. We construct a modified equi-spaced grid on the space of possible realizations of the short rate as follows: First, we compute the minimum and the maximum interest rates from a 100,000 length simulation of the true non-linear process. We then treat each integer interest rate as the mid-point of a cell, and we include a sufficient number of points within each cell to insure that the step-size between possible interest rates is smaller than the conditional standard deviation of the interest rate shock in that cell. We then add additional points until the total number of points per regime equals 550 (implying 1,650 possible states). For each of these discrete states, the transition probabilities are computed by evaluating the conditional density of next period's short interest rate (a state-dependent mixture-of-normals) and then normalizing these transition probabilities to sum to unity. The resulting Markov chain approximation to the law of motion given by equations (12) -(16) is highly accurate. In particular, when we simulate a sample of 100,000 observations using the Markov chain approximation, and re-estimate the model using this simulated data, all point estimates are within two standard errors

of the estimates given in Table 5. In fact, the only three parameters not within one standard error of the Table 5 value are a_{23} , b_{23} , and b_{33} , which govern the transition probabilities between regime 2 and regime 3.

We then derive the small sample distributions of the slope coefficients in the Campbell-Shiller (1991) regressions and the VAR statistics under the assumption that the short rates are generated by the estimated regime-switching model. Specifically, we simulate the estimated model of the three-month short rate to create an artificial time series of 297 months (the length of our sample). We then compute the long yields for 12, 36, and 60 month bonds implied by the expectations hypothesis, given the estimated law of motion of our model and compute the statistics reported in Tables 1 and 3. This exercise is performed 25,000 times to construct a distribution of the estimators for each of the statistics.

5.2 Monte Carlo results

A summary of the Monte Carlo exercise is reported in Table 7. As in Bekaert, Hodrick and Marshall (1997a), we find extreme positive bias in all the OLS slope estimators (as well as their implied counterparts from the VAR methodology). The high mean in the estimator of equation (7) is due in part to the use of $r(t+3,n)$ as a proxy for the unobserved $r(t+3,n-3)$. As a result, the probability limit of this estimator exceeds unity. (See Bekaert, Hodrick and Marshall (1997a) for a discussion of this point.) However, the population values for the slope coefficients of equation (8) are unity, and these display severe upward bias. While these upward biases make it more difficult to reconcile the expectations hypothesis with the empirical estimates reported in Table 1, the large standard deviation of these slope estimators reported in the fourth column of Table 7 makes it more difficult to reject the expectations hypothesis. These biases and estimator variances are more extreme than those reported in Bekaert, Hodrick and Marshall (1997a), which uses an alternative data generating process for the short rate and the spreads between long and short rates.

In order to re-consider the evidence on the expectations hypothesis across countries, we report in Table 7 the .5%, 2.5% and 5% quantiles of the empirical distributions. These quantiles correspond to the relevant critical values for two-sided tests with size 1%, 5% and 10% respectively.¹⁹ When the estimates in Tables 1 and 3 are evaluated using the small sample distributions in Table 7, the evidence against the expectations hypothesis appears weaker than when standard asymptotic inference is used. First, when the Campbell-Shiller regressions (7) and (8) are used to test the expectations hypothesis, the only rejection at the 1% level is for regression (7) at the 60 months horizon for the USD data. Although substantially negative values occur in the simulations, especially at shorter horizons, the severe positive bias in the estimators makes negative values very unlikely for longer horizons. When regression (8) is used, none of the estimates now reject the expectations hypothesis at the 1% level, and in only one case (36 month maturity with USD data) can we reject at the 5% level. The small sample distributions of the slope coefficients for regressions (7) and (8) implied by the VAR are very close to the distributions for the OLS estimators, indicating that a second order VAR is a good approximation to the data generating process.

Let us turn to the VAR-based correlation statistics and standard deviation ratio statistics, reported in Table 3. According to the small sample distributions summarized in Table 7, the correlation statistic does not reject the expectations hypothesis at conventional significance levels for any horizons in the GBP or DEM data. This statistic does reject the hypothesis in USD data at the 5% level for the 36-month horizon, and at the 1% level for the 60-month horizon. The standard deviation ratio statistic rejects the expectations hypothesis at the 1% level for the 12-month horizon with USD and DEM data, but does not reject at the 5% level with any of the three data sets for the longer horizons. Taken together, these results provide some evidence against the expectations

¹⁹ The small sample distributions are actually quite skewed which complicates the choice of the size of the test.

hypothesis, but this evidence is less uniformly unfavorable to this hypothesis than would be concluded using asymptotic distribution theory.

5.3 Interpretation of the results

Our data-generating process has several potential sources of small sample bias: persistence, conditional heteroskedasticity, skewness and kurtosis, as well as peso problems. The type of peso problem we focus on is where the high mean, high variance regime is under-represented in the sample relative to the population.²⁰ To determine which of these factors accounts for the biases and the enormous dispersion in the small sample distributions of Table 7, we conduct a series of Monte Carlo experiments using simpler regime-switching models as data generating processes. In the simplest of these (denoted "Model A"), the within-regime process for the short rate is Gaussian white noise, with identical disturbance variances in each regime. In Model A, the only thing that distinguishes the three regimes is the within-regime mean of the interest rate process. Furthermore, Model A has constant regime-switching probabilities. Model B is identical to Model A except that the regime-switching probabilities are state-dependent. Model C is an extension of Model B in which within-regime variances differ across regimes, and within-regime conditional heteroskedasticity is introduced. Finally, in Model D, we go back to the set-up of Model A, except we allow for the estimated within-regime persistence. To keep the various regime-switching models as close as possible to the full model, we alter the coefficients in a minimal way while ensuring that the process matches the population mean, standard deviation and frequencies of the three regimes for the short rate. The exact parametric specifications we use for the four simple models are given in Table 8. That table also reports the mean and standard deviation of the small sample distributions of the slope coefficients of regressions (7) and (8) for the four models.

²⁰ There are many other types of "peso problems". For example, one could also consider the number of switches from regime 2 to 3 (see Table 5). However, the frequency characterization we use is highly positively correlated with the "switches" definition.

It is clear from Table 8 that the main source of both bias and dispersion is the persistence of the short rate process. This is consistent with the results of Bekaert, Hodrick and Marshall (1997a), who use a very different data-generating process. More specifically, when the within-regime persistence is eliminated (in Models A, B and C), both the bias and dispersion decrease substantially relative to the full model. In the absence of within-regime persistence, adding state-dependent transition probabilities (Model B) or conditional heteroskedasticity (Model C) has little effect on the small sample means and standard deviations of the slope estimates. However, when state-dependent transition probabilities and conditional heteroskedasticity are added to a model that already incorporates the extreme persistence found in the data, the effect is dramatic. This can be seen by comparing Model D to the full model. For equation (7), the standard deviation of the slope coefficient estimator is doubled when we move from Model D to the full model, and, for equation (8), the corresponding standard deviation is increased three-fold or more. The small sample biases are also exacerbated substantially for the longer horizons.

Although the peso effect is less pronounced than the persistence effects, it may still be important to obtain negative coefficients in the first Campbell-Shiller regression. In Figure 1, we order the 25,000 slope coefficients of our experiments in bins of 1,000 ranked according to size and graph the average slope coefficients (for the 12 and 60 month maturities) over these bins relative to the average frequency of regime 3 over the relevant 1,000 samples. This figure illustrates the importance of peso effects in generating the tails of the small sample distributions. Negative slope coefficients are associated with a lower than average frequency of regime 3. Interestingly, peso effects of the kind that we thought would generate negative coefficients, also seem responsible for highly positive slope coefficients. (This is especially apparent in the second panel of Figure 1, which displays results for the 60-month maturity.) That is, they contribute to the dispersion of the small sample distributions in both tails.

Why are samples with low frequency of regime three sometimes associated with highly positive slope coefficients in the first Campbell-Shiller regression? First, when the third regime is under-represented in the sample, short rates are likely to be more persistent (since the third regime displays less persistence than the other two regimes). This increased persistence exacerbates the upward bias in small sample estimators of this slope coefficient. Second, peso effects are associated with an unusually low frequency of the third regime only if there are a substantial number of observations where short rate shocks significantly drive up the term spread. If these increases in the spread do not occur or they are not of sufficient magnitude, we are simply left with a sample of very persistent short rates and low-variance term spreads. Finally, the derivative of the regime transition probabilities with respect to the short rate is increasing in the level of the short rate. At low interest rates (which are more likely when the high-rate regime occurs infrequently), interest rate shocks are therefore less likely to generate large changes in term spreads.

This last point is related to a more general problem with the model of this section. A rise in the short rate has two effects on the current spread. The direct effect is to narrow the spread since the short rate enters with a negative sign. The indirect effect is to increase the long-rate through an increase in expected future short rates. Peso problems, if present, are generated through this indirect effect. This effect depends on the sensitivity of the transition probability to short rate changes, and is likely to dissipate for longer horizons. This explains why, in Table 7, longer maturity is associated with greater upward bias and fewer negative observations for the first Campbell-Shiller slope coefficients. Unfortunately, this pattern is opposite to that found in U.S. data (see Table 1), where longer maturities are associated with more negative slope coefficients.

One potential solution to this problem is to assume that agents observe a variable that affects the transition probability but is imperfectly correlated with the current short rate. This extension of the model raises too many technical complexities to be explored within the context of the current

paper. A second possibility is to move away from the strict expectations hypothesis, and to allow time-variation (in particular, regime-dependence) in term premiums. We explore this possibility in the next section.

6. Regime Switching and Term Premiums

6.1 Model and Calibration

To investigate the impact of regime-dependent term premiums, we present here an alternative model, based on the discrete-time affine class (see Backus, Foresi, Mozumdar and Wu (1997) and Campbell, Lo, and MacKinlay (1997)). We consider the following law of motion for M_t , the pricing kernel for nominal assets. It is assumed that $m_t \equiv \log(M_t)$ satisfies:

$$m_{t+1} = - \left[1 + \frac{\lambda^2}{2} \right] r_t - \lambda \sqrt{r_t} \epsilon_{t+1} \quad (17)$$

where λ is a parameter and $\{\epsilon_t\}$ is a sequence of independent standard normal random variables. In order to avoid introducing additional sources of noise into the model, we assume that the disturbance process ϵ_t in equation (17) is identical to the disturbance to r_t in equation (12). Note that $\exp(-r_t) = E[\exp(m_{t+1})]$, so the $\{m_t\}$ process correctly prices the short asset.

If there were but a single regime, r_t would be a Gaussian process, and the model given by equations (17) and (12) would fall into the discrete-time affine class. In the multi-regime case, r_t is not Gaussian, so this model does not inherit the analytic tractability of an affine model. None the less, its implications for long-term bond yields can be computed using our discrete Markov chain approximation to the short rate process given by equations (12) - (16). According to equation (1),

$$\begin{aligned} \exp[-r(t,n)n] &= E_t \left[\prod_{i=1}^n \exp[m_{t,i}] \right], \\ &= E_t \left\{ \exp(m_{t,1}) \cdot E_{t,1} \left[\exp(m_{t,2}) \cdot E_{t,2} \left(\exp(m_{t,3}) \cdot \dots \cdot E_{t,n-1} \left[\exp(m_{t,n}) \right] \cdot \dots \right) \right] \right\} \end{aligned} \quad (18)$$

The right-hand side of equation (18) is straightforward to evaluate. In the discrete state-space approximation, $E_{t,i}[\exp(m_{t,i+1})]$ can only take a finite number of values (one for each state). Once these values have been computed, the right-hand side of equation (18) can be evaluated by recursively applying the state transition matrix. The only additional parameter in this model is λ which determines the unconditional variance of m_t . We choose λ to match the relative volatilities of the short and long yields in our data. In particular, the average value across the U.S., U.K., and Germany of the ratio between the standard deviation of the three-month yield and the standard deviation of the five-year yield is 1.55. With our estimated r_t process, this value is matched by model (17) by setting $\lambda = 2.1$.

There is an additional technical issue to be addressed. The timing interval in equation (17) is one month. (Note that ϵ_t in equation (12) is a monthly process.) Therefore, the m_t process satisfying equation (17) is consistent with r_t if r_t is a one-month rate. However, for reasons of data availability, we estimate equation (12) using data for three-month interest rates. To accommodate this inconsistency between the timing interval in the model and that in the data, we treat the discretized r_t process generated by our estimates of model (12) - (16) as if this were the process for the one-month rate. In practice, the distortions thus induced are small, since one-month rates track three-month rates very closely in all countries for which we have data on both maturities. We then compute the (discretized) one-month nominal pricing kernel m_t according to equation (17), and generate (discretized) n -period rates according to equation (18) for n ranging from three months to five years. In our subsequent Monte Carlo analysis of equations (7) and (8), we use the three-month

rate generated by equation (18) as the short rate. This insures that all rates are generated in an internally consistent arbitrage-free manner. It also makes the results of this model comparable to those reported in section 5, above, since the three-month rate was used as the short rate in that analysis. Furthermore, we conduct an experiment where we use the three month short rate from the pricing kernel model to repeat the Monte Carlo experiment under the null of the expectations hypothesis, and our results are similar to the results reported in Table 7.²¹

6.2 Empirical results

General equilibrium models, and affine models in particular, have not been successful at generating population distributions for the term premiums that resolve the Campbell-Shiller puzzles. Our model, given by equations (12) - (17), has similar problems. The second and third columns of Table 9 show that our model does not deliver slope coefficients for equations (7) and (8) that are substantially below one in population. Table 9 also reports the population moments of the term premiums, the difference between the actual long-rate and its value predicted by the expectations hypothesis with term premiums set to zero. The term premium means are positive and increase with maturity, and the term premium is positively correlated with the level of the short rate. Note, however, that the standard deviation of the term premiums is rather small (varying between 26 basis points at 12 months to 62 basis points at 60 months), so this model represents a quantitatively small departure from the expectations hypothesis. Nevertheless, the presence of term premiums may exacerbate the small sample problems. One indication is a test we run to check our programs. We simulate very long samples from the Markov chain recording the average slope coefficients of the Campbell-Shiller regressions over three runs. The convergence to the population values is now much slower than before, requiring 1,000,000 observations (over 800 centuries of data).

Table 10 reports the results from a Monte Carlo experiment mimicking the experiment

²¹ Detailed results from this exercise are available upon request.

underlying the results in Table 7 for the new data generating process. The results are dramatic. The dispersion of the small sample distribution increases for all test statistics, but the increase is most dramatic for the longer maturities and the first Campbell-Shiller regression. The distributions also exhibit dramatic negative skewness, often exhibiting positive medians but negative means. The estimates of equation (7) and (8) for the U.S. data, displayed in Table 1, are not even below the 5% quantile of the empirical distribution. In particular, the distribution can accommodate the increasingly negative slope estimates for equation (7) as the maturity increases. The VAR continues to describe the dynamics of the data generating process rather well, with the implied Campbell-Shiller slope coefficients having very similar small sample distributions compared to their direct OLS counterparts.

The only statistic in the data that is not fully consistent with the model is the standard deviation ratio. The variance of the theoretical spread relative to the actual spread is too low in the data for $n=12$ (below the 0.5% quantile) and $n = 36$ (below the 2.5% quantile). The high mean of the empirical distribution of this statistic is not only caused by the upward bias we documented before, but also because the theoretical spread $s'(t,n)$, defined in equation (14), is more variable than the actual spread in population. The reason for this is as follows. As shown in Table 9, the term premium $c(t,n)$ is positively correlated with the level of the short rate. The short rate, in turn, is

negatively correlated with $s'(t,n) \equiv (1/n) \sum_{i=0}^{n-1} E_t r_{t+i} - r_t$, since r_t enters $s'(t,n)$ with a negative sign.

Together, these facts imply that the term premium is negatively correlated with $s'(t,n)$. Noting that the actual spread $s(t,n) \equiv s'(t,n) + c(t,n)$, the negative correlation between $s'(t,n)$ and $c(t,n)$ tends to reduce the variability of $s(t,n)$, relative to $s'(t,n)$.

How can the introduction of a rather small term premium generate such dramatic effects?

The reason is that the term premium considerably exacerbates the peso effects we described before. As before, a positive shock to interest rates has a direct negative effect on the spread, but the indirect positive effect through expected higher future short rates is now amplified by the positive term premiums. Hence, some samples may indeed experience dramatic peso effects, where large increases in spreads are not followed by corresponding increases in short rates. Given that the average term premium increases from 35 to 168 basis points (and becomes more variable), it is no surprise to see the most dramatic effects for long maturities.

To illustrate this effect, we perform an exercise for the model of this section analogous to that displayed in Figure 1. The results, displayed in Figure 2, show that extreme negative values of the estimated slope coefficient for equation (7) are unambiguously associated with those samples that have few realizations of the high-rate regime. Indeed, for the 12 month maturity, the relationship between the frequency of regime 3 and the magnitude of the estimated slope is virtually monotonic. For the 60 month maturity, the relationship is non-monotonic, with the humped shape that somewhat resembles Figure 1. However, the samples in which the frequency of regime 3 is below its population probability are unambiguously associated with extremely negative estimates for the slope coefficient.

Conclusions

Our paper contributes to the term structure literature in a number of ways. First, we review the evidence on the expectations hypothesis using the standard Campbell-Shiller (1991) regressions and a partially new VAR methodology for three countries: the U.S., the U.K., and Germany. The evidence is anomalous in that the expectations hypothesis is most strongly rejected for U.S. data but holds rather well for U.K. data.

Second, we argue that these anomalies may be due to a generalized peso problem in which a high-interest rate regime occurred less frequently in the sample of U.S. data than was rationally anticipated. We formalize this idea as a regime-switching model of short-term interest rates estimated

with data from seven countries. Technically, this model extends recent research on regime-switching models with state-dependent transitions to a cross-sectional setting. The regime-switching model reveals the existence of a high mean, high variance regime in which short rates are much more mean reverting than in the two "normal" regimes.

Third, when we use the small sample distributions generated by the regime-switching model to conduct inference, the evidence against the expectations hypothesis weakens considerably. Notable features of the distributions are a substantial upward bias and much larger dispersion than the asymptotic distributions. Whereas Bekaert, Hodrick and Marshall (1997a) show that such features are also present when the data generating process does not involve "peso effects," we demonstrate that peso effects contribute to the dispersion of the distribution, although the persistence of the short rate remains the most important determinant. Nevertheless, it remains somewhat implausible that our data-generating process generated the U.S. data.

Fourth, to reconcile the data for all countries with our model, we must allow for a small time-varying term premium. When we allow regime changes and other rate movements to be priced, the small sample distributions become much more dispersed, skewed and biased in the direction of explaining the results.

While we believe our results to be interesting, they also raise many interesting questions for further research. The effects of combining peso effects with time-varying risk premiums are dramatic. While our model developed here is only illustrative, we believe it deserves further exploration. For example, future work should attempt to allow for correlations across the short rate innovations of the various countries and use term structure data to identify the parameters of the regime-switching model. Introducing variables not perfectly correlated to the short rate (for example term spreads) in the transition probability also constitutes an interesting extension, but one that raises many technical issues.

Finally, the sensitivity of the Campbell-Shiller regressions to small sample problems raises the econometric issue of developing estimators less prone to severe small sample biases. This is all the more pressing since many important financial theories are tested with similar regressions. A good example is the Unbiasedness Hypothesis tests in foreign exchange, which suffer from similar problems (see Baillie and Bollerslev (1997)).

Appendix

A.1 Derivation of Equations (4), (5), and (6)

Equations (1) and (3) imply

$$r(t,n) = -\frac{1}{n} \log \left(E_t \exp(m_{t+n}^n) \right) \quad (19)$$

To evaluate the conditional expectation on the right-hand side of equation (19), take a Taylor-series expansion of $\exp(m_{t+n}^n)$ about $E_t m_{t+n}^n$:

$$\exp(m_{t+n}^n) = \exp(E_t m_{t+n}^n) \left[1 + \sum_{j=1}^{\infty} \frac{(m_{t+n}^n - E_t m_{t+n}^n)^j}{j!} \right], \quad (20)$$

Under the assumption that all conditional moments of m_{t+n}^n exist, we can apply the conditional expectation operator to equation (20), obtaining

$$E_t \exp(m_{t+n}^n) = \exp(E_t m_{t+n}^n) \left[1 + \sum_{j=2}^{\infty} \frac{v_t^n(j)}{j!} \right], \quad (21)$$

where $v_t^n(j) \equiv E_t (m_{t+n}^n - E_t m_{t+n}^n)^j$, the j^{th} conditional central moment of m_{t+n}^n conditional on date t information. Equations (19) and (21) imply

$$r(t,n) = -\frac{1}{n} E_t m_{t+n}^n - \frac{1}{n} \log \left[1 + \sum_{j=2}^{\infty} \frac{v_t^n(j)}{j!} \right]. \quad (22)$$

Applying the law of iterated expectations to equation (3), one obtains

$$E_t m_{t+n}^n = E_t \sum_{i=1}^n E_{t+i-1} m_{t+i} \quad (23)$$

Evaluating equation (23) at $n = 1$ for dates $t+i-1$, $i = 1, \dots, n$, implies

$$E_{t+i-t} m_{t+i} = -\tau(t+i-1, 1) - \log \left[1 + \sum_{j=2}^{\infty} \frac{v_{t+i-t}^1(j)}{j!} \right]. \quad (24)$$

To obtain equations (4) and (5), one first must substitute equation (24) into the right-hand side of equation (23), substitute the resulting expression into the first term on the right-hand side of equation (22), and combine terms.

We now derive equation (6) from equation (5), under the assumption that, conditional on date t information, m_{t+i} is normally distributed $\forall i \geq 1$. Using definition equation (3) of m_{t+n}^n , equation (6) can be written

$$c(t, n) \equiv \frac{1}{n} \left[\sum_{i=1}^n \frac{1}{2} E_t(\text{var}_{t+i-t}(m_{t+i}) - \frac{1}{2} \text{var}_t(m_{t+n}^n)) \right], \quad (25)$$

so equation (6) is equivalent to equation (5) if

$$\frac{1}{2} \text{var}_t(m_{t+n}^n) = \log \left[1 + \sum_{j=2}^{\infty} \frac{E_t(m_{t+n}^n - E_t m_{t+n}^n)^j}{j!} \right], \quad \forall n. \quad (26)$$

In fact, equation (26) holds under the conditional normality assumption. Let \tilde{x} be a normal random variable with mean μ and variance σ^2 , and let $\tilde{y} = \tilde{x} - \mu$. The variable \tilde{y} is normal, with moment generating function $M_{\tilde{y}}(t) = \exp\left[\frac{\sigma^2 t^2}{2}\right]$. Taking a MacLaurin Series expansion of $M_{\tilde{y}}(t)$, evaluated

at $t=1$, one obtains

$$\exp\left(\frac{\sigma^2}{2}\right) = M_{\tilde{y}}(1) = M_{\tilde{y}}(0) + \sum_{j=1}^{\infty} \frac{M_{\tilde{y}}^{(j)}(0)}{j!} = 1 + \sum_{j=2}^{\infty} \frac{E(\tilde{x}-\mu)^j}{j!} \quad (27)$$

where, in equation (27), $M_{\tilde{y}}^{(j)}(t)$ denotes the j^{th} derivative of $M_{\tilde{y}}$, evaluated at t , and we use $M_{\tilde{y}}(0)$

$= 1$ and $M_y^{(1)}(0) = 0$. Equation (27) implies equation (26) if we set $\bar{x} = m_{t+\tau}^n$ and treat the distribution in equation (27) as conditional on date t information.

A.2 Data Description

Our data set for USD, GBP and DEM zero-coupon bond yields is an updated version of the data originally used by Jorion and Mishkin (1991). We thank Philippe Jorion for generously providing us with the data. The Jorion-Mishkin data consist of monthly observations from 1972:01 through 1991:12 on implied zero-coupon yields with maturities of 3, 12, 24, 36, 48, and 60 months which are constructed from observations on outstanding government bonds. Data from 1990:1 to 1996:9 on zero-coupon bond yields with maturities of 3, 12, 36, and 60 months for the three currencies were obtained from a New York investment bank that wishes to remain anonymous.

The short rates used to estimate the regime-switching model are from a variety of sources. For the U.S., we use the three-month Treasury Bill data. For Germany, we use the three-month interbank rate from the Bank for International Settlements (BIS) database.²² For the U.K., we use three-month Treasury Bill data from the BIS.²³ The Australian short rate is the three month Treasury bill rate taken from Datastream from September 1972 onwards; for the first 8 months of 1972 we use the commercial paper rate from the BIS. The Italian short rate is the three month interbank rate from International Financial Statistics (IFS). The Japanese short rate is the three month Gensaki rate from the BIS. The Swedish short rate is the three month Treasury Bill rate from the IFS. Two outliers (at 1983:4 and 1983:5) were confirmed to be data errors by looking at interest

²² A missing observation at 1995:12 in this data for Germany was replaced by a linear interpolation between observations from 1995:11 to 1996:01.

²³ A missing observation at 1995:12 in this data for the U.K. was filled in with the observation for that date from the IFS database.

series reported in the Sveriges Riksbank Quarterly Review (1983-1984) and were replaced by Eurocurrency rates drawn from Datastream.

A.3 Joint Estimation of the VARs for Three Countries using GMM

We simultaneously estimate four-variable VARs for each country (U.S., U.K., and Germany) under the restriction that the VAR coefficients for each country are the same. Each VAR incorporates the 3-month rates and the spreads between three long rates (12-months, 36-months, and 60-months) and the 3-month rate. The estimation is to be done using GMM. This is a large GMM system since it requires 138 orthogonality conditions and 46 free parameters. The free parameters are the four constants, the thirty-two slope coefficients, and the ten elements of the innovation covariance matrix, which are constrained to be the same across the three currencies. The orthogonality conditions are the same as those imposed in the currency-by-currency estimation. Each of the four error terms for each currency is postulated to be orthogonal to the nine right-hand side variables associated with that currency, which provides 36 orthogonality conditions per currency. The remaining 30 orthogonality conditions are that the 10 parameters of the innovation covariance matrix should equal the usual cross-products of the innovations of the equations, again, for each of the three currencies.

Given the size of this GMM system relative to the number of available time-series observations, we chose to estimate the weighting matrix using the residuals from the unconstrained estimations, which are consistent estimates of the true residuals. We also chose to restrict the elements of the GMM weighting matrix using the assumptions that the cross-covariances of the innovations and squared innovations across currencies are zero. With these assumptions the weighting-matrix is block diagonal with three 36x36 blocks and three 10x10 blocks for each currency.

A.4 The Estimation of Cross-Sectional Regime Switching Models by Maximum Likelihood

In this appendix, we describe the construction of the likelihood function of the cross-sectional regime switching model in section 3. The model is basically a cross-sectional extension of the regime-switching model of interest rates proposed by Gray (1996), and we construct the likelihood function building on the framework in Gray (1995).

Let us first establish some notation. There are $N = 7$ countries, superscripted by j , and $K = 3$ regimes, sub-scripted by i . Here, the interest rate in country j will be denoted by r_t^j . Let $R_t = [r_t^1, \dots, r_t^N]'$, the cross-sectional collection of short rates at time t , and let $R_t^j = [r_t^1, \dots, r_t^{j-1}, r_t^{j+1}, \dots, r_t^N]'$, the cross-section of interest rates excluding country j . Let $\tilde{r}_t^j = [r_t^j, r_{t-1}^j, \dots, r_0^j]'$, the time-series vector of short rates for each country, and let $\tilde{R}_t = [\tilde{r}_t^1, \dots, \tilde{r}_t^N]'$, the cross-sectional collection of these short rate time series. The regime for country j 's rate will be denoted by s_t^j , with $s_t^j \in \{1, 2, \dots, K\} \forall j$. We will also use $s_{t,i}^j$ to denote $s_t^j = i(j)$ with $i(j) \in \{1, \dots, K\}, \forall j$. Since each of the N countries can be in K possible regimes, there are K^N possible values that we denote $S_{t,i} = [s_{t,i(1)}^1, \dots, s_{t,i(N)}^N]$. We will also define

$$S_{t,i}^j = [s_{t,i(1)}^1, \dots, s_{t,i(j-1)}^{j-1}, s_{t,i(j+1)}^{j+1}, \dots, s_{t,i(N)}^N]'. \quad (28)$$

The model for the short rate r_{t+1}^j is given by equations (12) and (19). For each country, s_t^j follows a Markov process with time-varying transition probabilities given in equations (20) through (16). Although s_t^j will evolve differently over time across countries, the parameters governing its stochastic process and the other parameters determining the short rate process are assumed to be the same for all countries. Without further parameter restrictions, there $4(2K-1)$ parameters to estimate:

$\{\mu_i, \beta_i, \sigma_i, \gamma_i, a_{ij}, b_{ij}\} i = 1, \dots, K$, and $\{a_{ij}, b_{ij}\} i = 2, \dots, K-1, j = i+1$.

Let the joint log-likelihood function for a sample of T observations be

$$l(\vec{R}_T) = \log[L(\vec{R}_T)]. \quad (29)$$

By conditioning on the initial observation, we can write:

$$L(\vec{R}_T) = \Pi_{t=1}^T L(R_t | \vec{R}_{t-1}). \quad (30)$$

We assume that agents observe the regime, but the econometrician does not. As in Hamilton (1989) and Gray (1995), the econometrician must condition on $s_{tj}^j, \forall j$, to evaluate $L(R_t | \vec{R}_{t-1})$.

Because there are K^N possible values for the states, we can write the conditional likelihood function as follows:

$$\begin{aligned} L(R_t | \vec{R}_{t-1}) &= \sum_{s_{t-1}}^{K^N} L(R_t, S_{t-1} | \vec{R}_{t-1}) \\ &= \sum_{s_{t-1}}^{K^N} L(R_t | S_{t-1}, \vec{R}_{t-1}) \cdot P(S_{t-1} | \vec{R}_{t-1}). \end{aligned} \quad (31)$$

We evaluate the two parts in equation (31) in turn. To evaluate the first part, we assume that the interest-rate innovations are conditionally independent across countries.

Assumption 1:

$$L(r_t^j | R_{t-1}^j, \vec{R}_{t-1}, S_{t-1}) = L(r_t^j | s_{t-1}^j, \vec{R}_{t-1}). \quad (32)$$

Assumption 1 implies

$$L(R_t | S_{t-1}, \vec{R}_{t-1}) = \Pi_{j=1}^N L(r_t^j | s_{t-1}^j, \vec{R}_{t-1}). \quad (33)$$

For the model described in equations (12)-(19), equation (33) represents a product of N univariate normal densities:

$$L(r_t^j | s_t^j, \bar{R}_{t-1}) = \frac{1}{\sqrt{2\pi} \sigma_i(r_{t-1}^j)^{\gamma_i}} \exp \left[\frac{(r_t^j - \mu_i - \beta r_{t-1}^j)^2}{2\sigma_i^2(r_{t-1}^j)^{2\gamma_i}} \right]. \quad (34)$$

To evaluate the second part, we attempt to write $P(S_{t,i} | \bar{R}_{t-1})$ as a recursion as in Gray's (1995) univariate model. To do so, we assume that regime variables s_t^j are conditionally independent across countries. This embeds two assumptions.

Assumption 2:

$$P[s_t^j = k | S_{t,i}^j, \bar{R}_{t-1}] = P[s_t^j = k | \bar{R}_{t-1}]. \quad (35)$$

Assumption 3:

$$P[s_t^j = k | R_{t,i}^j, \bar{R}_{t-1}] = P[s_t^j = k | \bar{R}_{t-1}]. \quad (36)$$

Using Assumption 2, it follows:

$$P(S_{t,i} | \bar{R}_{t-1}) = \Pi_{j=1}^N P[s_t^j = i(j) | \bar{R}_{t-1}]. \quad (37)$$

Now, we introduce some simplifying notation for the transition probabilities. Let

$$P(s_{t+1}^j = k | s_t^j = l, r_t^j) = P_{t,l,k}^j \quad (38)$$

We now write $P[s_t^j = i(j) | \bar{R}_{t-1}]$ as a recursion:

$$P[s_t^j = i(j) | \bar{R}_{t-1}] = \sum_{k=1}^K P[s_t^j = i(j) | s_{t-1}^j = k, \bar{R}_{t-1}] P[s_{t-1}^j = k | \bar{R}_{t-1}]. \quad (39)$$

From the model in equations (12)-(16), we can write $P[s_t^j = i(j) | s_{t-1}^j = k, \bar{R}_{t-1}] = P_{t,k,i}^j$. One can apply

Bayes's rule to $P[s_{t-1}^j = k | \bar{R}_{t-1}] = P[s_{t-1}^j = k | r_{t-1}^j, \bar{R}_{t-1}^j]$, to obtain

$$P(s_{t-1}^j = k | \tilde{\mathbf{R}}_{t-1}) = \frac{L(r_{t-1}^j | s_{t-1}^j = k, \mathbf{R}_{t-2}^j, \tilde{\mathbf{R}}_{t-2}) P(s_{t-1}^j = k | \mathbf{R}_{t-2}^j, \tilde{\mathbf{R}}_{t-2})}{\sum_{m=1}^K P(s_{t-1}^j = m | \mathbf{R}_{t-2}^j, \tilde{\mathbf{R}}_{t-2}) \cdot L(r_{t-1}^j | s_{t-1}^j = m, \mathbf{R}_{t-2}^j, \tilde{\mathbf{R}}_{t-2})} \quad (40)$$

where $L(\cdot | \cdot)$ represents the relevant conditional normal density, as in equation (33). Using assumptions (1) and (3) in equation (40) and substituting into (39), we obtain a recursive representation for the regime probability, which we will denote by $P_{t,\omega}^j$ for $\omega = \{1, \dots, K\}$,

$$P_{t,\omega}^j = P[s_t^j = \omega | \tilde{\mathbf{R}}_{t-1}] = \sum_{k=1}^K P_{t,\omega k}^j \frac{L(r_{t-1}^j | s_{t-1}^j = k, \tilde{\mathbf{R}}_{t-1}) \cdot P_{t-1,k}^j}{\sum_{m=1}^K P_{t-1,m}^j \cdot L(r_{t-1}^j | s_{t-1}^j = m, \tilde{\mathbf{R}}_{t-2})}, \quad (41)$$

and $L(\cdot)$ can be evaluated as in equation (33). Hence, the $P_{t,\omega}^j$ can be computed as a simple recursion, $\forall j$ and $\forall \omega$. There are a total of $N \times K$ such probabilities. Bringing the two pieces together, we can now fill out the likelihood function:

$$L(\mathbf{R}_t | \tilde{\mathbf{R}}_{t-1}) = \sum_{i=1}^{K^N} [\Pi_{j=1}^N L(r_t^j | s_t^j = i(j), \tilde{\mathbf{R}}_{t-1})] \cdot [\Pi_{j=2}^N P_{t,i(j)}^j]. \quad (42)$$

Hence,

$$l(\tilde{\mathbf{R}}_T) = \sum_{t=1}^T \ln \left\{ \sum_{i=1}^{K^N} [\Pi_{j=1}^N L(r_t^j | s_t^j = i(j), \tilde{\mathbf{R}}_{t-1})] \cdot [\Pi_{j=2}^N P_{t,i(j)}^j] \right\}. \quad (43)$$

References

- Ait-Sahalia, Y., 1996, "Testing Continuous-Time Models of the Spot Interest Rate," Review of Financial Studies 9, 385-426.
- Akaike, H., 1973, "Information Theory and the Extension of the Maximum Likelihood Principle." In 2nd International Symposium on Information Theory, B.N. Petrov and F. Csaki, eds., Budapest.
- Backus, D. K., 1993, "Cox-Ingersoll-Ross in Discrete Time," manuscript, New York University.
- Backus, D. K., S. Foresi, A. Mozumdar and L. Wu, 1997, "Predictable Changes in Yields and Forward Rates," manuscript.
- Backus, D. K., A Gregory and S. E. Zin, 1989, "Risk Premiums in the Term Structure: Evidence from Artificial Economies," Journal of Monetary Economics 24, 371-399.
- Backus, D. K., and S. E. Zin, 1994, "Reverse Engineering the Yield Curve," NBER Working Paper No. 4676.
- Baillie, R. T. and T. Bollerslev, 1997, "The Forward Premium Anomaly is not as Bad as you Think," manuscript.
- Bekaert, G., R. J. Hodrick, and D. Marshall, 1997a, "On Biases in Tests of the Expectations Hypothesis of the Term Structure of Interest Rates," Journal of Financial Economics 44, 309-348.
- Bekaert, G., R. J. Hodrick, and D. Marshall, 1997b, "The Implications of First-Order Risk Aversion for Asset Market Risk Premiums," Journal of Monetary Economics, forthcoming .
- Bollerslev, T., R. Y. Chou, and K.F. Kroner, 1992, "ARCH Modelling in Finance: A Review of the Theory and Empirical Evidence," Journal of Econometrics 52, 5-60.
- Bray, M. M., 1982, "Learning Estimation and the Stability of Rational Expectations," Journal of Economic Theory 26, 318-339.
- Campbell, J.Y., A. Lo and C. MacKinlay, 1997, The Econometrics of Financial Markets, Princeton University Press: Princeton.
- Campbell, J. Y., and R. J. Shiller, 1991, "Yield Spreads and Interest Rate Movements: A Bird's Eye View," Review of Economic Studies 58, 495-514.
- Conley, T., L. P. Hansen, E. Luttmer, and J. Scheinkman, 1994, "Estimating Subordinated Diffusions from Discrete Time Data," manuscript.
- Cox, J., J. Ingersoll, and S. Ross, 1985, "A Theory of the Term Structure of Interest Rates," Econometrica 53, 385-408.

- Duffie, D. and R. Kan, 1996, A Yield-factor Model of Interest Rates, Mathematical Finance 6, 379-406.
- Elliott, G., and J.H. Stock, 1994, "Inference in Time Series Regression When the Order of Integration of a Regressor is Unknown," Econometric Theory 10, 672-700.
- Evans, M. D. D., 1995, "Peso Problems: Their Theoretical and Empirical Implications," forthcoming in C.R. Rao, ed., Handbook of Statistics: Statistical Methods in Finance.
- Evans, M. D. D. and K. K. Lewis, 1994, "Do Risk Premia Explain It All? Evidence from the Term Structure," Journal of Monetary Economics 33, 285-318.
- Fama, E. F., 1984, "The Information in the Term Structure of Interest Rates," Journal of Financial Economics 13, 509-528.
- Fama, E. F. and R. Bliss, 1987, "The Information in Long-Maturity Rates," American Economic Review 77, 680-692.
- Fisher, M. and C. Gilles, 1996, "Term Premia in Exponential-Affine Models of the Term Structure," Board of Governors of the Federal Reserve System, manuscript.
- Flood, R. P. and P. M. Garber, 1980, "An Economic Theory of Monetary Reform," Journal of Political Economy 88, 24-58.
- Gerlach, S., and F. Smets, "The Term Structure of Euro-rates: Some Evidence in Support of the Expectations Hypothesis," 1997, Journal of International Money and Finance 16, 305-321.
- Gray, S. F., 1995, "An Analysis of Conditional Regime-Switching Models," manuscript.
- Gray, S. F., 1996, "Modelling the Conditional Distribution of Interest Rates as a Regime-Switching Process," Journal of Financial Economics 42, 27-62.
- den Haan, W., 1995, "The Term Structure of Interest Rates in Real and Monetary Economics," Journal of Economic Dynamics and Control 19, 909-40.
- Hamilton, J. D., 1988, "Rational Expectations Analysis of Changes in Regime: An Investigation of the Term Structure of Interest Rates," Journal of Economic Dynamics and Control 12, 385-423.
- Hamilton, J. D., 1989, "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle," Econometrica 57, 357-384.
- Hansen, L. P., 1982, "Large Sample Properties of Generalized Method of Moments Estimators," Econometrica 50, 1029-1054.
- Hardouvelis, G. A., 1994, "The Term Structure Spread and Future Changes in Long and Short Rates in the G7 Countries: Is There a Puzzle?" Journal of Monetary Economics 33, 255-284.

- Jorion, P., and F. Mishkin, 1991, "A Multicountry Comparison of Term-Structure Forecasts at Long Horizons," Journal of Financial Economics 29, 59-80.
- Kaminsky, G., 1993, "Is There a Peso Problem? Evidence from the Dollar/Pound Exchange Rate, 1976-1987" American Economic Review 83, 450-472.
- Kendall, M. G., 1954, "Note on Bias in the Estimation of Autocorrelation," Biometrika 41, 403-404.
- Krasker, W. S., 1980, "The 'Peso Problem' in Testing the Efficiency of Forward Exchange Markets," Journal of Monetary Economics 6, 269-276.
- Lee, L. and R. H. Porter, 1984, "Switching Regression Models With Imperfect Sample Separation Information - With an Application on Cartel Stability" Econometrica 52, 391-418.
- Lewis, K. K., 1989, "Changing Beliefs and Systematic Rational Forecast Errors with Evidence from Foreign Exchange," American Economic Review 79, 621-636.
- Lewis, K. K., 1991, "Was There a 'Peso Problem' in the U.S. Term Structure of Interest Rates: 1979-1982?" International Economic Review 32, 159-173.
- Mankiw, N. G., and J. A. Miron, 1986, "The Changing Behavior of the Term Structure of Interest Rates," Quarterly Journal of Economics, 101, 211-228.
- Marcet, A., and T. J. Sargent, 1989, "Convergence of Least Squares Learning Mechanisms in Self-Referential Linear Stochastic Models," Journal of Economic Theory 48, 337-368.
- McCallum, B. T., 1994, "Monetary Policy and the Term Structure of Interest Rates," manuscript.
- McCulloch, J. H., 1990, "U.S. Term Structure Data, 1946-87," Handbook of Monetary Economics Volume I, pp. 672-715.
- Naik, V. and M. H. Lee, 1994, "The Yield Curve and Bond Option Prices with Discrete Shifts in Economic Regimes," manuscript.
- Rietz, T. A., 1988, "The Equity Risk Premium: A Solution," Journal of Monetary Economics 22, 117-131.
- Roberds, W. and C. H. Whiteman, 1996, "Endogenous Term Premia and Anomalies in the Term Structure of Interest Rates: Explaining the Predictability Smile," Federal Reserve Bank of Atlanta, manuscript.
- Rudebusch, G. D., 1995, "Federal Reserve Interest Rate Targeting, Rational Expectations, and the Term Structure," Journal of Monetary Economics 35, 245-274.
- Schwarz, G., 1978, "Estimating the Dimension of a Model," Annals of Statistics 6, 461-464.
- Shiller, R. J., J. Y. Campbell, and K. L. Schoenholtz, 1983, "Forward Rates and Future Policy: Interpreting the Term Structure of Interest Rates," Brookings Papers on Economic Activity,

173-223.

Sola, M. and J. Driffill, 1994, "Testing the Term Structure of Interest Rates Using a Vector Autoregression with Regime Switching," Journal of Economic Dynamics and Control 18, 601-628

Stambaugh, R., 1988, "The Information in the Forward Rates: Implications for Models of the Term Structure," Journal of Financial Economics 21, 41-69.

Tauchen, G., and R. Hussey, 1991, "Quadrature-Based Methods for Obtaining Approximate Solutions to Nonlinear Asset Pricing Models," Econometrica 59, 371-396.

White, H., 1982, "Maximum Likelihood Estimation of Misspecified Models," Econometrica 50, 1-25.

Table 1

Slope Coefficient Estimates for Equations (7) and (8)

Horizon n (months)	USD		GBP		DEM	
	OLS Slope Coefficient Regression (7)	OLS Slope Coefficient Regression (8)	OLS Slope Coefficient Regression (7)	OLS Slope Coefficient Regression (8)	OLS Slope Coefficient Regression (7)	OLS Slope Coefficient Regression (8)
12	-2.012 (0.311)	0.152 (0.120)	0.095 (0.395)	0.581 (0.186)	-0.054 (0.161)	0.388 (0.125)
36	-3.098 (0.628)	0.448 (0.238)	0.894 (0.620)	1.300 (0.170)	-0.137 (0.390)	1.059 (0.227)
60	-4.211 (0.988)	0.678 (0.244)	0.874 (0.891)	1.426 (0.155)	-0.325 (0.543)	1.285 (0.108)

Note: The sample contains monthly data from 1972:01 through 1996:09. The short rate is the 3-month rate for each currency. The long rate is indicated in column 1. The maximum number of observations is used in each regression. Hence, for equation (7) there are 294 observations, while for equation (8) there are 297-(n-3). Hansen's (1982) GMM standard errors are in parentheses and are computed using the method of Newey-West (1987) to accommodate the overlapping error structure induced by using monthly observations with a multiperiod forecasting horizon. There are 3 Newey-West lags for equation (7) and n-3 lags for equation (8), which is one lag more than is necessary under the null hypothesis.

Table 2**VAR Order Selection Statistics**

Test Statistic	USD	GBP	DEM
Likelihood Ratio 1 vs. 2	201.8 ($<.00001$)	51.72 (.00001)	158.4 ($<.00001$)
Likelihood Ratio 2 vs. 3	55.80 ($<.00001$)	26.57 (.046)	26.32 (.050)
Likelihood Ratio 3 vs. 4	31.11 (.013)	19.25 (.256)	22.26 (.135)
Akaike Criterion 1	-8.736	-6.778	-14.51
Akaike Criterion 2	-9.337	-6.851	-14.96
Akaike Criterion 3	-9.427	-6.837	-14.94
Akaike Criterion 4	-9.431	-6.797	-14.91
Schwarz Criterion 1	-8.535	-6.577	-14.31
Schwarz Criterion 2	-8.935	-6.449	-14.56
Schwarz Criterion 3	-8.825	-6.234	-14.34
Schwarz Criterion 4	-8.627	-5.994	-14.11

Note: The sample contains monthly data from 1972:01 through 1996:09 with 293 observations in all VARs. The four variables in the VARs are the 3-month rate, and the spreads between the 12-month rate and the 3-month rate, the 36-month rate and the 3-month rate, and the 60-month rate and the 3-month rate. The likelihood ratio tests incorporate the degrees of freedom correction recommended by Sims (1980). The other lag-length criteria are from Akaike (1973) and Schwarz (1978).

Table 3

Implied Slope Coefficients and Model Diagnostics from VAR Analyses

	USD	GBP	DEM	Joint
Implied Slope Eq. (7) n=12	-1.948 (0.327)	-0.030 (0.309)	-0.066 (0.161)	-0.069 (0.269)
Implied Slope Eq. (8) n=12	0.070 (0.109)	0.486 (0.134)	0.257 (0.129)	0.375 (0.100)
Correlation n=12	0.144 (0.213)	0.769 (0.138)	0.539 (0.188)	0.807 (0.186)
Standard Dev. Ratio n=12	0.492 (0.128)	0.633 (0.143)	0.477 (0.124)	0.464 (0.067)
Implied Slope Eq. (7) n=36	-3.260 (0.772)	0.797 (0.523)	-0.136 (0.315)	0.203 (0.507)
Implied Slope Eq. (8) n=36	0.396 (0.236)	0.872 (0.191)	0.592 (0.192)	0.768 (0.095)
Correlation n=36	0.419 (0.221)	0.882 (0.078)	0.758 (0.139)	0.917 (0.070)
Standard Dev. Ratio n=36	0.945 (0.175)	0.989 (0.195)	0.782 (0.215)	0.838 (0.108)
Implied Slope Eq. (7) n=60	-4.204 (1.088)	0.858 (0.775)	-0.309 (0.419)	0.278 (0.428)
Implied Slope Eq. (8) n=60	0.600 (0.224)	0.953 (0.183)	0.722 (0.256)	0.900 (0.096)
Correlation n=60	0.515 (0.168)	0.841 (0.070)	0.812 (0.125)	0.921 (0.069)
Standard Dev. Ratio n=60	1.165 (0.157)	1.134 (0.195)	0.889 (0.262)	0.977 (0.124)

Note: The sample contains monthly data from 1972:01 through 1996:09 with 295 observations in the VARs. The short rate is the 3-month rate and the long rate is indicated by n. The Implied Slope Coefficients in the columns labelled "USD", "GBP", and "DEM" use data from the U.S., U.K., and Germany, respectively, and are computed from a four-variable, second-order VAR in the short rate and the spreads between the three longer rates and the short rate. Asymptotic standard errors (in parenthesis) use gradient techniques and are based on GMM standard errors for the slope coefficients and innovation variances of the VARs. Coefficients in the column labelled "joint" are from a joint estimation of these second-order VARs for all three countries under the restriction that the coefficients for all countries are the same. The Correlation and Standard Deviation Ratio statistics measure relations between the theoretical spread and the actual spread.

Table 4

The Effects of Large Spreads on Equation (7)

Horizon n (months)	USD		GBP		DEM	
	$s^N(t,n)$	$s^L(t,n)$	$s^N(t,n)$	$s^L(t,n)$	$s^N(t,n)$	$s^L(t,n)$
12	-1.186 (0.504)	-2.328 (0.318)	0.268 (0.503)	0.008 (0.432)	0.413 (0.291)	-0.222 (0.145)
36	-1.164 (0.949)	-4.692 (0.687)	1.224 (1.150)	0.723 (0.696)	0.256 (0.786)	-0.371 (0.346)
60	-2.177 (0.135)	-6.554 (1.522)	1.554 (1.919)	0.520 (0.935)	-0.034 (0.923)	-0.548 (0.533)

Note: The sample contains monthly data from 1972:01 through 1996:09 for 294 observations. The short rate is the 3-month rate for each currency. The long rate is indicated in column 1. The observations are split into two groups based on the size of the term spreads and regression (7) is re-estimated allowing for different slope coefficients for "normal" ($s^N(t,n)$) and "large" ($s^L(t,n)$) spreads. Spreads falling outside a band of 1.3 standard deviations of the mean spread are defined to be "large." The fraction of large spreads varies between 9.86% and 21.09% of the total sample, yielding a minimum of 29 observations to estimate the large spread slope coefficient. Hansen's (1982) GMM standard errors are in parentheses and are computed using the method of Newey-West (1987) to accommodate the overlapping error structure induced by using monthly observations with a multiperiod forecasting horizon. There are 3 Newey-West lags which is one lag more than is necessary under the null hypothesis.

Table 5

Estimates of Parameters of the Regime-Switching Model

Parameter	Estimate (Stand. error)	Parameter	Estimate (Stand. error)
μ_1	0.0037 (0.0074)	a_{11}	0.7805 (0.3447)
μ_2	0.0716 (0.0393)	b_{11}	0.0378 (0.0369)
μ_3	2.5443 (0.4988)	a_{22}	1.2833 (1.4754)
β_1	0.9991 (0.0016)	b_{22}	0.0749 (0.0552)
β_2	0.9828 (0.0050)	a_{23}	-0.8057 (0.9250)
β_3	0.8184 (0.0371)	b_{23}	0.1151 (0.0800)
σ_1	0.0332 (0.0076)	a_{33}	-3.6232 (0.8610)
σ_2	0.2144 (0.0313)	b_{33}	0.3419 (0.0822)
σ_3	1.1457 (0.3060)		
γ_1	0.4760 (0.1186)		
γ_2	0.3469 (0.0685)		
γ_3	0.1312 (0.1077)		

Note: Data are monthly observations on the three-month interest rate for seven countries, Australia, Germany, Italy, Japan, Sweden, the United Kingdom, and the United States, from 1972:01 through 1995:12. Maximum likelihood robust standard errors (see White (1982)) are in parentheses.

Table 6
Model Diagnostics

	Australia	Germany	Italy	Japan	Sweden	U.K.	U.S.
Min. C.V. 1	0.004	0.003	0.005	0.0002	0.002	0.004	0.003
Mean C.V. 1	0.010	0.007	0.013	0.006	0.009	0.010	0.007
Max. C.V. 1	0.018	0.014	0.020	0.018	0.019	0.016	0.016
Min. C.V. 2	0.117	0.103	0.135	0.015	0.074	0.126	0.095
Mean C.V. 2	0.224	0.171	0.271	0.156	0.216	0.224	0.175
Max. C.V. 2	0.358	0.297	0.380	0.358	0.369	0.319	0.319
Min. C.V. 3	1.870	1.781	1.974	0.850	1.575	1.921	1.730
Mean C.V. 3	2.370	2.138	2.553	2.045	2.335	2.377	2.159
Max. C.V. 3	2.855	2.660	2.918	2.855	2.887	2.730	2.731
Switches 1- 2 or 2-1	25	39	31	30	28	23	22
Switches 2-3 or 3-2	22	20	28	8	32	32	10
% time 1	.31	.29	.30	.35	.22	.18	.23
% time 2	.31	.64	.52	.53	.69	.68	.68
% time 3	.38	.07	.18	.12	.09	.14	.09

Note: Min. (Mean or Max.) C.V. *i* denotes the minimum (mean or maximum) conditional variance in regime *i*, *i* = 1,2,3. Switches 1-2 or 2-1 (2-3 or 3-2) denotes the estimated number of switches between regimes 1 and 2 (2 and 3). The percentage of the time that the model estimates each country spent in regime *i* is denoted % time *i*.

Table 7

Monte Carlo Distributions of the OLS Slope Coefficients, and VAR-based Statistics

Using the Regime-Switching Model as the Data Generating Process

n	Mean	Median	σ	0.5%	2.5%	5%
Panel A: Equation (7) (With Approximation Error)						
12	1.761	1.864	2.070	-4.524	-2.965	-2.160
36	3.173	2.629	2.750	-5.186	-1.476	0.278
60	3.843	2.920	3.050	-0.624	0.655	0.909
Panel B: Equation (8)						
12	1.286	1.257	0.683	-0.550	-0.082	0.181
36	1.939	1.723	1.015	-0.220	0.548	0.724
60	2.111	1.804	1.185	-0.433	0.540	0.761
Panel C: VAR-Implied Equation (7) (With Approximation Error)						
12	1.665	1.789	2.035	-4.459	-2.938	-2.157
36	3.022	2.507	2.689	-5.086	-1.512	0.106
60	3.719	2.840	2.985	-0.157	0.553	0.824
Panel D: VAR-Implied Equation (8)						
12	1.174	1.152	0.596	-0.379	0.021	0.216
36	1.615	1.457	0.751	0.218	0.517	0.655
60	1.648	1.425	0.811	0.202	0.509	0.657

Panel E: VAR Statistics Correlation Coefficient						
n	Mean	Media n	σ	0.5%	2.5%	5%
12	0.694	0.808	0.282	-0.177	0.012	0.119
36	0.863	0.909	0.133	0.355	0.504	0.584
60	0.928	0.952	0.078	0.519	0.725	0.784

Panel F: VAR Statistics Standard Deviation Ratio						
12	1.802	1.715	0.718	0.497	0.698	0.822
36	1.954	1.607	1.076	0.393	0.658	0.768
60	1.790	1.492	0.940	0.329	0.617	0.743

Note: The Monte Carlo evidence is based on 25,000 replications. The data generating process is the discretized, regime-switching model based on the parameters reported in Table 5. There are 297 total observations in each experiment. The columns labelled Mean, Median, σ , 0.5%, 2.5%, and 5% are the sample mean, the median, the standard deviation, and the respective quantiles of the empirical distributions. Panel A reports the OLS slope coefficients from Equation (7) of the change in the yield on an n-period bond on $[3/(n-3)]$ times the term spread between the n-period yield and the short rate, with the n-3 period yield approximated by the n-period yield at time $t+3$. Panel B (8) reports the OLS slope coefficients from Equation (8) of the weighted average of changes in future short rates on the term spread. Panels C through F report statistics based on a second-order VAR that is estimated for $y(t) = [r(t,3), s(t,12), s(t,36), s(t,60)]'$, where $r(t,3)$ is the three-month rate at time t and $s(t,n)$ represents the term spread at time t between the n-month rate and the three-month rate. Panels C and D report implied slope coefficients corresponding to Equations (16) and (15), respectively. Panels E and F report the correlation between the theoretical spread in Equation (14) that satisfies the expectations hypothesis and the actual spread and the ratio of the standard deviation of the theoretical spread to the standard deviation of the actual spread.

Table 8

Bias and Dispersion Using Alternative Regime-Switching Models as the Data Generating Process

	Model A	Model B	Model C	Model D	Full Model
Panel A: Equation (7) (With Approximation Error) - Mean					
12	1.140	1.187	1.224	2.249	1.761
36	1.069	1.103	1.126	2.423	3.173
60	1.053	1.079	1.097	2.258	3.843
Panel B: Equation (8) - Mean					
12	1.019	1.025	1.031	1.310	1.286
36	1.017	1.023	1.027	1.388	1.939
60	1.017	1.023	1.026	1.366	2.111
Panel C: Equation (7) (With Approximation Error) - Standard Deviation					
12	0.1431	0.1663	0.1960	0.9299	1.864
36	0.1474	0.1785	0.2076	1.0853	2.629
60	0.1475	0.1788	0.2070	1.0130	2.920
Panel D: Equation (8) - Standard Deviation					
12	0.0804	0.0970	0.1113	0.4506	1.257
36	0.0398	0.0505	0.0562	0.4455	1.723
60	0.0278	0.0346	0.0378	0.3569	1.804

Note: Models A to D represent simplified versions of the model that imply the same mean, standard deviation and regime frequencies for the short rate in population as the full model. In particular: Model A eliminates the persistence within-regimes, restricts the innovation variance to be constant across and within regimes, and imposes constant regime-switching probabilities. Model B introduces state-dependent regime switching probabilities into Model A. Model C introduces regime-specific innovation variances and conditional heteroskedasticity into Model B. Finally, Model D introduces within-regime persistence into Model A. The exact parametric specifications used are as follows:

Model A: $\mu_1 = 6.16050; \mu_2 = 6.16052; \mu_3 = 16.0082; \beta_1 = \beta_2 = \beta_3 = 0; \sigma_1 = \sigma_2 = \sigma_3 = 2.5; \gamma_1 = \gamma_2 = \gamma_3 = 0; a_{11} = 0.5805; a_{22} = 1.4833; a_{23} = -1.3732; a_{33} = 1.1943; b_{11} = b_{22} = b_{23} = b_{33} =$

0.

Model B: $\mu_1 = 6.16050$; $\mu_2 = 6.16052$; $\mu_3 = 16.0082$; $\beta_1 = \beta_2 = \beta_3 = 0$; $\sigma_1 = \sigma_2 = \sigma_3 = 2.5$; $\gamma_1 = \gamma_2 = \gamma_3 = 0$; $a_{11} = 0.7805$; $a_{22} = 1.3833$; $a_{23} = -3.6232$; $a_{33} = 0.3557$; $b_{11} = 0.0378$; $b_{22} = 0.0749$; $b_{23} = 0.3419$; $b_{33} = 0.1151$.

Model C: $\mu_1 = 6.16050$; $\mu_2 = 6.16052$; $\mu_3 = 16.0082$; $\beta_1 = \beta_2 = \beta_3 = 0$; $\sigma_1 = 0.7827$; $\sigma_2 = 1.1610$; $\sigma_3 = 1.9937$; $\gamma_1 = 0.4760$; $\gamma_2 = 0.3469$; $\gamma_3 = 0.1312$; $a_{11} = 0.7805$; $a_{22} = 1.3833$; $a_{23} = -3.6232$; $a_{33} = 0.3557$; $b_{11} = 0.0378$; $b_{22} = 0.0749$; $b_{23} = 0.3419$; $b_{33} = 0.1151$.

Model D: $\mu_1 = 0.00300$; $\mu_2 = 0.05725$; $\mu_3 = 2.0351$; $\beta_1 = 0.9991$; $\beta_2 = 0.9828$; $\beta_3 = 0.8184$; $\sigma_1 = 0.80000$; $\sigma_2 = 0.79988$; $\sigma_3 = 0.80239$; $\gamma_1 = \gamma_2 = \gamma_3 = 0$; $a_{11} = 0.5805$; $a_{22} = 1.4833$; $a_{23} = -1.3732$; $a_{33} = 1.1943$; $b_{11} = b_{22} = b_{23} = b_{33} = 0$.

The statistics reported in the last column are for the full regime-switching model, with parameters reported in Table 5. These statistics are identical to those displayed in panels A and B of Table 7. They are reproduced here for ease of comparison with the simple models. The Monte Carlo evidence is based on 25,000 replications, with 297 observations in each replication. Panels A and B report the mean of the empirical distributions, whereas Panels C and D report the standard deviation. Panels A and C report small sample statistics for the OLS slope coefficients from Equation (7) of the change in the yield on an n-period bond on $[3/(n-3)]$ times the term spread between the n-period yield and the short rate, with the n-3 period yield approximated by the n-period yield at time t+3. Panels B and D report small sample statistics for the OLS slope coefficients from Equation (8) of the weighted average of changes in future short rates on the term spread.

Table 9

Characteristics of Regime-dependent Term Premiums

n	Equation (7)	Equation (8)	Mean Term premium	Standard Deviation Term premium	Correlation between term premium and short interest rate
12	0.986	1.021	0.354	0.260	0.907
36	1.201	1.136	1.108	0.524	0.966
60	1.317	1.197	1.680	0.615	0.953

Note: The data generating process is described in Section 6 and embeds time-variation in the term premiums. Equation (7) and Equation (13) refer to the slope coefficients of the Campbell-Shiller regressions described in the article, but there is no approximation error for Campbell-Shiller (12). The term premium is the difference between the actual long yield and the yield predicted by the expectations hypothesis. The OLS slope coefficients and the means and standard deviations of the term premiums are population values, that is, they are computed directly from the Markov chain approximation to the regime-switching-model and the corresponding pricing kernel, described in section 6.

Table 10

Monte Carlo Distributions of the OLS Slope Coefficients, and VAR-based Statistics
Using a Pricing Kernel and the Regime-Switching Model as the Data Generating Process

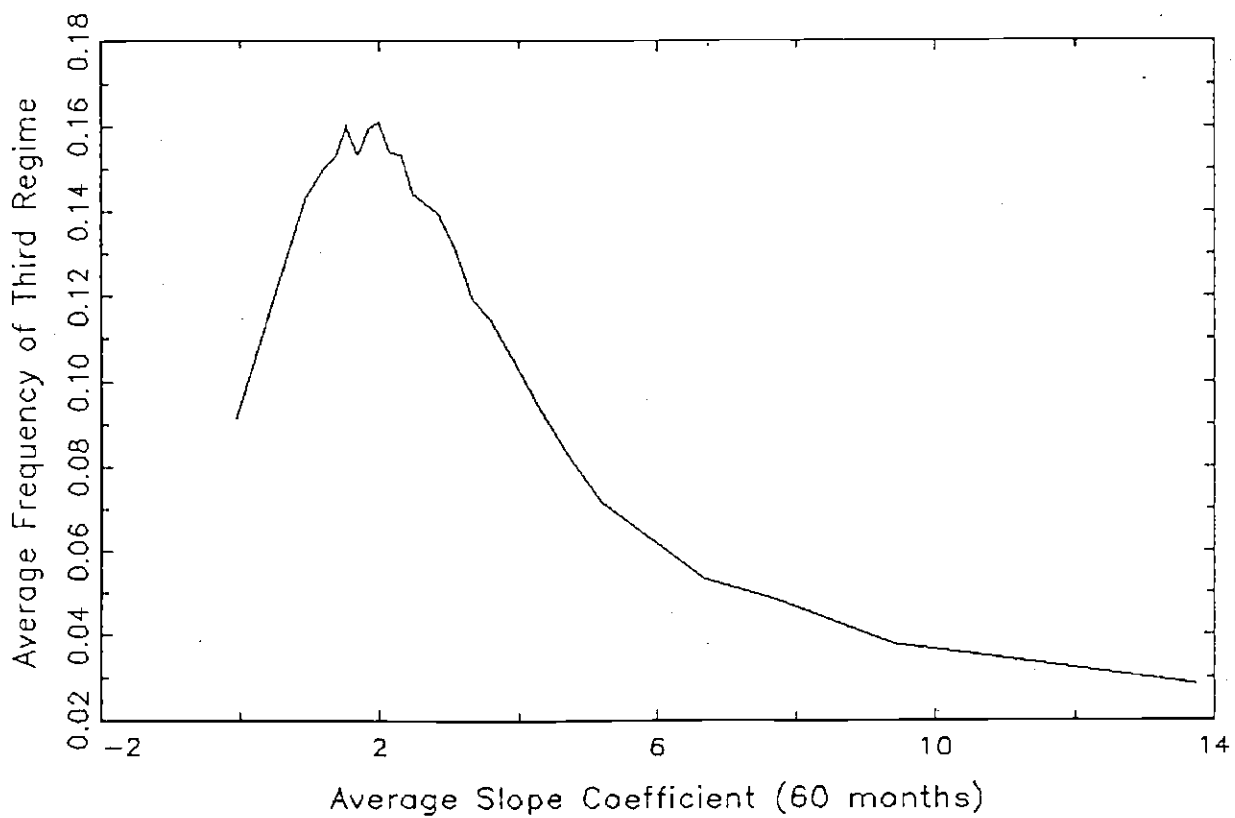
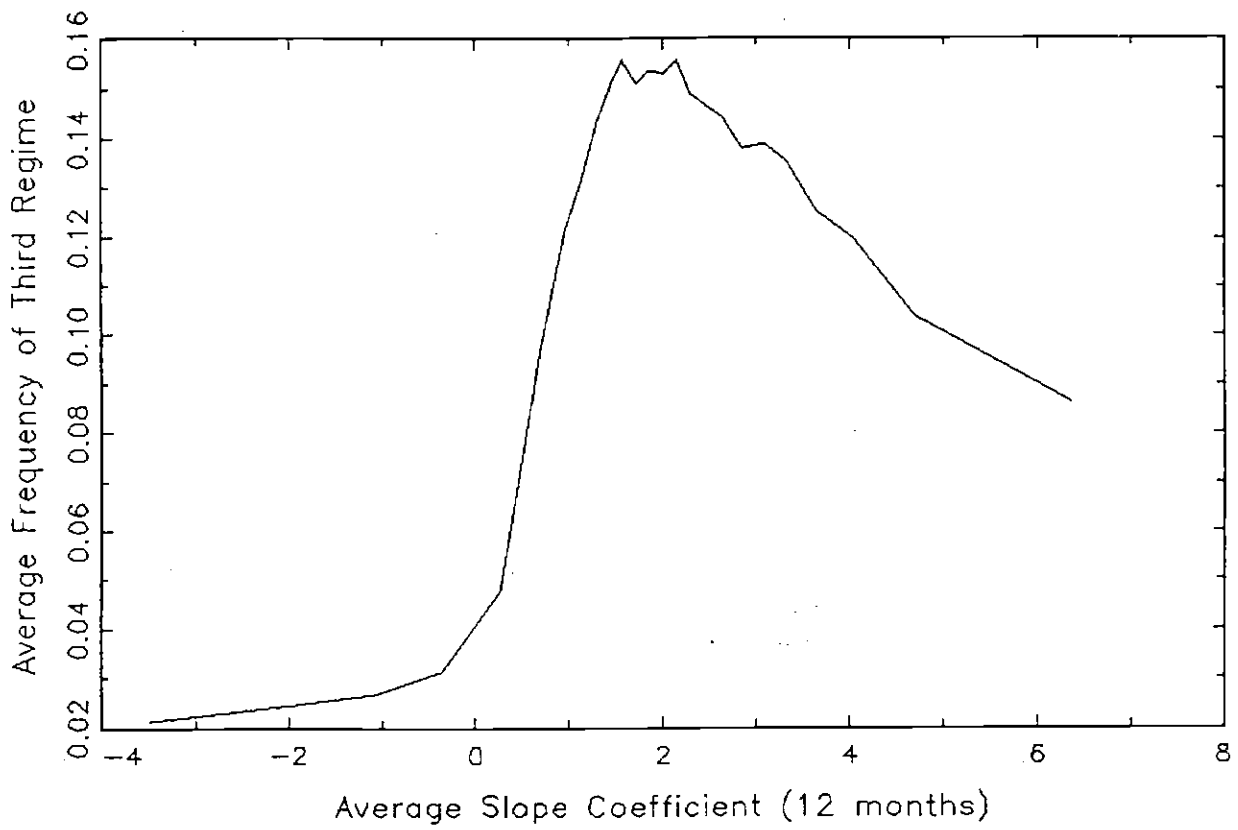
n	Mean	Median	σ	0.5%	2.5%	5%
Panel A: Equation (7) (With Approximation Error)						
12	-0.752	-0.290	3.961	-10.256	-8.262	-7.321
36	-1.526	1.850	8.597	-27.108	-21.979	-18.998
60	-1.553	2.578	12.210	-44.956	-34.879	-29.459
Panel B: Equation (8)						
12	0.233	0.248	1.420	-2.988	-2.289	-1.972
36	0.140	1.229	2.443	-6.248	-4.977	-4.398
60	0.192	1.526	3.243	-9.205	-7.447	-6.565
Panel C: VAR-Implied Equation (7) (With Approximation Error)						
12	-0.749	-0.339	3.873	-9.794	-8.025	-7.152
36	-1.566	1.791	8.499	-26.645	-21.633	-18.947
60	-1.649	2.466	12.112	-44.019	-34.523	-29.405
Panel D: VAR-Implied Equation (8)						
12	0.201	0.191	1.278	-2.432	-1.946	-1.723
36	0.158	1.082	1.991	-4.524	-3.848	-3.496
60	0.345	1.317	2.388	-5.995	-5.242	-4.771

Panel E: VAR Statistics Correlation Coefficient						
n	Mean	Media n	σ	1%	5%	10%
12	0.113	0.070	0.536	-0.820	-0.727	-0.663
36	0.255	0.527	0.657	-0.886	-0.835	-0.802
60	0.423	0.765	0.651	-0.907	-0.866	-0.836

Panel F: VAR Statistics Standard Deviation Ratio						
12	2.589	2.587	0.832	0.818	1.054	1.209
36	3.011	2.826	1.405	0.765	0.997	1.118
60	3.191	2.461	1.972	0.689	0.948	1.061

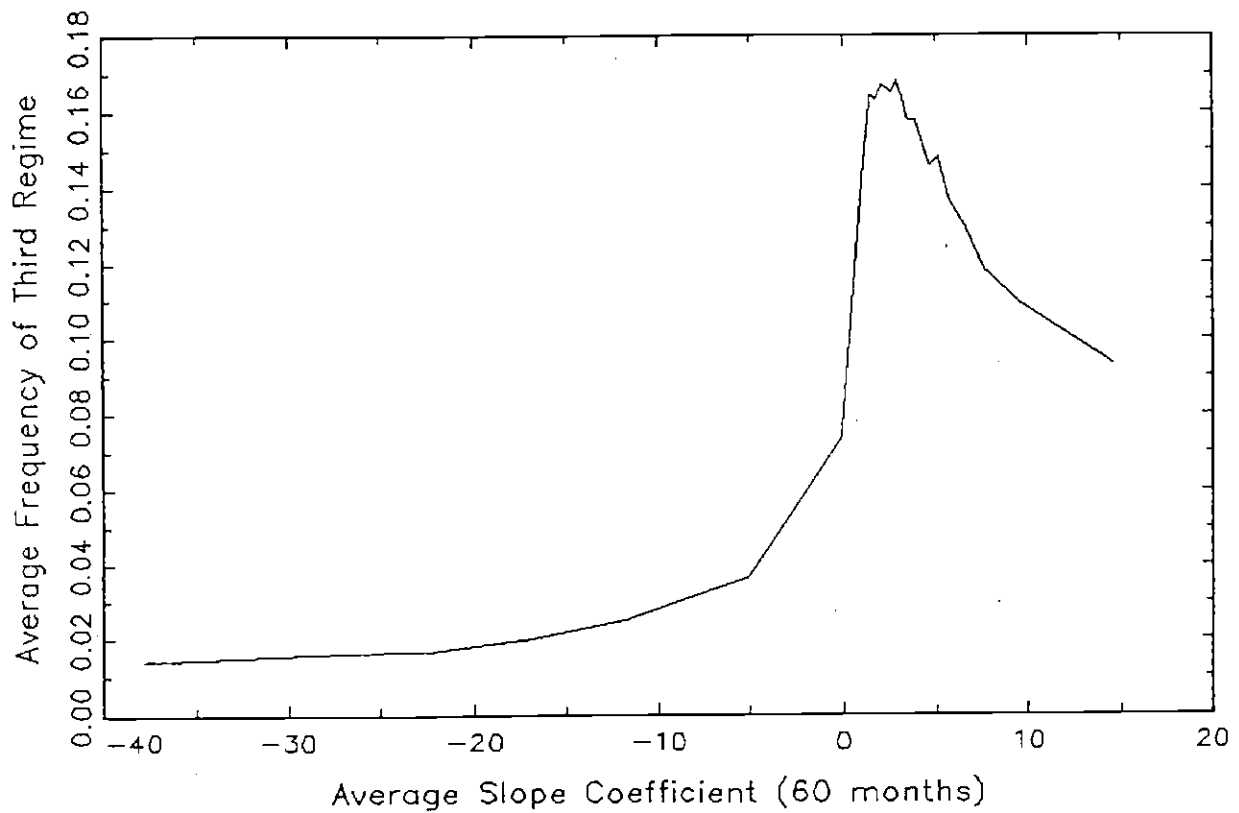
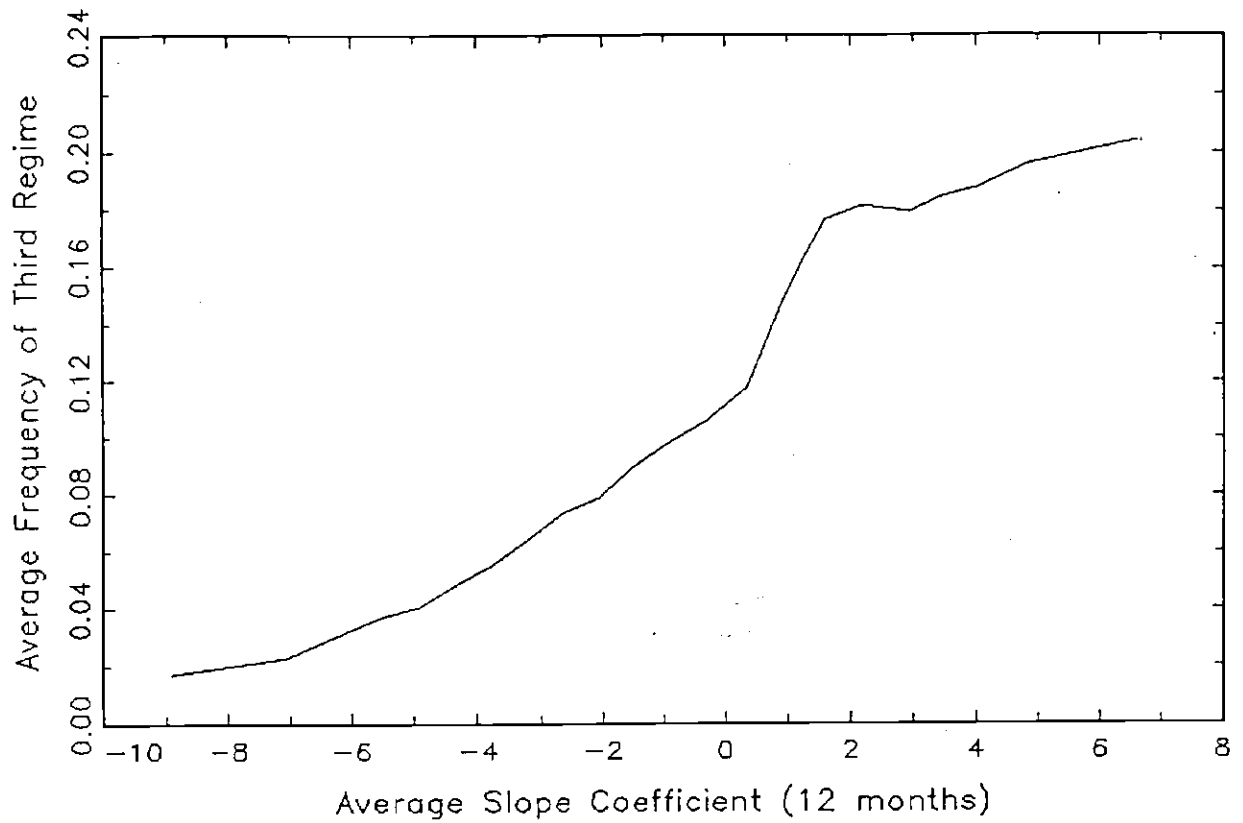
Note: The Monte Carlo evidence is based on 25,000 replications. The data generating process is described in Section 6 and embeds time-variation in the term premiums. There are 297 total observations in each experiment. The columns labelled Mean, Median, σ , 0.5%, 2.5%, and 5% are the sample mean, the median, the standard deviation, and the respective quantiles of the empirical distributions. Panel A reports the OLS slope coefficients from Equation (7) of the change in the yield on an n-period bond on $[3/(n-3)]$ times the term spread between the n-period yield and the short rate, with the n-3 period yield approximated by the n-period yield at time $t+3$. Panel B (13) reports the OLS slope coefficients from Equation (8) of the weighted average of changes in future short rates on the term spread. Panels C through F report statistics based on a second-order VAR that is estimated for $y(t) = [r(t,3), s(t,12), s(t,36), s(t,60)]'$, where $r(t,3)$ is the three-month rate at time t and $s(t,n)$ represents the term spread at time t between the n-month rate and the three-month rate. Panels C and D report implied slope coefficients corresponding to Equations (16) and (15), respectively. Panels E and F report the correlation between the theoretical spread in Equation (14) that satisfies the expectations hypothesis and the actual spread and the ratio of the standard deviation of the theoretical spread to the standard deviation of the actual spread.

Figure 1: Peso Effects in the Expectations Hypothesis - Model



Notes to Figure 1: We simulate the model of section 5, in which the short interest rate is generated by the three-regime switching process given by equations (12) - (16) with $K=3$, using the parameter estimates reported in Table 5, and the long rates are generated by the expectations hypothesis. We conduct 25,000 independent Monte Carlo simulations, with 297 observations in each simulation. For each simulation, we estimate the slope coefficient α_1 in equation (7) with $m=3$. The 25,000 simulations are sorted into bins of 1,000 each, ranked according to the size of α_1 . The graphs in this figure plot the average frequency of regime 3 in each bin against the average value of α_1 in that bin. The upper graph displays results for the 12-month maturity ($n=12$), and the lower graph displays results for the 60-month maturity ($n=60$).

Figure 2: Peso Effects in the Model with Time-Varying Term Premiums



Notes to Figure 2: We simulate the model of section 6, in which the short interest rate is generated by the three-regime switching process given by equations (12) - (16) with $K=3$, using the parameter estimates reported in Table 5, and the long rates are generated by the model of time-varying term premiums given by equations (17) and (18). We conduct 25,000 independent Monte Carlo simulations, with 297 observations in each simulation. For each simulation, we estimate the slope coefficient α_1 in equation (7) with $m=3$. The 25,000 simulations are sorted into bins of 1,000 each, ranked according to the size of α_1 . The graphs in this figure plot the average frequency of regime 3 in each bin against the average value of α_1 in that bin. The upper graph displays results for the 12-month maturity ($n=12$), and the lower graph displays results for the 60-month maturity ($n=60$).