NBER WORKING PAPERS SERIES

DO RISK PREMIA EXPLAIN IT ALL? EVIDENCE FROM THE TERM STRUCTURE

Working paper No. 3451

Martin D.D. Evans

Karen K. Lewis

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
September 1990

We are grateful for useful comments from Ken Froot, John Huizinga, Greg Mankiw, Angelo Melino, Rick Mishkin, and Jim Stock as well as participants in the NBER Summer Institute. Lewis gratefully acknowledges research support from the Olin Foundation and the National Science Foundation under grant #89-02794. This paper is part of NBER's research program in Financial Markets and Monetary Economics. Any opinions expressed are those of the authors and not those of the National Bureau of Economic Research.

NBER Working Paper #3451
September 1990

DO RISK PREMIA EXPLAIN IT ALL?
EVIDENCE FROM THE TERM STRUCTURE

ABSTRACT

Most studies of the expectations theory of the term structure reject the model. However, the significance of the rejections depend strongly upon the form of the test. In this paper, we use the pattern of rejection across maturities to back out the implied behavior of time-varying risk premia and/or market forecasts. We then use a new technique to test whether stationary risk premia alone can be responsible for these rejections. Surprisingly, this test is rejected for short maturities up to 6 months, suggesting that time-varying risk premia do not explain it all. We also describe how this method can be used to test other asset pricing relationships.

Martin D.D. Evans
University of Michigan
611 Tappan
Lorch Hall, Room 240
Ann Arbor, MI 48109

Karen K. Lewis
Department of Economics
University of Pennsylvania
Room 160 McNeil Building
3718 Locust Walk
Philadelphia, PA 19104

1. Introduction

Much of the empirical investigation of the interest rate term structure has focused upon the "expectations theory" that relates the yield on long term bonds to expected future short rates. This theory has frequently been tested using regression tests of the slope of the yield curve (the "yield spread") or forward rates.[1] In general, these tests reject the expectations theory citing standard explanations that either (a) risk premia are time-varying, or (b) market forecasts are biased when viewed *ex post*.[2] In this paper, we investigate these explanations with two goals in mind.

Our first goal is to ask what these rejections must imply about the behavior of bond returns in two polar cases. In the first case, we assume the conventional paradigm that forecasts are always unbiased so that time-varying risk premia must explain all of the rejections. Using the pattern of rejection across regression tests and maturities together with a unifying framework, we back out a number of restrictions on the behavior of risk premia including the following:

(a) the variance of the risk premia increases with the maturity of the bond;

(b) this variance increases faster than the variance of the yield spreads across maturities;

(c) the covariance of the forward premium with the current risk premium is greater than the sum of covariances of the forward premium with all future

---

[1]For only a few examples, see Fama (1984), Shiller (1979), Mankiw (1986), Shiller, Campbell, and Schoenholtz (1983). Fama and Bliss (1987) examine the relationship using long-term forward rates. For a more extensive list of references, see Campbell and Shiller (1989) or the surves by Melino (1986) and Shiller (1987). In addition to the regression tests that are the focus of this paper, Campbell and Shiller (1987) also calculate the yield spread that would be implied by the expectations theory and find that it is highly correlated with the actual yield spread.

[2]See Campbell and Shiller (1989), for example. As discussed below, forecast errors may appear biased when observed *ex post* even though agents use information efficiently and in this sense may be considered to have "rational expectations." Therefore, we call these forecasts *ex post* biased and not "irrational."

risk premia over the maturity of the bond.

Any model of risk premia that assumes expectations are unbiased *ex post* must incorporate these features.

On the other hand, recent research has shown that market participants may make forecasts that appear biased when viewed *ex post* even though they are using all information efficiently *ex ante*. They may make these forecasts if they believe discrete changes in the process of rates are possible, if they are learning about a recent change in the process of rates, or if the nature of the process is non-ergodic.[3] In view of these results, we next examine the polar opposite case to the previous one by supposing that risk premia are constant. Under this assumption, we document the behavior of expectations necessary to explain the pattern of rejections from the regression tests. We find that when the yield curve is upward-sloping, the market on-average predicts a higher future interest rate than occurs *ex post*. Furthermore, we find sufficient conditions for forecast errors to explain all of the rejections. Generally speaking, we would observe the pattern of rejections if current forecast errors have greater covariation with current information than do future forecast errors.

Although we document these stylized facts under the two polar cases of either time-varying risk premia with unbiased forecasts or constant risk premia with *ex post* biased forecasts, actual returns may contain both time-varying risk premia and biased forecasts. Nevertheless, much of the literature has assumed the first case. This observation leads us to the second goal of our paper. Namely, to ask: can we find any evidence against the presumption that time-varying risk premia explain it all?

To ask this question, we test a condition that must necessarily hold when

---

[3]The impact of future discrete change on expectational errors was first pointed out by Rogoff (1980), subsequently by Krasker (1980), and was recently empirically investigated in interest rates by Lewis (1991). The effects of learning on forecast errors after changes in policy are described in Lewis (1989). Hodrick (1990) notes that non-ergodic processes may account for expectational errors.

time-varying risk premia are both solely responsible for rejecting the theory and are stationary.[4] Intuitively, if *ex post* biased forecast errors are not present in the sample, the unexplained component of excess bond returns would be comprised of a white noise forecast error and a stationary time-varying risk premium. In this case, the unexplained residual must necessarily be stationary. On the other hand, if *ex post* biased forecast errors are present in the sample, then the residual may include runs of serially correlated forecast errors. Since interest rates are non-stationary, these runs within the sample may make the excess bond returns appear non-stationary as well. This intuition forms the basis for testing the null hypothesis that no *ex post* biased forecasts appear in the sample. Surprisingly, this test is rejected for short maturities up to 6 months, suggesting that time-varying risk premia do not provide the only source of rejecting the expectations theory.

An interesting feature of our test is that it may be applied more generally than the specific focus of this paper. Our test uses very recent results in time series analysis that allow inference about parameters from cointegrating regressions.[5] Similar tests may be useful in future research.

The structure of the paper is as follows. Section 2 develops the unifying framework that allows for time-varying risk premia and evaluates existing regression test results in the presence of these premia. Section 3 describes how *ex post* biased forecasts affect these results. A summary of what the regression results imply about either time-varying risk premia or expectational errors is contained in Section 4. Section 5 examines whether risk premia can be solely responsible for rejections of the expectations theory. Concluding remarks follow.

---

[4]Standard models of time-varying risk premia imply that these are stationary since they depend upon the time-series properties of the change in consumption. See, for example, Campbell (1987) and Grossman and Shiller (1981).

[5]See Stock and Watson (1989) and Hansen and Phillips (1989).

2. Regression tests and time-varying risk premia

2.1 A unifying framework

The "expectations theory" of the term structure of interest rates relates the equilibrium yield of long bonds to the expected value of short rates over the maturity of the bond. In order to evaluate how time-varying risk premia would affect regression tests, we consider a unifying framework that incorporates the expectations theory as a special case. For the case of pure discount bonds, this relationship is:[6]

$$R^k_t = (1/k) \sum_{i=0}^{k-1} E_t R^1_{t+i} + (1/k)\ \theta^k_t \tag{1}$$

where R^k_t is the yield on a k-period bond purchased at time t, E_t denotes the market's expectations conditional upon information available at time t, and θ^k_t is a time-varying risk premium on holding the k-period bond relative to rolling over one period bonds. The expectations theory implies a special case of equation (1) where the risk premium is a constant, θ^k.

A slightly different form of the risk premia will also prove useful for the investigation below. In particular, we will also define the one-period holding premia:

$$\phi^k_t = k\ R^k_t - (k-1)\ E_t R^{k-1}_{t+1} - R^1_t \tag{2}$$

where ϕ^k_t is the time-varying premium relative to the risk-free one-period rate (R^1_t) on a risky position of holding a k-period bond for one period and then selling the proceeds at the prevailing rate. Iterating (2) forward verifies that $\theta^k_t = \Sigma_{i=1}^{k-2} E_t \phi^{k-i}_{t+i}$, or that the risk premia on holding a long bond relative to rolling over short bonds is equal to the expected value of the sum of holding premia from today until the maturity period of the long bond.

---

[6]See Campbell and Shiller (1989), for example. In this paper, we will be using data series for discount bonds rather than coupon bonds.

2.2 Yield Spread Regressions

According to the expectations theory, the difference between the long k-period bond and the short 1-period bond should be the market's forecast of the change in the long bond. Therefore, tests of the expectations theory have frequently used this yield spread as a regressor. These regression tests have typically taken two forms. In the first, the yield spread predicts the one-period change in the long bond:

$$(k-1)\ (R^{k-1}_{t+1} - R^k_t) = a_0 + a_1\ (R^k_t - R^1_t) + u_{1,t+1}, \tag{3}$$

where under the expectations theory, $a_0 = -\phi^k$ and $a_1 = 1$.[7]

The second form of the yield spread regression relates the yield spread to the *ex post* changes in the short rate over the maturity of the bond:

$$\sum_{i=1}^{k-1} (1 - (i/k))\ \Delta R^1_{t+i} = b_0 + b_1\ (R^k_t - R^1_t) + u_{2,t+k}. \tag{4}$$

Here the expectations theory requires $b_0 = -(1/k)\ \theta^k$ and $b_1 = 1$.[8]

By contrast, when risk premia vary over time, the generalized framework in (1) implies that the coefficients a_1 and b_1 will be biased away from one. In particular, the probability limits are:

$$\text{plim } a_1 = 1 - [\text{Cov}(\phi^k_t,\ R^k_t - R^1_t)/\ \text{Var}(R^k_t - R^1_t)] \tag{5}$$

and

$$\text{plim } b_1 = 1 - (1/k)\sum_{i=0}^{k-2} [\text{Cov}(\phi^{k-i}_{t+i}, R^k_t - R^1_t)/\text{Var}(R^k_t - R^1_t)]. \tag{6}$$

[7]Other studies with estimates of this regression include Campbell and Shiller (1989), Shiller (1979), and Mankiw (1986).

[8]See, for example, Campbell and Shiller (1989).

Thus, when the holding premia are positively correlated with the yield spread, the coefficients in the yield spread regressions will be biased downwards.

The first and second columns of Table 1 report the coefficient estimates of a_1 and b_1, respectively, for one-month through 12-month U.S. T-Bill rates. These series come from the *Center for Research in Security Prices* for the end of the month over the period of availability from June, 1964 to December, 1988 and are used throughout the empirical analysis below. As the table indicates, the parameter estimates for a_1 are negative and increase in absolute value with maturity k. All of the coefficients are significantly less than the hypothesized value of one and even become significantly negative as the maturity horizon lengthens. Furthermore, this pattern continues with longer maturities up to 10 years, as described in Campbell and Shiller (1989). Therefore, this regression test points to a strong rejection of the expectations theory.

By contrast, the estimates for b_1 are positive, though less than one at values near .4. Since the residuals contain k overlapping forecast errors under the null hypothesis, the reported standard errors are corrected for a moving average component of order k-1. Using these standard errors, the hypothesis that the coefficients are equal to one are rejected in all cases. But for the longer maturities investigated in Campbell and Shiller (1989), the point estimates increase and become insignificantly different than one.

Clearly, these two regression tests suggest different results for the expectations theory. In terms of the generalized framework in (1), however, this rejection pattern tells us about the behavior of risk premia across maturities. To see why, note first that the a_1 coefficients less than zero imply from (5):

$$\mathrm{Cov}(\phi^k{}_t,\ R^k{}_t - R^1{}_t) \ > \mathrm{Var}(R^k{}_t - R^1{}_t) > 0,$$

and from (6) that,

$$(1/k)\ \Sigma_{i=0}^{k-2}\ [\mathrm{Cov}(\phi^{k-i}{}_{t+i},\ R^k{}_t - R^1{}_t)] > \mathrm{Var}(R^k{}_t - R^1{}_t) > 0.$$

Taken together, the negative estimates for a_1 and the positive estimates for b_1 imply that $Cov(\phi^k_t, R^k_t - R^1_t) > (1/(k-1))\Sigma_{i=1}^{k-2} Cov(\phi^{k-i}_{t+i}, R^k_t - R^1_t)$. In other words, the covariance between the yield spread and the current k-period holding premium is larger than the average of the covariances between the yield spread and the expected future holding premia.

Such a pattern would likely occur if (a) the $Cov(\phi^j_t, R^k_t - R^1_t)$ were to increase with holding premia, *j*, and if (b) holding premia were stationary so that the covariances between future premia and current yield spreads tend to fall with the future forecast horizon. To consider the validity of (a), excess holding returns for bills of maturity *j* were regressed on each yield spread, $R^k_t - R^1_t$. These excess returns are defined by rewriting (2) and substituting the actual for expected *t+1* rate giving,

$$h^j_{t+1} = j\, R^j_t - (j-1)\, R^{j-1}_{t+1} - R^1_t = \phi^j_t - (j-1)(R^{j-1}_{t+1} - E_t R^{j-1}_{t+1}).$$

Therefore, the coefficients in regressions of holding returns on yield spreads provide the ratio of $Cov(\phi^j_t, R^k_t - R^1_t)$ to the variance of yield spread.

Table 2 reports the results in these regressions. Each column provides the regression coefficient for different excess returns on the same yield spread. In each case, the covariances of premia with the yield spread increase systematically with the maturity of the premia, consistent with the relationship described above. Furthermore, if risk premia are stationary, then the covariances between the current yield spread and the future risk premia would tend to fall with the forecast horizon.

More can be learned from the estimates of a_1. As Fama (1984) has pointed out, a finding that a_1 is negative indicates that the variance of the risk premium is greater than the variance of the yield spread.[9] However, as the third column in Table 1 indicates, the variance of the yield spread also increases with maturity. These results therefore imply that the variances of the risk premia grow across maturities at a rate even faster than the variance in yield spreads. Therefore, if time-varying risk premia are responsible for the pattern of rejecting the expectations theory, any model of risk premia must explain why variances increase so dramatically across maturities.

2.3 Forward Rate Regressions

A second type of regression test involves the "forward rate" for a contract to buy a bond in the future. Using (1) this rate is given by:

$$F^{k-1}{}_t = k\,R^k{}_t - (k-1)\,R^{k-1}{}_t = E_t\,R^1{}_{t+k-1} + E_t \sum_{i=0}^{k-2}[\phi^{k-i}{}_{t+i} - \phi^{k-i-1}{}_{t+i}]$$
$$= E_t\,R^1{}_{t+k-1} + \theta^k{}_t - \theta^{k-1}{}_t$$

where $F^{k-1}{}_t$ is defined as the rate at time t to buy a one-period bond at t+k-1. This rate equals the expected future spot rate plus the difference in holding premia between a k period bond and the k-1 period bond over the maturity of the bond. Hence, the second type of regression test uses as the regressor the "forward premia", or the difference between the forward rate and the current spot rate.

These forward rate regressions typically use two different left-hand side variables: (a) the change in the long bond yield over a short horizon; and (b) changes in the short rate over a long horizon. In particular, the first form regresses the excess holding returns on the forward premia:

---

[9]To see this, redefine $Y^k = (k-1)\,(E_tR^{k-1}{}_{t+1} - R^k{}_t)$. Then, using (2) and (5), we have:

$$\text{plim}\,(a_1) = [(Var(Y^k) + Cov(Y^k, \phi^k))/(Var(Y^k) + Var(\phi^k) + 2\,Cov(Y^k, \phi^k))].$$

Then $a_1 < 0$ implies (i) $Cov(Y^k, \phi^k) < 0$, (ii) $|\,Cov(Y^k, \phi^k)\,| > Var(Y^k)$, and (iii) $Var(\phi^k) > |\,Cov(Y^k, \phi^k)\,| > Var(Y^k)$. Combining (i) and (iii) together with the variance of the yield spread implies: $Var(\phi^k) > Var(R^k - R^1)$.

$$h^k_{t+1} = c_o + c_1 (F^{k-1}_t - R^1_t) + e_{1,t+1} \tag{7}$$

where, according to the expectations theory, $c_o = \phi^k$ and $c_1 = 0$. The second form regresses the *ex post* realizations of the future short rate on the forward premia.[10] In other words,

$$R^1_{t+k-1} - R^1_t = d_o + d_1(F^{k-1}_t - R^1_t) + e_{2,t+k-1} \tag{8}$$

where under the expectations theory, $d_o = \theta^k - \theta^{k-1}$ and $d_1 = 1$.

The generalized framework in (1) implies:

$$\text{plim}(c_1) = \text{Cov}(\phi^k_t, F^{k-1}_t - R^1_t)/\text{Var}(F^{k-1}_t - R^1_t) \tag{9}$$

and

$$\text{plim}(d_1) = 1 - \text{Cov}(\theta^k_t - \theta^{k-1}_t, F^{k-1}_t - R^1_t)/\text{Var}(F^{k-1}_t - R^1_t). \tag{10}$$

In other words, c_1 will differ from zero and d_1 will differ from one if the risk premia are correlated with the forward premia.

The first column of Table 3 reports the coefficient estimates for c_1. Except for the eight month bond, all of the coefficients are significantly positive, thus rejecting the expectations hypothesis. The second column of Table 3 reports the estimated coefficients for d_1. The standard errors are corrected for the MA(k-1) process in the k period overlapping forecast error. As the results indicate, all of the coefficients are significantly less than one.

Comparing the rows of Table 3 provides further information about the behavior of implied risk premia. Here we see that the coefficient estimates for c_1 are always greater than their counterparts for d_1. Furthermore, on maturities greater than 3 months, c_1 is also greater than $1 - d_1$. Comparing (9) and (10), these estimates imply that, if time-varying risk premia are to blame for

---

[10] For example, Fama and Bliss (1987) find that the prediction of the forward rate increases with the forecast horizon.

rejecting the expectations theory,

$$\text{Cov}(\phi^k{}_t,\ F^{k-1}{}_t - R^1{}_t) \ > \ \text{Cov}(\theta^k{}_t - \theta^{k-1}{}_t,\ F^{k-1}{}_t - R^1{}_t).$$

Recall that θ^k are the sum of expected risk premia ϕ^k over the maturity of the bond. In other words, we would find this relationship if the *difference* across risk premia of adjacent maturities have lower covariability with the forward premium than does the *level* of the risk premia.

In summary, if time-varying risk premia are solely responsible for rejecting the expectations theory, the forward rate and yield spread regression results place limits on the dynamics of the risk premia. Among the necessary requirements implied by the regression results are:

(a) the variance of the risk premia increase with maturity of the bond;

(b) this variance increases faster than the variance of the yield spreads;

(c) the covariance of the forward premium with the current risk premium is greater than the sum of covariances of the forward premium with all future risk premia over the maturity of the bond.

Any model of risk premia must therefore incorporate these features.

3. Regression Tests and Ex Post Biased Forecasts

Our interpretation of the regression results above assumes the forecast errors are uncorrelated with current information. This would be true if there were no actual or anticipated switches in the time series process of interest rates. However, recent studies have shown that this assumption may be incorrect if agents rationally anticipate future discrete changes in policy or if they are rationally learning about past discrete changes.[11] Furthermore, studies based upon survey data find that expectations may be on-average incorrect for

---

[11]In addition, they may appear biased if the transition process for policy regimes is not ergodic. Hodrick (1990) cites non-ergodicity as a source of anomalies in financial markets.

significant periods of time.[12] Clearly, this explanation need not exclude the presence of time-varying risk premia discussed above. But in order to focus upon the potential effects of *ex post* biased forecasts in this section, we will assume that risk premia are constant. Later, in Section 5, we will allow for both time-varying risk premia and *ex post* biased forecasts simultaneously.

As described above, *ex post* biased forecasts may arise in small samples even though market participants use information efficiently. These forecasts may be illustrated with an example. Suppose the market believes that the policy process generating interest rates may change by period t+1 with probability, π_t. Therefore, the market's assessed forecast at time *t* of the rate at *t+1* is:

$$E_tR_{t+1} = (1 - \pi_t)\ E^o{}_tR_{t+1} + \pi_t\ E^A{}_tR_{t+1} \tag{11}$$

where $E^o{}_t$, $E^A{}_t$ are the expectations operators conditional upon the current, and alternative policy processes, respectively. But now suppose that *ex post* the interest rate process continues to be driven by the current process *o*.[13] In this case, the observed rates would be uncorrelated with the forecasts that were conditioned upon the current process *o*. In other words,

$$E\{(R^k{}_{t+j} - E^o{}_t\ R^k{}_{t+j}) \mid I_t\} = 0 \tag{12}$$

where I_t is the set of all information available at time t. But the actual market forecast error may be correlated with current information. The market's error in forecasting when viewed *ex post* is the difference between the actual realized rate and the forecast given in equation (11),

---

[12]For example, see Friedman (1979,1980) and more recently Froot (1989).

[13]For the following analysis, this condition is more restrictive than we need. As will be discussed below in Section 4, we require only that any switches in the process occur infrequently within the sample. We use condition (12) only to develop the example in this section.

$$\epsilon^{k}_{t,t+1} = R^{k}_{t+1} - E_tR^{k}_{t+1} = (R^{k}_{t+j} - E^{\circ}_{t} R^{k}_{t+j}) + \pi_t(E^{\circ}_{t}R_{t+1} - E^{A}_{t}R_{t+1}). \qquad (13)$$

Although the first term is uncorrelated with current information, the second term may depend upon current information when π is not zero. As discussed in Lewis (1989), a similar relationship arises if the market believes that a policy shift may have recently taken place with probability π_t.

If expectations appear biased *ex post*, then the correlation between the forecast errors and the yield spreads or forward premia will affect the point estimates in the regression tests above. Then even if the risk premia were constant, the limits of the regression coefficients in (4) and (5) would be:

$$\text{plim } a_1 = 1 + (k-1)\ [\text{Cov}(\epsilon^{k-1}_{t,t+1},\ R^k_t - R^1_t)/\text{Var}(R^k_t - R^1_t)] \qquad (4')$$

and

$$\text{plim } b_1 = 1 + \sum_{i=1}^{k-1} ((k-i)/k)\ [\text{Cov}(\epsilon^{1}_{t+i-1,t+i}, R^k_t - R^1_t)/\ \text{Var}(R^k_t - R^1_t)]. \qquad (5')$$

In this case, both of these sets of estimates in Table 1 indicate that the the yield spread is negatively correlated with the *ex post* forecast errors. On average, when the yield spread was predicting higher interest rates, interest rates turned out to be lower than expected.

Before proceeding, we should note a caveat to the limits described in (4') and (5'). If the sample contains *ex post* biased forecast errors for reasons discussed above, we would expect this bias to disappear asymptotically. Therefore, (4') and (5') are the limits if the current sample were representative of the true asymptotic distribution. We will address this issue more carefully in Section 5 below.

With this caveat, the conflicting evidence between the signs of the a_1 and b_1 estimates imply that:

$$k(k-1) > \sum_{i=1}^{k-1} (k-i)\ [\text{Cov}(\epsilon^{1}_{t+i-1,t+i}, R^k_t - R^1_t)\ /\text{Cov}(\epsilon^{k-1}_{t,t+1},\ R^k_t - R^1_t)].$$

Thus, even if the covariance between the yield spread and the forecast errors were the same so that the ratio of covariances were unity, this relation would hold since $k(k-1) > \Sigma_{i=1}^{k-1}(k-i) = (\frac{1}{2})\, k(k-1)$. Thus, a sufficient condition to find the rejection pattern is that:

$$|Cov(\epsilon^{1}_{t+i-1,t+i}, R^{k}_{t} - R^{1}_{t})| \leq |Cov(\epsilon^{k-1}_{t,t+1}, R^{k}_{t} - R^{1}_{t})|.$$

This condition seems plausible since the variance of the excess holding returns increases with maturity as the last column of Table 2 shows.

With *ex post* biased forecasts and constant risk premia, the coefficient in the forward premia regressions given in (9) and (10) would instead be:

$$plim(c_1) = -(k-1)\, Cov(\epsilon^{k-1}_{t,t+1}, F^{k-1}_{t} - R^{1}_{t})/Var(F^{k-1}_{t} - R^{1}_{t}) \qquad (9')$$

and

$$plim(d_1) = 1 + Cov(\epsilon^{1}_{t,t+k-1}, F^{k-1}_{t} - R^{1}_{t})/Var(F^{k-1}_{t} - R^{1}_{t}). \qquad (10')$$

The positive estimates for c_1 in Table 3 imply that periods of higher forward premia are also periods when people forecast interest rates higher than occur *ex post*. As noted above, Table 3 also suggests that $c_1 > 1 - d_1$ for the maturities 3 months and greater. But even if the covariances between the yield spreads and the forecast errors were the same across maturities, we would tend to find this result since for $k \geq 3$,

$$(k-1) > 1 > Cov(\epsilon^{1}_{t,t+k-1}, F^{k-1}_{t} - R^{1}_{t}) / Cov(\epsilon^{k-1}_{t,t+1}, F^{k-1}_{t} - R^{1}_{t}).$$

In summary, if forecast errors within the sample period are correlated with the yield spread and the forward premia, the conflicting pattern of coefficient estimates found in the literature would arise under fairly weak conditions.

4. Summary of Stylized Facts

To this point, we have discovered that empirical evidence in the literature based upon regression tests places explicit restrictions upon the behavior of bond returns. Using a unifying framework, we have documented what these results must imply about either time-varying risk premia or expectational errors under two polar cases.

Table 4 summarizes these regressions and their implications under each extreme. Part A gives the results for regressions on the yield spreads. The first column refers to the left-hand side variable in the regression, while the second summarizes the point estimates found in Table 1. The third, fourth, and fifth columns list the implied value of the coefficient under the "Expectations Theory," "If Risk Premia Alone," and "If Forecast Errors Alone." Part B in the table lists these same relationships when the right-hand side variables are forward premia. Together, the results indicate the implied behavior of risk premia or expectational errors. Risk premia are capable of explaining the pattern of rejection if they obey the necessary conditions listed in Table 4. Generally speaking, expectational errors could explain this pattern if the covariance between short-term forecast errors and current information were less than the covariance between long-term forecast errors and current information.

In reality, both time-varying risk premia and *ex post* biased market forecasts could cause rejections of the expectations theory. However, since these two effects are not separately identifiable, the typical approach has been to assume that forecasts are *ex post* unbiased and treat all of the unexplained component as time-varying risk premia. The rest of this paper will follow this presumption and ask: can time-varying risk premia explain it all? We will address this question by testing a condition that must necessarily hold when time-varying risk premia are both solely responsible for rejecting the theory and are stationary. Standard models of time-varying risk premia imply that these must be stationary since they depend upon the time-series properties of the change in consumption.[14]

For this purpose, we will develop a new test that may also be used to examine other asset pricing relationships. The basic intuition behind the test is straightforward. If *ex post* biased forecast errors are not present in the sample, the unexplained component of excess bond returns would be the sum of a white noise forecast error and a stationary time-varying risk premium. In this case, the unexplained residual must necessarily be stationary. On the other hand, if *ex post* biased forecast errors are present in the sample, then the residual may include runs of serially correlated forecast errors. Since interest rates are non-stationary, these runs within the sample may make the excess bond returns appear non-stationary as well. This insight forms the basis of our tests below.

5. Do risk premia explain it all?

5.1 Time-Varying Risk Premia and Biased Forecasts

As long as time-varying risk premia are stationary, then standard rational expectations conditions imply that the sum of the risk premia and forecast errors must also be stationary. Therefore, this minimal condition must hold if there are no episodes of *ex post* unbiased forecasts. As an alternative hypothesis, consider the example above in equation (11) where market participants anticipate a future shift in policy or are learning about one that occured in the past. To provide further structure, suppose that the forecast of interest rates conditional upon a switch in policy can be written as:

$$E^{A}_{t}R^{k}_{t+1} = \mu_{t}\; E^{o}_{t}R^{k}_{t+1}. \qquad (14)$$

That is, the forecast of the interest rate conditional upon a switch in the interest rate process, denoted *A*, depends upon the forecast conditional upon the

[14]See, for example, Campbell (1987) and Grossman and Shiller (1981).

current process, denoted o, according to a time-varying parameter, μ. Substituting (14) into (11) and rearranging implies,

$$E_tR^k_{t+1} = (1 + \pi_t(\mu_t - 1))\ E^o_tR^k_{t+1}. \tag{15}$$

Or, defining $\delta^k_t = (1 + \pi_t(\mu_t - 1))$, we have $E_tR^k_{t+1} = \delta^k_t\ E^o_tR^k_{t+1}$. In other words, the actual expected future rate depends upon the efficient forecast conditional upon the current process according to a parameter that reflects both the probability of a switch, π_t, and the difference between forecasts, $(\mu_t - 1)$.

Equation (15) provides a simple and tractable interpretation of expectations. When δ is always and everywhere equal to one, we have the standard rational expectations condition through (12). But if δ_t varies over time, its mean within the sample would give the direction of switch people anticipated on-average. For example, if on-average people anticipated that rates would rise so that $E^A_tR^k_{t+1} > E^o_tR^k_{t+1}$, then (14) shows that the mean of μ would be greater than one. In this case, the mean value of δ_t would also exceed unity.

We can also consider this relationship by forming the *ex post* forecast error based upon the expectations in (15) (and subsuming the k superscript)

$$\epsilon_{t,t+1} = (R_{t+1} - E_tR_{t+1}) = (R_{t+1} - E^o_{t+1}R_{t+1}) + (1 - \delta_t)\ E^o_tR_{t+1} \tag{16}$$

As long as interest rates follow the current process "o", the first term on the right-hand side has mean zero. Thus, the expected mean of forecast errors equals the expected mean of $(1 - \delta_t)\ E^o_tR_{t+1}$. Clearly, δ may vary above and below one so that forecast errors could in fact still have mean zero. However, if shifts in the interest rate process are expected, we may find within any finite sample that δ systematically deviates on one-side of unity for significant periods of time. Since the interest rate process appears non-stationary, forecast errors within the sample would look highly persistent in this case.

If switches occur within the sample, but market participants believe rates may revert back to the old processs (as in Lewis (1991)) or they are learning about the new process (as in Lewis (1989)), then forecast errors would continue to have the same basic form as (16).[15]

5.2 A test of unbiased forecasts with time-varying risk premia

In the empirical results below, we will test a necessary condition for the hypothesis that δ_t is always and everywhere equal to one. In our empirical analysis, we will require for econometric identification that $\delta^k{}_t = \delta^{k-1}{}_t$ for all adjacent pairs of k and k-1. Clearly, this is not restrictive since under the null hypothesis, these parameters are constants where $\delta^k = \delta^{k-1} = 1$. Furthermore, for clarity, we will subsume the k superscript on δ for the remaining discussion.

When expectations incorporate process switches, the risk premium may be written in terms of the current observed process, denoted *o*. Specifically, substituting (15) into (2) implies:

$$\phi^k{}_t = k\, R^k{}_t - (k-1)\ \delta_t\ E^o{}_t\ R^{k-1}{}_{t+1} - R^1{}_t. \tag{17}$$

Substituting the actual for the expected next period interest rate into (17) and rearranging, the equation can be rewritten.

$$(k-1)\ R^{k-1}{}_{t+1} = \alpha_o + \alpha_1\ (k\ R^k{}_t - R^1{}_t) + v_{t+1} \tag{18}$$

where $\alpha_1 = (1/\delta_t)$, and $v_{t+1} = -(1/\delta_t)\ \phi^k{}_t + (k-1)(R^{k-1}{}_{t+1} - E^o{}_t R^{k-1}{}_{t+1}) - \alpha_o$. Thus, α_1 is the inverse of δ_t, capturing the difference between expected future interest rates under the alternative processes. When expectations are *ex post* unbiased as under standard rational expectations conditions, δ and α_1 are always

[15]In this case, we would have:

$$\epsilon_{t,t+1} = (R_{t+1} - E^A{}_t R_{t+1}) + \ (1 - (1/\delta_t))\ E^A{}_t R_{t+1}.$$

With interest rate realizations arising from process "A", $(R_{t+1} - E^A{}_t R_{t+1})$ would be white noise, while systimatic forecast errors would arise from the second term.

and everywhere equal to one. If δ_t varies over time, α_1 would also vary. Within any given sample, the average level of α_1 would not in general equal one, although it might coincidentally equal one as described above in (4.2). Therefore, at a minimum, the null hypothesis of *ex post* unbiased forecasts requires that α_1 be equal to one. We will test this hypothesis below.

If the variables in (18) were stationary, then we could not obtain an asymptotically consistent estimate from this regression due to correlation between the right-hand side variable and the risk premium in the composite residual, v_{t+1}. This estimation problem arose repeatedly when examining the standard regression tests above. But here we formed the regression in terms of variables that are typically considered non-stationary.[16] We developed this equation in order to exploit a result from recent advances in time-series analysis of non-stationary variables. Specifically, as long as the residual is stationary, a regression of a non-stationary variable on another non-stationary variable will provide an asymptotically consistent coefficient estimate *even though the residual may be correlated with the right-hand side variable.*[17]

This result means that, as long as v_{t+1} is stationary, the regression in (18) provides an estimate of $(1/\delta)$ that is asymptotically consistent and independent of the risk premium. But by the definition of v_{t+1} in (18), this residual equals a risk-premium, that is stationary on theoretical grounds, plus a conditional forecast error that is white noise by construction. Therefore, we expect v_{t+1} to be stationary *a priori*. In the results below, we also test this condition with Dickey-Fuller tests.

---

[16]See, for example, Mishkin (1989). To verify this non-stationarity with our data, we tested the hypothesis of unit roots for interest rates using modified Dickey-Fuller tests.

[17]See West (1988) for a derivation of the property of asymptotic normality when non-stationary regressors have drift. For a more general discussion, see Pagan and Wickens (1989) and references therein.

In summary, we can test a minimum requirement for risk premia to explain all of the expectations theory rejections by estimating α_1 in equation (18) and testing whether it is equal to one. Since we have written equation (18) in the form of non-stationary variables, this provides coefficient estimates of α_1 that are asymptotically consistent even though the yield spread may be correlated with the risk premia. On the other hand, this estimation introduces some additional econometric issues to which we now turn.

5.3 Cointegrating Regression Results

Although the regression in (18) provides asymptotically consistent estimates of α_1, correlation between the residual and the right-hand-side variable will bias these estimates in small samples. Two different methods have recently been introduced to adjust for this bias. First, Hansen and Phillips (1989) use the data to generate the asymptotic variance-covariance matrix and then use this matrix to adjust the small sample bias in the coefficient estimate. Second, Stock and Watson (1990) propose a method that purges the small sample bias by including as regressors sufficient leads and lags of (the first-difference) of the right-hand side variable. We use both methods to test the hypothesis that $\delta = 1$ below. Details of these methods are provided in the appendix.

Before proceeding with the analysis, we should emphasize the relationship between the asymptotic distribution of equation (18) and the sample distribution. As long as market participants use information efficiently (and the distribution is ergodic), the true asymptotic mean of δ_t must be equal to one. However, in this section, we treat the sample data as representative of the long run distribution to test whether δ is equal to one. In the next section, we address the small sample problem more directly with Monte Carlo experiments.

The first column of Table 5 reports the coefficient estimates of α_1 for different (k month) maturities of Treasury Bills. All of the point estimates are less than one. The second and third columns report the test statistics for the null hypothesis that $\alpha_1 = 1$ using the Hansen and Phillips (H-P) method. As described in the appendix, this so-called "G-statistic" is a modified Wald statistic based upon an estimate of the asymptotic variance-covariance matrix. When the risk premium and, therefore, v_t are serially correlated, a consistent estimator of the covariance matrix must incorporate lagged autocovariances of the error process. Since we do not know the order of the process for the risk premium, we calculated the "G-statistics" for two extreme truncation lags. In the first, reported as G_0 in column (2), the truncation lag was assumed to be zero with no serial correlation in v_t. In the second, reported as G_6 in column (3), the process could be correlated as many as 6 months. Inspection of the regression residuals indicated that this provides a conservatively large truncation lag of autocorrelation.

These results are reported in the table together with their marginal significance levels. For the statistics assuming no serial correlation, under G_0, the hypothesis that α_1 equals one is rejected with marginal significance levels less that 10% for 6 of the 11 maturities. The more conservative covariance estimates incorporating autocovariances up to 6 months tend to blow up the covariance matrix. As a result, the G_6 statistics cannot reject the hypothesis except for the very short term 2 and 3 month maturities.

Stock and Watson (1990) propose an alternative approach that purges the small sample bias due to simultaneity between the regressors and the residuals. They include leads and lags of the first difference of the right-hand side variable in equations such as (18). In the present case, this yields:

$$(k-1)\ R^{k-1}{}_{t+1} = \alpha_0 + \alpha_1\ (k\ R^k{}_t - R^1{}_t) + \sum_{i=-n+1}^{n+1} \beta_i\ \Delta(k\ R^k{}_{t+i} - R^1{}_{t+i}) + u_{1t}. \qquad (19)$$

Intuitively, including the lagged and future changes in the regressor "soaks up" the correlation between the regressor and the right-hand side variable. This regression makes the residual independent of the entire sequence of interest rate regressors, $(k R^k_t - R^1_t)$, and thus renders normal the asymptotic distribution of the OLS estimator for α_1. As before, the estimated variance-covariance matrix of parameters must include autocorrelation lags in order to incorporate potential serial correlation in the residual.

Column (4) of Table (5) reports the $\chi^2(1)$ Wald test that $\alpha_1 = 1$. These statistics were calculated assuming that n, the number of leads/lags of the regressors, equals 2. These results also assume that the truncation lag for estimating the variance-covariance matrix is 1. The results are not sensitive to these choices, however. As the results indicate, the hypothesis that α_1 equals one is rejected only for the two and three month maturities.

The empirical results in this section rely heavily upon the assumption that, if we were to view an infinite number of observations, they would look like the sample data. However, this condition may not hold. If agents make forecasts that appear biased when viewed *ex post* but use all information efficiently, the systematic nature of forecast errors would disappear with a sufficiently large number of observations. Therefore, in the next section, we use Monte Carlo experiments to calculate the standard errors of α_1 directly from the observed sample.

5.4 Monte Carlo Simulations

When forecasts appear biased *ex post* due to beliefs about future or past shifts in the interest rate process, forecast errors will be systematically biased in one direction for a period before a future shift or after a past shift. Since interest rates appear non-stationary, runs of these systematic forecast errors will likely appear very persistent within the sample. Thus, even if these episodes disappear asymptotically, we may use the data to "bootstrap" the standard errors of α_1 conditional upon the sample.

For this purpose, we generated time-series for the future interest rate based upon the data. We constructed the series in two ways that reflect different assumptions about conditional heteroscedasticity. We considered both cases in order to examine the finite sample sensitivity of our Monte Carlo experiments to assumptions upon heteroscedasticity. As a first case, we generated a benchmark series allowing for unconditional heteroscedasticity but assuming conditional homoscedasticity. Specifically, we followed these steps: (1) Estimate eqn. (19) by OLS, saving the residuals and the estimates of α_o, β_i for all i. (2) For each run, choose alternative values for:

α_1 =	0.9	Model A	(M_A)
	0.95	Model B	(M_B)
	0.99	Model C	(M_C)
	1.00	Model D	(M_D)

(3) Draw randomly from the distribution of residuals in the data. Using α_o, the β_i and the constrained value of α_1, generate a time-series for (k-1) $R^{k-1}{}_{t+1}$ equal to the number of observations, 294. (4) Estimate the cointegrating regression (18) by OLS and save α_1. (5) Repeat (3) and (4) 1000 times.

The second way we calculated the data was to parameterize the conditional heteroscedasticity with an ARCH process. In this case, we followed the same steps as above except for the data generating step (3). Here, we substituted the following step: (3') Take the residuals from (19) and estimate an ARCH process. Scale the residuals by the predictions of the ARCH model. Then, draw randomly from the scaled distribution. Rescale the distribution of residuals using predictions from the ARCH process and use these to generate the time series process of rates.

In estimating the ARCH process, we used a simple rule of thumb to be consistent across maturities. We first estimated an ARCH process of order six and checked to make sure that the implied variances were positive. In the few instances where negative variances were encountered, we reduced the order of the ARCH process until variances were always positive.

In Table 5, columns (5), (6), (7), and (8) report the results of these Monte Carlo experiments for Models A, B, C, and D, respectively. The first numbers in each row are the mean of the distribution of α_1 coefficients. None of the mean estimates of the α_1 coefficients from the heteroscedastic Monte Carlo estimates differed from the homoscedastic Monte Carlo estimates by more than 10^{-3}. Therefore, we report only mean values of α_1 from the homoscedastic case for clarity. The estimates demonstrate the downward-bias in small samples due to the correlation between the residuals and the right-hand side variables. In all cases, the Monte Carlo estimates of α_1 are below the true α_1 used to construct the time-series.

The numbers in parentheses in columns (5) to (8) report the p-values for the hypothesis that the point estimates in column (1) are significantly different from the values of α_1 at the top of each column. The first number in parentheses is the p-value from the distribution of coefficients when the data are generated from the conditionally homoscedastic case. The lower number in parenthesis is the corresponding p-value based upon the conditionally heteroscedastic case. These p-values describe the nature of the distribution. We would almost never reject that the point estimates in column (1) are significantly different than $\alpha_1 = .9$ as in column (5) or $\alpha_1 = .95$ as in column (6) since most of the p-values are near 1. Furthermore, for most of the maturities, we would not reject α_1 is different than .99.

The numbers in parentheses in column (8) give the p-values for our test that $\alpha_1 = 1$. As the table shows, we would reject this hypothesis at the 5% marginal significance level for all maturities except for 7, 8, and 9 months. At the 10% level, we would also reject for the 9 month maturity. Therefore, the null hypothesis that δ is always and everywhere equal to one is rejected for most maturities.

These results indicate that stationary risk premia alone do not fully explain the departure of returns from the expectations theory regression tests. Intuitively, since interest rates are non-stationary, systematic misprediction of rates within a finite sample appear non-stationary as well. Hence, the empirical finding that forward rates at short maturities deviate from the subsequent realization of rates by a non-stationary component suggests systematic *ex post* bias.

5.5 Are yield spreads and excess returns non-stationary?

According to the cointegrating regressions in Table 5, the market forecasts future rates slightly differently than the *ex post* statistical forecast implies. If correct, this implies that the residuals should appear to contain non-stationary components in standard regressions of excess holding returns. Put differently, substituting the forecast error (16) into the definition of the holding returns yields:

$$h^k_{t+1} = k\, R^k_t - (k-1)\, R^{k-1}_{t+1} - R^1_t = \phi^k_t - (k-1)(R^{k-1}_{t+1} - E^o_t\, R^{k-1}_{t+1}) - (k-1)(1-\delta_t)E^o_t\, R^{k-1}_{t+1}. \qquad (20)$$

The first two terms on the right-hand side are, respectively, the risk premia and the statistical innovation to forecasts of the k-1 bond. Hence, both are stationary. But the last term is proportional to the statistical forecast of the long rate. If δ_t systematically deviates from one for prolonged periods of time, this component may appear non-stationary since long rates are themselves non-stationary. Therefore, the evidence for our alternative hypothesis seems to imply that the holding returns and yield spreads should appear non-stationary. Because other studies in the literature find no evidence of this behavior, our alternative model might appear inconsistent with these other studies.

To investigate the effects of the non-stationary expectational errors, we included additional steps to the Monte Carlo experiments described above. Using the process for future rates generated by the data, we calculated both the excess holding returns, h^k_{t+1}, as in (20), and the yield spreads, $R^k_t - R^1_t$. For each of these, we conducted a modified Dickey-Fuller test for unit roots using the method in Phillips and Perron (1987).

Table 6 reports the results of these tests for the case of Model A where at $\alpha_1 = .9$, the parameter deviates the most from one and is therefore the most likely to appear non-stationary. Column (1) reports the first-order autocorrelation multiplied by 100. All of these estimates are far away from one. Columns (2) and (3) provide the Dickey-Fuller unit-root tests for the excess holding returns and the yield spreads, respectively. Since the 5% marginal significance level for 250 observations is -3.43, all of these test statistics strongly reject the hypothesis of unit roots. These statistics in columns (1) to (3) were also calculated for Models B, C, and D where $\alpha_1 = .95$, .99, and 1, respectively. In all cases, the hypothesis of unit roots would be strongly rejected. Not surprisingly, the first-order correlation coefficients become even smaller. For $\alpha_1 = .95$, the maximum autocorrelation is .04 while for $\alpha_1 = .99$, the first-order autocorrelations are all negative and near zero. Therefore, our finding that δ is significantly different than one (at short maturities) is completely consistent with standard findings that excess holding returns and yield spreads are stationary.

5.6 How do long rates react when expectations include process switches?

In Section 3, we used results from regression tests to infer the behavior of expectations under the extreme case that risk premia are constant. Under this view, we found that on-average market participants predicted a higher interest rate than was realized *ex post*. This result is consistent with the general finding in the literature that realized rates are lower than predicted by the yield spread. But in this section, we have found that this phenomenon is highly persistent in the data. The persistence of systematic forecast errors suggests that market participants may anticipate switches in the process of interest rates.

In order to evaluate the effects of these beliefs on the yield spread, we calculated the "theoretical spread" that would occur at the average level of δ_t as found from the estimates in Table 5. Setting δ_t equal to its mean, δ, and iterating the risk premium relationship in equation (17), implies,

$$R^k_t - R^1_t = (1/k) \sum_{i=0}^{k-2} \delta^i E^o_t R^1_{t+i} - R^1_t + (1/k) \sum_{i=0}^{k-2} \delta^i E_t \phi^{k-i}_{t+i}. \quad (21)$$

Since $\delta > 1$, the market's forecasts of long rates include the probability that a shift to higher rates will occur. The probability of the switch increases with the forecast horizon, pushing up the levels of long term bonds. When $\delta = 1$, we have the standard model as in equation (1).

To compare with standard expectations theory assumptions, we calculated the theoretical yield spread using (21) at both $\delta = 1$, the standard case, and the average level of δ_t in the sample. For this purpose, we identified and estimated the process for the short rate as:

$$\Delta R^1_t = \underset{(.161)}{0.792}\, \Delta R^1_{t-1} + \epsilon_t - \underset{(.133)}{0.862}\, \epsilon_{t-1} \quad \text{(Stand. Error)} \qquad R^2 = 0.91 \;\; \text{S.E. } 0.83 \quad (22)$$

For both the standard model and at the mean δ, Figure 1 depicts the plots of the expectations components of the spread between the 11 month bond and the one month bond, i.e., $(1/k) \Sigma^{k-2}_{i=0} E_t R^1_{t+i} - R^1_t$. Similar pictures were also found for the other maturities of the long bond. As the figure indicates, the yield spread varies less than in the standard expectations theory case. In a sense, long rates *underreact* to expected future short-rates. This finding is consistent with the description of long rates found in Mankiw and Summers (1984) and Campbell and Shiller (1984). Our results suggest that this behavior may be partially due to market participants anticipating switches in the process for interest rates.

6. Concluding Remarks

In this paper, we have asked what the pattern of rejecting the expectations theory implies about the behavior of time-varying risk premia and/or *ex post* biased forecasts of interest rates. We uncovered dynamics in either the risk premia or forecast errors that must necessarily hold if either variable were solely responsible for the rejections. Since time-varying risk premia are likely to be present regardless of expectational errors, we tested a necessary condition when risk premia are solely responsible. Surprisingly, we rejected this hypothesis at short maturities. Overall, our evidence using data since 1964 suggests the presence of systematic forecast errors on bonds of short maturities together with time-varying risk premia. Therefore, a challenge for future research will be to explain this behavior.

References

Campbell, John Y., (1987), "Stock returns and the term structure," Journal of Financial Economics 7, 265-296.

Campbell, John Y. and Robert J. Shiller (1984), "A Simple Assount of the Behavor of Long-Term Interest Rates," American Economic Review Papers and Proceedings 74: 44-4.

Campbell, John Y. and Robert J. Shiller (1989), "Yield Spreads and Interest Rate Movements: A Bird's Eye View," N.B.E.R Working Paper No. 3153.

Campbell, John Y. and Robert J. Shiller (1987), "Cointegration and Test of Present Value Models," Journal of Political Economy, 95: 1062-1088.

Fama, Eugene F. (1984a), "The Information in the Term Structure," Journal of Financial Economics, 13: 509-28.

Fama, Eugene F. (1984b), "Term Premiums in Bond Returns," Journal of Financial Economics, 13: 529-46.

Fama, Eugene F. and Robert R. Bliss (1987), "The Information in Long Maturity Forward Rates," American Economic Review, 77: 680-692.

Friedman, Benjamin M., (1979) "Interest Rate Expectations Versus Forward Rates: Evidence from an Expectations Survey," Journal of Finance, 34: 965-73.

Friedman, Benjamin M., (1980), "Survey Evidence on the Rationality of Interest Rate Expectations," Journal of Monetary Economics, 6: 453-65.

Froot, Kenneth A., (1989), "New Hope for the Expectations Hypothesis of the Term Structure of Interest Rates," Journal of Finance, 44: 283-305.

Grossman, Sanford, and Robert Shiller, (1981), "The Determinants of the Variability of Stock Market Prices," American Economic Review, 71: 222-227.

Hansen B. and P. Phillips, (1989), "Statistical Inference in Instrumental Variables Regression with I(1) Processes," Review of Economic Studies, 57 No. 189: 99-126.

Hodrick, Robert J., (1990), "Volatility in the Foreign Exchange and Stock Markets: Is It Excessive?" American Economic Review Papers and Proceedings, 80:

Krasker, William S., (1980), "The Peso Problem in Testing the Efficiency of Foward Exchange Markets," Journal of Monetary Economics, 6: 269-276.

Lewis, Karen K., (1989), "Changing Beliefs and Systematic Forecast Errors," American Economic Review, 79: 621-636.

Lewis, Karen K., (1991), "Was There a 'Peso Problem in the U.S. Term Structure of Interest Rates: 1979-1982?", International Economic Review, forthcoming.

Mankiw, N. Gregory, (1986), "The Term Structure of Interest Rates Revisited," Brookings Papers on Economic Activity, 61: 61-96.

Mankiw, N. Gregory, and Lawrence H. Summers, (1984), "Do Long-Term Interest Rates Overreact to Short-Term Interest Rates?" Brookings Papers on Economic Activity, 1: 223-242.

Melino, Angelo, (1986), "The Term Structure of Interest Rates: Evidence and Theory," N.B.E.R. Working Paper No. 1828.

Mishkin, Frederic, (1989), "Is the Fisher Effect for Real?" Mimeo, Columbia University.

Pagan, Adrian R. and M.R. Wickens, "A Survey of Some Recent Econometric Methods," The Economic Journal, 99: 962-1025.

Phillips, Peter and Pierre Perron, (1987) "Testing for a unit root in a time series regression," Biometrika, 74: 335-46.

Rogoff, Kenneth S., (1980) "Essays on Expectations and Exchange Rate Volatility," Unpublished Ph.D. Dissertation, Massachusetts Institute of Technology.

Shiller, Robert J., (1987), "The Term Structure of Interest Rates," N.B.E.R. Working Paper No. 2341.

Shiller, Robert J., (1979), "The Volatility of Long-Term Interest Rates and Expectations Models of the Term Structure," Journal of Political Economy, 87: 1190-19.

Shiller, Robert J., John Y. Campbell, and Kermit L. Schoenholtz, (1983), "Forward Rates and Future Policy: Interpreting the Term Structure of Interest Rates," Brookings Papers on Economic Activity, pp. 173-217.

Stock, James and Mark Watson, (1989) "A simple MLE of Cointegrating vectors in higher order integrated systems," NBER Technical Working Paper No. 83.

West, Kenneth D., (1988), "Asymptotic Normality When Regressors Have a Unit Root," Econometrica, pp. 1397-417.

Appendix

In this appendix, we describe the methods used in Section 4 to adjust for simultaneous equation bias in the cointegrating regression (18). For more detailed and thorough discussions, see Hansen and Phillips (1989) and Stock and Watson (1990). For notational simplicity, note that equation (18) may be written as,

$$y_t = \gamma x_t + u_{1t} \tag{A1}$$

$$\Delta x_t = u_{2t}, \tag{A2}$$

where u_{1t} and u_{2t} are stationary, $y_t = (k-1)R^{k-1}{}_t$, $x_t = (k\ R^k{}_t - R^1{}_t)$, and the constant term is omitted for simplicity. We are interested in testing $\gamma = 1$. Since x_t is endogenous, it is likely that $\mathrm{Cov}(x_t, u_{1t}) \neq 0$. In this case, γ will be biased in any finite sample. Therefore, test statistics on γ must take account of this bias. This bias arises even though the estimate of γ remains consistent with this simultaneity. We use two methods to adjust for the bias. Below, we give the steps for estimating (A1) and (A2) using each of these methods.

A. Hansen and Phillips method

(1) Estimate (A1) and (A2) to get the estimates of u_{1t} and u_{2t} and the OLS estimate of γ. For future reference, define the vector of residual estimates as: $u_t = [u_{1t}, u_{2t}]'$.

(2) Calculate

$$y_t^+ = Y_t - \Omega_{12}\, \Omega_{22}^{-1}\, \Delta x_t$$

where $\Omega = [\Omega_{ij}] = T^{-1}\Sigma_1^T u_t u_t' + T^{-1}\Sigma_{\tau=1}^{\ell} \omega_{\tau\ell} \Sigma_{t=\tau+1}^{T}(u_t u_{t-\tau}' + u_{t-\tau} u_t'),$

i.e., Ω is the Newey-West estimator of the variance-covariance matrix with ℓ lags of autocovariances and weights ω.

(3) Calculate the "bias-adjusted" estimate of γ as:

$$\gamma^{+} = [\ \Sigma_1^T y_t^{+} x_t - T\ [I,\ -\Omega_{12}\Omega_{22}^{-1}][\Lambda_{21},\ \Lambda_{22}]'\ ][\Sigma_1^T x_t^2]^{-1}$$

where $\Lambda = [\Lambda_{ij}] = T^{-1}[\Sigma_1^T u_t u_t' + \Sigma_{\tau=1}^{\ell}\omega_{\tau\ell}\Sigma_{t=\tau+1}^{T}(u_t\ u_{t-\tau}')]$.

(4) Calculate a modified Wald statistic, known as the G-statistic, to test $\gamma^{+} = 1$. This G-statistic is:

$$G_\ell = (\gamma^{+} - 1)^2\ [\ \Omega_{11.2} \otimes (\Sigma\ x_t^2)^{-1}]^{-1} \sim \chi^2_{(1)}$$

where $$\Omega_{11.2} = \Omega_{11} - \Omega_{12}\ \Omega_{22}^{-1}\ \Omega_{21}.$$

Note that subscript ℓ refers to the number of lags included in the estimators Ω and Λ.

Table 5 reports the results for $\ell = 0$ and $\ell = 6$ in columns (2) and (3).

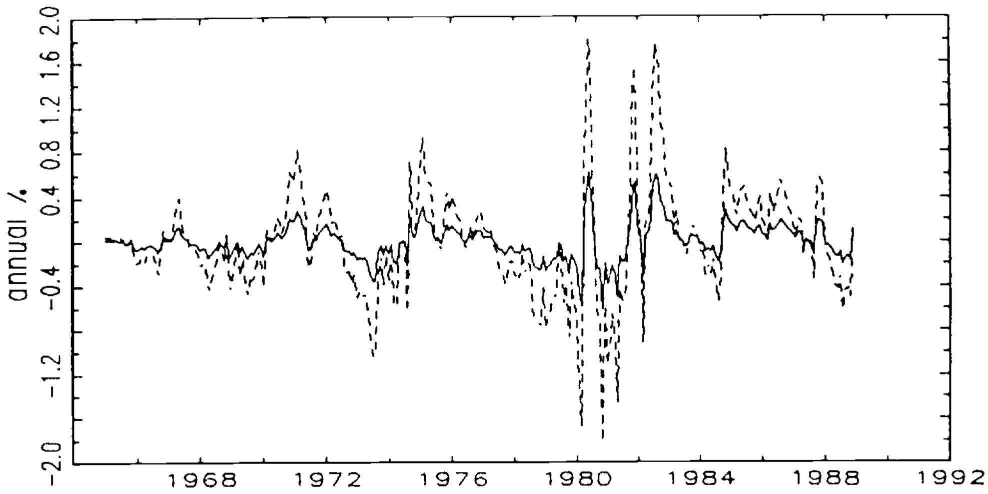
B. Stock and Watson Procedure

Rewrite equation (A1) according to,

$$y_t = \gamma_0 + \gamma\ x_t + \beta(L)\ \Delta x_t + u_{1t} \tag{A3}$$

where $\beta(L)$ is a polynomial in the lag operator, L, i.e., $\beta(L) = (L^n + L^{n-1} + \cdots + L + 1 + L^{-1} + \cdots + L^{-n+1} + L^{-n})$. The idea to rewriting (A1) in this form is to include as many of leads and lags of Δx on the right-hand side of the equation to make u_{1t} independent of x. This implies that the asymptotic distribution of the OLS estimator of γ is normal. Intuitively, including the leads and lags of Δx_t on the right-hand side gets rid of the simultaneous equation bias problem. Note that since u_{1t} will be serially correlated in general, the Wald test of $\gamma=1$ from (A3) should also use the Newey-West estimator for the covariance matrix. Column (4) of Table 5 reports the results when $n=2$ and $\ell=1$. The results are not sensitive to these choices.

Figure 1
Theoretical Yield Spreads



Key: Dashed line denotes the yield spread between the 11 and 1-month bond under standard expectations [δ=1]. Solid line denotes the same spread with *ex post* biased expectations [$\delta > 1$].

Table 1
Regression Tests Using Yield Spreads

Regressor: (R^k-R^1) for k =	Regression Coefficients: Change in Long-Bond (1) a_1	Perfect Foresight Spread (2) b_1	Variances: (R^k-R^1) (3)	$(R^k-R^1)/(k-1)$ (4)
2	-.17[a] (.24)	.42[a] (.12)	.10	.100
3	-.43[a] (.48)	.32[a] (.19)	.15	.075
4	-.70[a] (.61)	.39[a] (.19)	.20	.066
5	-1.15[a] (.68)	.36[a] (.17)	.24	.060
6	-1.27[a] (.72)	.38[a] (.20)	.27	.054
7	-1.48[a] (.79)	.40[a] (.18)	.31	.051
8	-1.52[a] (.83)	.41[a] (.20)	.34	.047
9	-1.72[a] (.79)	.37[a] (.17)	.37	.046
10	-1.89[a] (.86)	.39[a] (.20)	.39	.043
11	-1.90[a] (.88)	.40[a] (.22)	.41	.041

Notes: "a" indicates significantly less than one at the 95% confidence level.

Column (1) is the regression coefficient using $(k-1)(R^{k-1}_{t+1}-R^k_t)$ as dependent variable.

Column (2) is the regression coefficient using $(1/k \sum_{i=1}^{k-1} (k-i)\ \Delta R^1_{t+i}$ as dependent variable. Standard errors are corrected for an MA(k-1) error.

Covariance matrix estimated using Hansen's (1982) sample moments method.

Table 2

Regression Coefficients of Excess Holding Returns on Yield Spreads

$[Cov(h^j_{t+1}, R^k_t - R^1_t)/Var(R^k_t - R^1_t)]$

Return Maturity (j)	Yield Spread Maturity (k) 2	3	4	5	6	7	8	9	10	11	Variance h^j
2	1.17 (.24)	.82 (.23)	.64 (.22)	.54 (.20)	.47 (.19)	.37 (.18)	.32 (.17)	.31 (.16)	.29 (.15)	.27 (.15)	.74
3	1.44 (.54)	1.43 (.48)	1.20 (.44)	1.06 (.39)	.94 (.37)	.79 (.34)	.68 (.32)	.68 (.30)	.65 (.29)	.61 (.28)	2.32
4	1.84 (.80)	1.82 (.70)	1.70 (.61)	1.47 (.54)	1.28 (.50)	1.08 (.47)	.92 (.44)	.95 (.41)	.91 (.40)	.85 (.37)	5.09
5	2.38 (1.08)	2.34 (.92)	2.20 (.80)	2.15 (.68)	1.92 (.64)	1.67 (.59)	1.46 (.56)	1.51 (.51)	1.47 (.50)	1.40 (.47)	8.95
6	2.72 (1.28)	2.65 (1.06)	2.47 (.92)	2.43 (.78)	2.27 (.72)	1.98 (.67)	1.74 (.63)	1.81 (.57)	1.76 (.56)	1.68 (.53)	13.00
7	2.91 (1.45)	2.86 (1.18)	2.69 (1.02)	2.70 (.85)	2.48 (.79)	2.29 (.71)	1.99 (.68)	2.05 (.61)	2.02 (.60)	1.92 (.56)	17.60
8	3.36 (1.64)	3.22 (1.34)	2.99 (1.16)	2.96 (.99)	2.71 (.92)	2.52 (.83)	2.26 (.78)	2.33 (.72)	2.28 (.70)	2.15 (.67)	22.72
9	3.57 (1.85)	3.58 (1.51)	3.36 (1.30)	3.38 (1.10)	3.10 (1.01)	2.89 (.91)	2.57 (.87)	2.72 (.79)	2.66 (.77)	2.52 (.73)	30.66
10	3.57 (2.05)	3.53 (1.70)	3.37 (1.47)	3.49 (1.22)	3.21 (1.12)	3.01 (1.01)	2.66 (.96)	2.82 (.88)	2.89 (.86)	2.73 (.81)	39.37
11	3.49 (2.25)	3.48 (1.86)	3.29 (1.64)	3.56 (1.34)	3.29 (1.22)	3.11 (1.11)	2.75 (1.06)	2.94 (.96)	2.99 (.94)	2.90 (.88)	46.76
Variance $(R^k - R^1)$	.10	.15	.20	.24	.27	.31	.34	.37	.39	.39	.41

Notes: Standard errors in parentheses are calculated using Hansen's (1982) sample moments method

Table 3
Regression Using Forward Premia

Right-hand Side Variable $[F_t^{(k,k-1)}-R_t^1]$	Regression Test: Left-hand Side Variable	
	Excess Return $(k-1)R_{t+1}^{(k-1)}-kR_t^k-R_t^1$	Ex Post Short Rate $R_{t+k-1}^{(1)}-R_t^1$
k =		
2	.08[a]	-.02[c]
	(.03)	(.03)
3	.78[a]	.12[c]
	(.26)	(.19)
4	.96[a]	.40[c]
	(.34)	(.19)
5	1.05[a]	.19[c]
	(.28)	(.08)
6	.88[a]	.29[c]
	(.39)	(.12)
7	.95[a]	.30[c]
	(.36)	(.16)
8	.79	.27[c]
	(.58)	(.17)
9	1.92[a]	.22[c]
	(.52)	(.11)
10	1.51[a]	.18[c]
	(.49)	(.12)
11	1.15[a]	.39[c]
	(.45)	(.18)

Notes: [a] signif. diff. than 0 at 95% level
[c] " " " 1 at 95% level

Table 4
Summary of Regression Tests

Left-Hand Side Variables	Point Estimates	If Expectations Theory,	If Risk Premia Alone,	If Forecast Errors Alone
			A. Yield Spread: $Y_t = (R^k_t - R^1_t)$	
1. $(k-1)(R^{k-1}_{t+1} - R^k_t)$	$a_1 < 0,$	$a_1 = 1,$	$a_1 = 1 - \left[\frac{Cov(\phi^k_t, Y_t)}{Var(Y_t)}\right].$	$a_1 = 1 + (k-1)\left[\frac{Cov(\epsilon^{k-1}_{t,t+1}, Y_t)}{Var(Y_t)}\right]$
2. $\Sigma_{i=1}^{k-1}(1-(i/k))\Delta R^1_{t+i}$	$0 < b_1 < 1,$	$b_1 = 1,$	$b_1 = 1 - \Sigma_{i=0}^{k-2}\left[\frac{Cov(\phi^{k-1}_{t+i}, Y_t)}{k\ Var(Y_t)}\right]$	$b_1 = 1 + \Sigma_{i=1}^{k-1}(k-i)\left[\frac{Cov(\epsilon^1_{t+i-1}, Y_t)}{k\ Var(Y_t)}\right]$
Comparison of (1) and (2)	$a_1 < b_1.$	$a_1 = b_1.$	$Cov(\phi^k_t, Y_t) >$ $(1/k-1)\Sigma_{i=1}^{k-2} Cov(\phi^{k-1}_{t+i}, Y_t)$	$k(k-1)Cov(\epsilon^{k-1}_{t,t+1}, Y_t) >$ $\Sigma_{i=1}^{k-1}(k-i)\ Cov(\epsilon^1_{t+i-1}, Y_t)$
			<u>Other Necessary Conditions</u>:	<u>Sufficient Conditions</u>:
			$Cov(\phi^k_t) > Var(Y_t)$	$\lvert Cov(\epsilon^1_{t+i-1,t+i}, Y_t)\rvert$
			$Var(\phi^j_t) > Var(\phi^{j-1}_t)\ \forall\ j$	$< \lvert Cov(\epsilon^{k-1}_{t,t+1}, Y_t)\rvert$
			$\left[\frac{Var(\phi^j_t)}{Var(\phi^{j-1}_t)}\right] > \left[\frac{Var(R^j_t - R^1_t)}{Var(R^{j-1}_t - R^1_t)}\right] > 1$	
			B. Forward Premia: $Z_t = (F^{k-1}_{t+1} - R^1_t)$	
1 h^k_{t+1}	$0 < c_1$	$c_1 = 0,$	$c_1 = \left[\frac{Cov(\phi^k_t, Z_t)}{Var(Z_t)}\right]$	$c_1 = -\left[\frac{Cov(\epsilon^{k-1}_{t,t+1}, Z_t)}{Var(Z_t)}\right]$
2. $R^1_{t+k-1} - R^1_t$	$d_1 < 1$	$d_1 = 1,$	$d_1 = 1 - \Sigma_{i=0}^{k-2}\left[\frac{Cov(\phi^{k-i}_{t+i} - \phi^{k-i-1}_{t+i}, Z_t)}{Var(Z_t)}\right]$	$d_1 = 1 + \left[\frac{Cov(\epsilon^1_{t,t+k-1}, Z_t)}{Var(Z_t)}\right]$
Comparison of (1) and (2) ($k \geq 3$)	$c_1 > 1 - d_1$	$c_1 = d_1.$	$Cov(\phi^k_t, Z_t) >$ $\Sigma_{i=0}^{k-2} Cov(\phi^{k-i}_{t+i} - \phi^{k-i-1}_{t+i}, Z_t)$	$(k-1)\lvert Cov(\epsilon^{k-1}_{t,t+1}, Z_t)\rvert >$ $\lvert Cov(\epsilon^1_{t,t+k-1}, Z_t)\rvert$
				<u>Sufficient Condition;</u>
				$\lvert Cov(\epsilon^{k-1}_{t,t+1} Z_t)\rvert >$ $\lvert Cov(\epsilon^1_{t,t+k-1}, Z_t)\rvert$

Notes: ϕ^k_{t+i} is the risk premium on holding a k-month bond for one month beginning i months from now. $\epsilon^k_{t+n,t+i} = (R^k_{t+i} - E_{t+n}R^k_{t+i})$, or the error in forecasting a k-month bond i months from now based upon information n periods from now.

Table 5
Cointegration Results

Maturity k	(1) α_1	(2) G_0	(3) G_6	(4) S-W	(5) M_A (α_1=.9)	(6) M_B (α_1=.95)	(7) M_C (α_1=.99)	(8) M_D (α_1=1.0)
2	.921	15.43	4.52	5.21	.870	.920	.960	.970
		(<.001)	(0.34)	(.022)	(1.000)	(.544)	(.001)	(.000)
					(.921)	(.523)	(.017)	(.003)
3	.929	32.51	9.35	3.51	.869	.919	.959	.969
		(.000)	(.002)	(.061)	(1.000)	(.941)	(.000)	(.000)
					(1.000)	(.806)	(.003)	(.000)
4	.951	8.84	1.64	0.26	.866	.916	.956	.966
		(.003)	(.203)	(.610)	(1.000)	(1.000)	(.175)	(.001)
					(1.000)	(1.000)	(.236)	(.015)
5[a]	.956	3.82	0.01	0.11	.867	.917	.957	.967
		(.051)	(.999)	(.740)	(1.000)	(1.000)	(.407)	(.009)
					(1.000)	(1.000)	(.407)	(.019)
6	.960	2.62	0.33	0.93	.867	.917	.957	.967
		(.106)	(.566)	(.335)	(1.000)	(1.000)	(.812)	(.032)
					(1.000)	(1.000)	(.738)	(.084)
7	.967	0.09	0.05	0.03	.868	.918	.958	.968
		(.764)	(.823)	(.862)	(1.000)	(1.000)	(.999)	(.420)
					(1.000)	(1.000)	(.987)	(.429)
8[b]	.967	0.25	0.01	0.03	.869	.919	.959	.969
		(.617)	(.999)	(.862)	(1.000)	(1.000)	(.997)	(.203)
					(1.000)	(1.000)	(.991)	(.295)
9	.962	2.88	0.15	0.06	.869	.917	.957	.967
		(.090)	(.699)	(.806)	(1.000)	(1.000)	(.978)	(.029)
					(1.000)	(1.000)	(.952)	(.070)

(Continued)

Table 5 (continued)
Cointegration Results

Maturity k	(1) α_1	(2) G_0	(3) G_6	(4) S-W	(5) M_A (α_1=.9)	(6) M_B (α_1=.95)	(7) M_C (α_1=.99)	(8) M_D (α_1=1.0)
10	.962	3.16	0.12	0.00	.868	.918	.958	.968
		(.075)	(.729)	(1.00)	(1.000)	(1.000)	(.940)	(.009)
					(1.000)	(1.000)	(.862)	(.044)
11	.968	0.01	0.03	0.06	.868	.918	.958	.968
		(.999)	(.862)	(.806)	(1.000)	(1.000)	(1.000)	(.537)
					(1.000)	(1.000)	(1.000)	(.538)

Notes: [a]Order in ARCH process for heteroscedasticity correction = 3.

[b]Order in ARCH process for heteroscedasticity correction = 4.

[c]Order in ARCH process for heteroscedasticity correction = 5.

α_1 is the parameter estimate in the cointegrating regression: $(k-1)R_{t+1}^{(k-1)} = \alpha_0 + \alpha_1(kR_t^{(k)} - R_t^{(1)} + u_{t+1}^{(k-1)}$.

G_0 and G_6 are $\chi_2(1)$ test statistics of the hypothesis that $\alpha_1 = 1$, for Cov $(u_t u_{t-k}) = 0$ for all $k > \ell$ for $\ell = 0$ and $\ell = 6$, respectively. S-W is the test that $\alpha_1 = 1$ using the $\chi^2(1)$ test statistic in Stock and Watson (1990). M_A, M_B M_C, and M_D are regression coefficient estimates, α_1, using Monte Carlo simulations when α_1 = .9, .95, .99 and 1.0, respectively. The numbers in parentheses in columns (2) and (3) are the marginal significance levels. The numbers in parentheses in columns (5)-(8) are the p-values for the hypothesis that the point estimates from column (1) equal the .9, .95, .99, and 1, respectively. The upper p-value is calculated assuming conditional heteroscedasticity whil the lower p-value assume conditional heteroscedasticity from an ARCH(6) process (unless otherwise indicated).

Table 6
Monte Carlo Results

Maturity	Model A (α_1 = .90)			Model D (α_1 = 1.0)		
k	1st Order Auto:[a] h^k	Unit Root: h^k	Unit Root: R^k-R^1	1st Order Auto:[a] h^k	Unit Root: h^k	Unit Root: R^k-R^1
2	9.41	-18.01	N/A	-0.10	-17.90	N/A
3	12.67	-18.16	-12.13	-0.34	-18.08	-11.97
4	12.32	-18.19	-11.79	-0.32	-18.05	-11.72
5	12.62	-18.09	-11.19	-0.30	-18.07	-11.15
6	12.98	-18.09	-10.45	-0.12	-18.00	-10.54
7	13.63	-17.95	-9.92	-0.44	-18.12	-10.05
8	13.86	-18.14	-9.48	-0.43	-18.07	-9.67
9	13.66	-18.13	-9.32	-0.51	-18.11	-9.49
10	13.31	-18.09	-9.17	0.05	-17.87	-9.33
11	14.15	-17.91	-9.08	-0.01	-18.20	-9.29

Notes: [a]First-order autocorrelations of the h_t^k, the excess holding returns multiplied by 100.

[b]Dickey-Fuller unit root tests.