

# **Endogenous Term Premia and Anomalies In the Term Structure of Interest Rates: Explaining the Predictability Smile**

William Roberds and Charles H. Whiteman

Federal Reserve Bank of Atlanta  
Working Paper 96-11  
October 1996

**Abstract:** Recent studies have documented the existence of a “predictability smile” in the term structure of interest rates: spreads between long maturity rates and short rates predict subsequent movements in interest rates provided the long horizon is three months or less or if the long horizon is two years or more, but not for intermediate maturities. Accounts for portions of the smile involve interest rate smoothing by the Fed, time-varying risk premia, “Peso problems,” and measurement error. We take a more nearly general equilibrium approach to explaining this phenomenon and show that despite its highly restrictive nature, the Cox-Ingersoll-Ross (1985) model of the term structure can account for the predictability smile.

JEL classification: E43, G12

---

The authors thank Andrew Filardo for helpful comments. Whiteman gratefully acknowledges the support of the Federal Reserve Bank of Atlanta and of the National Science Foundation through grant SBR 9422873 to the University of Iowa. The views expressed here are those of the authors and not necessarily those of the Federal Reserve Bank of Atlanta or the Federal Reserve System. Any remaining errors are the authors' responsibility.

Please address questions of substance to William Roberds, Research Officer and Senior Economist, Research Department, Federal Reserve Bank of Atlanta, 104 Marietta Street, N.W., Atlanta, Georgia 30303-2713, 404/521-8970, 404/521-8956 (fax), [wroberds@frbatlanta.org](mailto:wroberds@frbatlanta.org), and Charles H. Whiteman, Department of Economics, University of Iowa, Iowa City, Iowa 52242, [whiteman@uiowa.edu](mailto:whiteman@uiowa.edu).

Questions regarding subscriptions to the Federal Reserve Bank of Atlanta working paper series should be addressed to the Public Affairs Department, Federal Reserve Bank of Atlanta, 104 Marietta Street, N.W., Atlanta, Georgia 30303-2713, 404/521-8020, <http://www.frbatlanta.org>.

## I. Introduction

What accounts for the observed relationship between yields on short- and long-term bonds? A popular framework for the study of this issue involves the *expectations hypothesis* of the term structure of interest rates. According to this hypothesis, the interest rate on a long-term bond is the average of expected short-term interest rates over the duration of the long-term bond. An implication is that when the long rate is above the short rate, short rates should rise by the amount of the long-short spread. Another implication, noted by Macaulay over fifty years ago, is that if the long-short spread is positive, subsequent long rates should rise by a particular fraction of the spread. Conversely, when the short rate is above the long rate, subsequent long and short rates should fall. Numerous tests of these implications have appeared in the literature.

Studies of the implications of yield spreads for movements in short rates indicate that there is a “predictability smile” in the term structure for post-war U.S. data. That is, when the maturity of the long bond is three months or less, short rates generally move as predicted by the expectations hypothesis; for maturities between about three months and two years, short rates do not on average react to long-short spreads; and for maturities beyond two years, the long-short spread again predicts future short rate movements. A graph of the slope coefficients in a regression of subsequently realized average short rates on the long-short spread resembles a “smile”: the coefficients are near unity at short horizons, near zero at intermediate horizons, and return toward unity at long horizons.

For movements in long rates, the “smile” is like Mona Lisa’s “smirk.” For very short maturities, long rates seem unaffected by the long-short spread, and for somewhat longer maturities, positive long-short spreads seem to signal *reductions* in future long rates.<sup>1</sup>

There have been a number of attempts to explain features of the predictability smile and smirk. Because the complete image has become apparent only recently, these studies tend to focus on either the “trough” (the very low predictability at intermediate maturities) or the “left half” of the smile (for short and intermediate maturities). For example, Mankiw and Miron (1986) hypothesize that aspects of the predictability smile for post-1914 short rates can be accounted for by interest-rate smoothing by the Federal Reserve. McCallum (1994), Hardouvelis (1994), Rudebusch (1995), Dotsey and Otrok (1995), and Bekaert, Hodrick, and Marshall (1995), extend the smoothing idea or offer related explanations involving time-varying risk premia, “Peso problems,” and measurement error.

In this paper we take an alternative approach to explaining these empirical regularities, by working out the implications of two benchmark theoretical models of the term structure for the relationship between long-short spreads and subsequent movements in long and short rates. These models, developed by Cox, Ingersoll, and Ross (1985, CIR) and Chen and Scott (1992, CS), provide parsimonious characterizations of the dynamics of the term structure, and have been widely studied in the finance literature. However, no attempt has been made to work out analytically the implications these models carry for the ability of yield spreads to predict future interest rates. Our approach can be seen as complementary to the simulation approach of

---

<sup>1</sup> For evidence on the predictability smile, see, e.g., Wood (1964), Fama (1984), Mankiw and Miron (1986), Fama and Bliss (1987), Hardouvelis (1988), Mishkin (1988), Simon (1989), Campbell and Shiller (1991), Campbell (1995), and Roberds, Runkle, and Whiteman (1996). For the Mona Lisa smirk, see, e.g., Campbell and Shiller (1991), Bekaert, Hodrick, and Marshall (1995), and Campbell (1995).

Backus, Gregory, and Zin (1989) and related papers. These studies examine various term-structure anomalies using simulations of discretized CIR-type models. Our continuous-time approach, which is directed specifically at the predictability smile, trades off the generality obtained through discretization in favor of analytical tractability.

Our primary motivation in undertaking this project is to determine whether there is an “off-the-shelf” explanation for the smile and the smirk, using well-understood models of the term structure. A secondary goal is to investigate to what extent these empirical regularities may be used to restrict the range of permissible parameter values of these and related models.

Our results suggest first that both models are relatively successful in replicating the smile and the smirk. There is of course a large literature which suggests that the models are empirically implausible along a number of dimensions. What we have shown is that the endogenous term premia embedded in even the simple CIR-type models are enough to explain most of the deviations from the expectations hypothesis which have been found to characterize the term structure data.

Second, consistent with the conjecture of Mankiw and Miron (1986) and subsequent work, we find that there must be substantial persistence in the short rate (or one component of the short rate) to generate these patterns. Third, the cost of matching the patterns is that other characteristics of the short rate cannot be matched to the data, though the difficulties in doing so are smaller for the two-factor model.

## **II. Econometric Studies of the Expectations Hypothesis**

According to the expectations hypothesis, long rates can be written as averages of expected future short rates. This implies that current spreads between interest rates at different maturities

predict future interest rate changes. To see this, let  $R_{t,n}$  denote a longer-term,  $n$ -period rate of interest, and let  $R_{t,m}$  denote a shorter,  $m$ -period rate of interest, where  $m$  divides  $n$ . The risk-adjusted expectations hypothesis then states that the  $n$ -period interest rate at time  $t$ ,  $R_{t,n}$ , is the average of the current  $m$ -period interest rate  $R_{t,m}$  and current expectations about future  $m$ -period rates, plus a time-invariant risk premium; that is,

$$R_{t,n} = \frac{1}{k} \sum_{i=0}^{k-1} E_t R_{t+mi,m} + c, \quad k = \frac{n}{m} \quad (1)$$

where  $E_t R_{t+k,m}$  is the expectation at time  $t$  of the  $m$ -period interest rate starting in period  $t+k$  and  $c$  is a term premium which may depend on  $m$  and  $n$  but not  $t$ .

To use the hypothesis to predict short-term rates, we follow the standard approach in the literature (e.g., Campbell and Shiller (1991)) by subtracting  $R_{t,m}$  from both sides of (1), giving

$$\frac{1}{k} \sum_{i=0}^{k-1} E_t R_{t+mi,m} - R_{t,m} = R_{t,n} - R_{t,m} - c.$$

The right side is just the current spread between  $n$ - and  $m$ -period interest rates. This indicates that the difference between the average expected  $m$ -period rate and the current  $m$ -period rate is equal to the current spread between  $n$ - and  $m$ -period rates plus a risk premium.

The expectations hypothesis can therefore be tested by regressing the realized difference between the average  $m$ -period rate and the current  $m$ -period rate,  $(1/k)\sum R_{t+mi,m} - R_{t,m} \equiv S^{(n,m)}$ , on the current spread,  $R_{t,n} - R_{t,m} \equiv S^{(n,m)}$ , i.e.

$$\frac{1}{k} \sum_{i=0}^{k-1} R_{t+mi,m} - R_{t,m} = \alpha + \beta(R_{t,n} - R_{t,m}) + e_{t+mi,m}. \quad (2)$$

The expectations theory implies that  $\beta$  should be unity. Thus, the current spread should be a good predictor of the future average change in short-term rates.

To see the implications of the expectations hypothesis for future long rate changes, note that if (1) holds, the  $n$ -period interest rate should also equal an appropriate weighted average of the  $m$ -

period rate and the  $(n-m)$ -period rate. This implies, after scaling, that the spread should also predict short-run changes in the long rate, i.e.,

$$s_t^{(n,m)} \equiv \frac{m}{n-m} S_t^{(n,m)} = E_t R_{t+m,n-m} - R_{t,n} . \quad (3)$$

This implication of the expectations hypothesis can be tested by regressing the realized value  $R_{t+m,n-m} - R_{t,n}$  on  $s_t^{(n,m)}$ :

$$R_{t+m,n-m} - R_{t,n} = \gamma + \delta \frac{m}{n-m} (R_{t,n} - R_{t,m}) + u_{t+m,n-m} \quad (4)$$

The expectations theory predicts that  $\delta$  will be unity.

Existing evidence on the performance of equation (2) is summarized in Figures 1 and 2. Figure 1 displays estimates of  $\beta$  from regressions using postwar U.S. data, in which  $m = n/2$ ; i.e., the maturity of the long rate is twice that of the short rate. Figure 2 displays estimates of  $\beta$  from regressions in which the maturity of the short rate is one month or less.

The predictability smile is evident in each of the figures: estimated values of  $\beta$  from postwar U.S. data follow a characteristic U-shaped pattern. That is, estimates of  $\beta$  initially fall with increasing  $n$  from a value near unity to values close to zero by the time  $n$  equals six months. For  $n$  between six months and two years,  $\beta$  is close to zero. At maturities  $n$  greater than two years,  $\beta$  rises slowly with increasing  $n$ .

Very different results have been obtained in studies that estimate equation (2) for different data sets. Mankiw and Miron (1986), for example, estimate (2) using pre-1914 U.S. data and obtain values of  $\beta$  close to unity for  $n$  equal to six months and  $m$  equal to three months. Roberds, Runkle, and Whiteman (1996) estimate (2) using daily data from the fed funds market from November 1979 to October 1982, and again obtain estimates of  $\beta$  close to unity for  $m$  equal to one day and  $n$  equal to one, two or three months. Kugler (1988, 1990) also finds that estimated

values of  $\beta$  closer to unity, when (2) is estimated using postwar data from Germany and Switzerland. Marty (1990) finds similar results for 1975-1984 data on Euro interest rates for England, France, and the Netherlands.

Existing evidence on the performance of equation (4) is summarized in Figures 3 and 4. Again, in Figure 3 the long rate maturity is twice that of the short rate and in Figure 4 the short rate maturity is fixed at one month (or less).

We characterize the patterns in Figures 3 and 4 as a "smirk". In particular, if  $m$  is fixed at one month, then  $\delta$  falls monotonically with increasing  $n$ , from a value near zero for  $n$  equal to two months, to values as low as  $-4$  for  $n$  equal to ten years. On the other hand, if  $m=n/2$ , then estimates of  $\delta$  display a U-shaped pattern as a function of  $n$ , falling from a value near zero for  $n$  equal to two months to a minimum value of  $-1$  around  $n$  equal to two years, and slowly rising thereafter.

Mankiw and Miron (1986) argue that the failure of long-short spreads to predict future short rates for (intermediate maturities) results from a combination of time-varying term premia and an unpredictable short rate. Following Mankiw and Summers (1984), Mankiw and Miron note that in the presence of time-varying term premia, when changes in the short rate are not predictable (i.e., the short rate is a random walk), the estimate of the slope coefficient  $\beta$  in equation (2) will approach zero. As the variance of changes in the short rate increase, the coefficient estimate approaches the "expectations hypothesis value" of unity. Mankiw and Miron note that changes in short term rates became much less predictable after the founding of the Fed, and they attribute the more recent "failure" of the expectations theory to the smoothness of the short rate caused by Fed actions. Though Mankiw and Miron were specifically interested in the low slope coefficient estimates for intermediate maturities, one of their results (their Figure II)

carries potential to address the entire predictability smile: when the correlation between the term premium and the expected change in the short rate is negative, slope coefficient estimates of  $\beta$  in (2) initially fall from zero as the variance of the change in short rates rises, and then eventually turn up, become positive, and approach unity. Thus with the right sort of correlation between unobserved term premia and expected changes in short rates, a wide range of slope coefficient estimates can be accommodated.

Rudebusch (1995) provides an explicit model of a smooth short rate. Specifically, he models the funds rate as temporary deviations from a gradually changing and persistent target.<sup>2</sup> This makes the fed funds rate follow a type of Markov-switching process which embodies nonlinear dynamics coupled with substantial persistence. After calibrating a model for the short rate which incorporates these features, Rudebusch simulates short rate data, and builds simulated long rate data using the expectations hypothesis. The simulated data are used to calculate estimates of equation (2). The t-statistics from the resulting estimates display a declining pattern as the maturity of the short rate increases, thus producing a version of half of the predictability smile. Rudebusch does not report coefficient estimates of  $\beta$  from equation (2) because the estimates depend on the variance of term premia (as in Mankiw and Miron) which Rudebusch chose not to include in his model.

Dotsey and Otrok (1995) utilize a short rate process much like Rudebusch's, but also provide an explicit treatment of time-varying term premia. Dotsey and Otrok first present simulations of the term structure in which the short rate follows a persistent, nonlinear process and long rates are modeled as weighted averages of expected future short rates. For these

---

<sup>2</sup> Balduzzi, Bertola, and Foresi (1993) use a similar setup to explain features of the smile.



simulations, long-short interest rate spreads on average correctly predict future movements in short rates. The same result holds when white noise term premia are added to long rates. Dotsey and Otrok then construct time-varying term premia by simulating ARCH-M models which have been fit to observed term premia in postwar U.S. data. When these simulated premia are added to long rates, the resulting estimates of versions of equation (2) roughly reproduce the predictability smile.

McCallum (1994) offers a related but more behavioral explanation involving the Fed's objectives. In McCallum's setup, the Fed favors stable short rates, but is willing to tolerate some short rate volatility in order to pursue countercyclical policy. That is, the Fed seeks to smooth the short rate, but will adjust it when the absolute long-short spread is high. As McCallum notes, for the purpose of explaining term structure anomalies, the source of the feedback is unimportant. It matters not, for example, whether policy tightening in the face of a large long-short spread is because the spread is a good predictor of future output growth or an indication that recent policy has been loose. Upon coupling his short rate process with the expectations hypothesis and a time-varying term premium, McCallum shows that the probability limit of the estimate of  $\beta$  in (2) depends on the degree of persistence in the term premium and the Fed's responsiveness to long-short spreads in setting the short rate. Even with highly persistent term premia and responsive short rates, the estimates can still fall far short of unity.

Relatively few studies have advanced explanations for the long-rate "smirk." In one which has, Bekaert, Hodrick, and Marshall (1995) hypothesize that this pattern can be explained by a regime-switching model of the short-rate, coupled with a version of the "peso problem," in which agents rationally anticipate a high-inflation regime that is underrepresented in historical

data. They find this effect is not strong enough to explain the long-rate smirk in postwar U.S. data. Hardouvelis (1994) considers a number of explanations for this pattern, including white-noise measurement error in the long rate, additive fads, market overreactions, and time-varying risk premia. He finds that none of these explanations can account for the deviations from the expectations hypothesis observed in postwar U.S. data.

The most common themes in these attempts to explain the term structure anomalies are that smooth short rates and time-varying term premia seem necessary. In the next section, we begin our study of whether the most ready source of both--the typical implementation of the term structure model of Cox, Ingersoll, and Ross (1985)--delivers the predictability smile and smirk.

### III. Implications of the Cox-Ingersoll-Ross Model for Spread Regressions<sup>3</sup>

In the continuous-time CIR model, a representative agent with constant relative risk aversion faces production opportunities which evolve according to movements in a single state variable, which is in turn described by a first-order stochastic differential equation. This implies that the instantaneous interest rate is proportional to the state variable (and thus can be thought of itself as the state variable) and evolves according to a process of the form

$$dr = \kappa(\theta - r)dt + \sigma\sqrt{r}dz \quad (5)$$

where  $r$  is the interest rate,  $\kappa$  is the "speed of adjustment" of the interest rate toward its long-run value  $\theta$ ,  $\sigma$  is the instantaneous variance, and  $z$  is a one-dimensional Wiener process. This

---

<sup>3</sup> CIR was developed as a general equilibrium model of the real term structure. In this section we follow the tradition of the finance literature (e.g., Longstaff and Schwartz (1992) or Chen and Scott (1993)) in using CIR as a benchmark model of the nominal term structure.

process has the feature that interest rates cannot become negative, and the absolute variance of the interest rate increases with the interest rate itself.

With this process for the short rate and the assumptions made concerning preferences, CIR show that the time- $t$  yield to maturity on a pure discount bond paying one unit of the consumption good in  $\tau$  periods can be written

$$R_{t,\tau} = (B(\tau)r_t - \ln A(\tau)) / \tau \quad (6)$$

where

$$A(\tau) \equiv \left[ \frac{2\gamma e^{(\kappa+\lambda+\gamma)\tau/2}}{(\kappa+\lambda+\gamma)(e^\tau - 1) + 2\gamma} \right]^{2\kappa\theta/\sigma^2}$$

$$B(\tau) \equiv \frac{2(e^\tau - 1)}{(\kappa + \lambda + \gamma)(e^\tau - 1) + 2\gamma} \quad (7)$$

$$\gamma \equiv ((\kappa + \lambda)^2 + 2\sigma^2)^{1/2}$$

Thus the yield is an affine function of the instantaneous short rate, and depends upon the long-run level of the short rate  $\theta$ , the degree of mean reversion  $\kappa$ , the volatility of the short rate  $\sigma$ , and the “market price of risk”  $\lambda$  – the covariance between changes in the interest rate and the market portfolio.

The structure (5)-(7) is highly restrictive, and some observed yield curves cannot be produced using it. For this reason, a variety of modifications to the setup have been pursued. One we shall explore presently was provided by Chen and Scott (1992). Adopting an approach which has become common in the literature, Chen and Scott begin with a specification of the

instantaneous short rate (rather than the underlying state variables.) The short rate they consider is simply the sum of two independent processes of the form of (5), each with its own  $\theta$ ,  $\kappa$ , and  $\sigma$ .

In this case, the yield curve is of the form

$$R_{t,\tau} = (B_1(\tau)r_{1t} + B_2(\tau)r_{2t} - \ln A_1(\tau) - \ln A_2(\tau)) / \tau \quad (8)$$

with  $A_i(\tau)$  and  $B_i(\tau)$  given by formulas analogous to (7).<sup>4</sup>

To determine the implications the CIR model carries for estimates of coefficients in regressions (2) and (4), it will prove useful to work through a simple example often studied in the literature, namely the case in which the maturity of the long rate is twice that of the short rate. Thus consider equation (2) for this special case, i.e.,

$$\frac{1}{2} \left[ R_{\frac{n}{2}, t+\frac{n}{2}} - R_{\frac{n}{2}, t} \right] = \alpha \left( n, \frac{n}{2} \right) + \beta \left( n, \frac{n}{2} \right) \left( R_{n,t} - R_{\frac{n}{2}, t} \right) + e_{t+\frac{n}{2}}.$$

The population estimate of  $\beta \left( n, \frac{n}{2} \right)$  is given by the covariance of the right-hand-side variable with the left, divided by the variance of the right-hand-side variable. Using (7),

$$\text{var} \left( R_{n,t} - R_{\frac{n}{2}, t} \right) = \left[ \frac{B(n)}{n} - \frac{2B\left(\frac{n}{2}\right)}{n} \right]^2 \text{var}(r_t)$$

and

$$\text{cov} \left\{ \frac{1}{2} \left[ R_{\frac{n}{2}, t+\frac{n}{2}} - R_{\frac{n}{2}, t} \right], \left( R_{n,t} - R_{\frac{n}{2}, t} \right) \right\} = \left[ \frac{2B\left(\frac{n}{2}\right)B(n)}{n^2} - \frac{4B\left(\frac{n}{2}\right)}{n^2} \right] \left[ E \left( r_t r_{t+\frac{n}{2}} \right) - E r_t^2 \right].$$

Thus

<sup>4</sup> A very similar model is presented in Longstaff and Schwartz (1992).

$$\beta\left(n, \frac{n}{2}\right) = \frac{B\left(\frac{n}{2}\right)}{B(n) - 2B\left(\frac{n}{2}\right)} \left[ e^{-\kappa \frac{n}{2}} - 1 \right].$$

An analogous procedure may be used to derive a population estimate for the slope coefficient in equation (4):

$$\delta\left(n, \frac{n}{2}\right) = \frac{2B\left(\frac{n}{2}\right)e^{-\kappa \frac{n}{2}} - B(n)}{B(n) - 2B\left(\frac{n}{2}\right)}.$$

The limiting behavior of these estimates for the more general case in which the maturity of the long rate is a multiple (other than 2) of the short rate maturity is explored in detail in a set of technical appendices (available from the authors on request, attached to this version of the paper). For purposes of comparing our setup to the expectations hypothesis as embodied in (1), perhaps the most important results are that as  $n \rightarrow \infty$ ,

$$\beta\left(n, \frac{n}{2}\right) \rightarrow 1; \delta\left(n, \frac{n}{2}\right) \rightarrow 1 \quad (9)$$

and when  $n \downarrow 0$

$$\beta\left(n, \frac{n}{2}\right) \rightarrow \frac{\kappa}{\kappa + \lambda}; \delta\left(n, \frac{n}{2}\right) \rightarrow \frac{\kappa - \lambda}{\kappa + \lambda}. \quad (10)$$

The limiting value for  $\beta$  in (9) implies that CIR will at least replicate the right endpoint of the predictability smile. As the maturity of the long rate  $n$  lengthens, then  $\beta$  will be driven to unity, irrespective of the values of the model parameters. Similarly, if the market price of risk  $\lambda \approx 0$ , i.e., if the model does not deviate too far from "local" risk neutrality, then (10) reveals that the left endpoint of the predictability smile will be approximately equal to unity.

Limits (9) and (10) are also informative with respect to the ability of the CIR model to replicate the smirk. According to (10), it should be possible to obtain estimates of  $\delta$  near unity for very short maturities, so long as the market price of risk remains sufficiently small. However, (9) shows that for very long maturities, estimates of  $\delta$  will tend to unity irrespective of the model parameter values. Hence the CIR model cannot replicate the negative long-maturity estimates of  $\delta$  obtained for postwar U.S. data.

The behavior of  $\beta\left(n, \frac{n}{2}\right)$  and  $\delta\left(n, \frac{n}{2}\right)$  at intermediate maturities can be parameterized (the technical appendices provide details) by the ratio  $q_n$ , defined as

$$q_n = \left( \frac{\text{var}(\theta_{t,n})}{\text{var}(E_t \Delta R_{t+n/2,n/2})} \right)^{\frac{1}{2}}$$

where the risk premium  $\theta_{t,n}$  is defined by

$$\theta_{t,n} = R_{t,n} - R_{t,n/2} - \frac{1}{2} E_t (R_{t+n/2,n/2} - R_{t,n/2})$$

and where  $\Delta$  is the gapped difference operator  $(1 - L^{n/2})$ . In words,  $q_n$  is the ratio of the standard deviation of the risk premium incorporated into the long rate, to the standard deviation of the forecast change in the short rate. Mankiw-Miron (1986) and others have noted that  $\beta\left(n, \frac{n}{2}\right)$  can be written as

$$\beta\left(n, \frac{n}{2}\right) = \frac{1 + 2\rho_n q_n}{1 + 2\rho_n q_n + 4q_n^2}$$

where

$$\rho_n \equiv \text{corr}(E_t \Delta R_{t+n/2,n/2}, \theta_{t,n}).$$

The one-factor structure of CIR implies perfect correlation between movements in the term premia and forecast changes in the short rate, implying that  $\rho_n = 1$ , and

$$\beta_n = \frac{1}{1+q_n}. \quad (11)$$

Hence the "smile" pattern in  $\beta_n$  requires a fairly sharp peak in  $q_n$  at  $n$  approximately equal to one year.

It is also possible to show (see technical appendix *D* for details) that  $\delta$  depends on  $q_n$  via

$$\delta\left(n, \frac{n}{2}\right) = \frac{1-2q_n}{1+2q_n}. \quad (12)$$

Equation (12) shows that it is possible to obtain negative values of  $\delta$  so long as  $q_n$  exceeds two. Since  $\delta$  is decreasing in  $q_n$ , a sharp peak in  $q_n$  at  $n$  approximately equal to one year, i.e., the pattern necessary to produce the smile in  $\beta$ , will also produce a "smile" in  $\delta$  with a trough at the same maturity.

Another special case of equations (2) and (4) often considered in the literature is where the maturity of the short rate is fixed at 3 months or less, and the maturity of the long rate is allowed to vary. For this case we can show that the population value of  $\beta$  in regression (2) is given approximately by

$$\beta = \hat{\beta}(n, m) \equiv \frac{\frac{1}{n\kappa}(1 - e^{-\kappa n}) - \frac{B(m)}{m}}{\frac{B(n)}{n} - \frac{B(m)}{m}}$$

where the approximation becomes exact as  $m \downarrow 0$ . The approximation of  $\beta$  results from the following computationally convenient approximation which will be valid for small  $m$ :

$$\frac{1}{k} \sum_{i=0}^{k-1} R_{t+mi, m} \approx \frac{1}{n} \int_0^n r_{t+s} ds, \quad \text{where } n = km.$$

It is also straightforward to show that the population value of the slope coefficient in equation (4) will be given by

$$\delta = \delta(n, m) \equiv \frac{n-m}{m} \frac{e^{-\kappa m} \frac{B(n-m)}{n-m} \frac{B(n)}{n}}{\frac{B(n)}{n} \frac{B(m)}{m}}.$$

For small values of the short maturity  $m$ , the limiting values of  $\hat{\beta}(n, m)$  and  $\delta(n, m)$ , the right and left endpoints of the smile, and the left endpoint of the smirk are the same as in the case  $n=2m$  given by equations (9) and (10). (See the technical appendices.) For this case, the right endpoint of the smirk is given by

$$\lim_{n \rightarrow \infty} \lim_{m \downarrow 0} \delta(n, m) = \frac{2\kappa}{\kappa + \lambda + \gamma}.$$

It is also possible to show that  $\hat{\beta}(n, m)$  and  $\delta(n, m)$  can be written as

$$\begin{aligned} \hat{\beta} &= \frac{1}{1+q} \\ \delta &= \frac{1}{1-Q} \end{aligned} \tag{13}$$

where  $q$  ( $Q$ ) is the ratio of the standard deviations of the appropriate risk premium to the standard deviations of forecast changes in the short (long) rates.

The analytical results, taken together, suggest that explaining the smile and the smirk will require that the CIR model generate highly variable risk premia for some maturities, yet not deviate too far from local risk neutrality, i.e.,  $\lambda = 0$ . If local risk neutrality is violated, from (10) it is impossible to obtain a left endpoint of the smile close to unity. However, equations (11) and



(12) show that in order to generate the smile and the smirk, movements in risk premia must be large, relative to forecast changes in interest rates. Since the market price of risk is necessarily small, this suggests that the short rate should display a high degree of persistence, i.e. that  $\kappa$  should be close to zero.

Equations (13) will hold exactly for one-factor models such as the CIR model, but will also hold approximately for multifactor models when (1) the maturity of the long rate  $n$  is "large," and (2) one of the factors is more persistent than the others (which is typically the case). The inability of various term structure models to explain the smirk at the long end of the U.S. term structure results from fact that empirical estimates of  $Q$  are quite large (Hardouvelis(1994) reports estimates on the order of 30), whereas equation (13) implies that values of  $Q$  slightly larger than unity are required in order to generate the highly negative estimates of  $\delta$  that are observed in the data. Efforts to explain the smirk pattern have, in effect, explored the possibility that  $Q$  is mismeasured. Hardouvelis (1994), for example, explores the possibility that observed variations in risk premia are augmented by factors such as measurement error, fads, or overreactions (i.e., that the numerator of  $Q$  is too large), while Bekaert, Hodrick, and Marshall (1995) suggest that small variations short rates can lead to large variations in expected long rates (i.e., that the denominator of  $Q$  is too small).

#### **IV. Empirical Results**

*IV.1 Single-factor model.* To judge how well CIR fits the empirical regularities displayed in Figures 1-4, we conducted a search over the CIR parameters  $\kappa$ ,  $\lambda$ , and  $\sigma$  with the

objective of minimizing squared deviations of the model from the average of empirical estimates displayed in Figure 1.<sup>5</sup> This procedure yielded the following parameter values:

$$\begin{aligned}\kappa &= 0.049 \\ \lambda &= -0.015 \\ \sigma^2 &= 0.81.\end{aligned}\tag{14}$$

As suggested by our analytical results, the fitted value of  $\kappa$  is quite close to zero, implying a very persistent, though still stationary short rate process: the implied one-month autocorrelation of the term structure is  $e^{-\kappa/12} = .9959$ . The fitted value of  $\lambda$  is negative and quite close to zero. The implied left endpoint of the smile is  $\kappa/(\kappa+\lambda) = 1.4511$ .

In the CIR model the sum  $\kappa + \lambda$  has a special interpretation. From equations (6) and (7) it is evident that the CIR model prices the term structure “as if” local risk neutrality held and the short rate followed a process

$$dr = \kappa\theta dt - (\kappa + \lambda)r dt + \sigma\sqrt{r} dz\tag{15}$$

Thus  $\kappa + \lambda$  represents the rate of mean reversion for the “equivalent martingale measure” in (15). The fitted parameter values (14) imply that  $(\kappa + \lambda) = .0340$ , which means that in order to replicate the smile and the smirk the term structure must be priced as if the short rate were almost nonstationary.

The values of  $\beta$  and  $\delta$  implied by the parameter values in (14) are displayed as “CIR” values in Figures 1-4. As might be expected, the fitted parameter values do a reasonable job of replicating the smile when the maturity of the short rate is half that of the long rate (Figure 1). Less obviously, the fitted parameter values do a reasonable job of replicating the smirk when the

---

<sup>5</sup> Note that the CIR parameter  $\theta$  does not enter into the expressions for  $\beta$  or  $\delta$ .

maturity of the short rate is half that of the long rate (Figure 3). When the maturity of the short rate is fixed at one month, the model fit deteriorates somewhat. Figure 2 shows that the values of  $\beta$  implied by the fitted CIR model fit the data for long maturities less than one year, but that they rise too slowly with the long maturity, relative to the data for long maturities beyond one year. Likewise, Figure 4 shows that the implied values of  $\delta$  fall with increasing maturity, for maturities up to roughly one year. For longer maturities, the model implies that  $\delta$  will rise slowly towards its positive asymptotic value, whereas the data require that  $\delta$  continues to fall with increasing maturity.

Although the fitted CIR model is consistent with many of the empirical regularities embodied in the smile and the smirk, it is well-known that the one-factor CIR model does not provide a good approximation to the term structure. In terms of the fitted model, the empirical shortcomings of the CIR model are strikingly apparent along several dimensions.

First, the fitted values of  $\kappa$  and  $\sigma^2$  in (14) are not close to those obtained by researchers estimating discretized versions of equation (5) using postwar U.S. time series on various short rates.<sup>6</sup> In these studies, reported estimates of  $\kappa$  tend to be much larger and estimates of  $\sigma^2$  much smaller than those required to match the smile and the smirk. As a consequence it is impossible to match the first and second moments of actual short rates using the values of  $\kappa$  and  $\sigma^2$  given in (14). To see this, note that the unconditional mean of the short rate process in CIR is given by  $E(r) = \theta$ , while the unconditional variance of the short is rate is given by  $\text{var}(r) = \sigma^2\theta / (2\kappa)$ . A realistic value for first moment requires a value of  $\theta$  say, on the order of  $\theta = 5\%$ . However this

<sup>6</sup> See, for example, Chan et al. (1992); Bliss and Smith (1994); and Pagan, Hall, and Martin (1994).

value, combined with the values of  $\kappa$  and  $\sigma^2$  in (14), implies an unconditional standard deviation for the short rate of 64%. Forcing the second moment of the short rate to a reasonable value, say corresponding to an unconditional standard deviation of 2.5%, implies a mean value for the short rate of less than .01%.

Second, the CIR model cannot produce the declining pattern of estimated t-statistics for  $\beta$  in equation (2) noted by Rudebusch (1995). To check on the model's implications for these t-statistics, we approximated the asymptotic distribution of the estimated slope in (2) as

$$\sqrt{T}(\beta_{OLS} - \beta) \sim N(0, \sigma_{OLS}^2) \quad (16)$$

where  $\sqrt{T}$  is the appropriate sample size,  $\beta_{OLS}$  is the standard OLS estimate, and  $\sigma_{OLS}^2$  is the OLS estimator of the asymptotic variance of  $\beta_{OLS}$ , i.e.,

$$\sigma_{OLS}^2 = \frac{\text{var}(LHS) - \beta^2 \text{var}(RHS)}{\text{var}(RHS)} \quad (17)$$

Where *LHS* and *RHS* refer to the dependent and independent variables in (2), respectively. Carrying out this exercise for the parameter values in (14), taking  $n=2m$  and  $T=400/n$  (a realistic value for postwar sample sizes where  $n$  is measured in months) yields an asymptotic t-statistic of roughly 0.64 at all maturities of less than 10 years. This flat pattern is in sharp contrast with the "smile" pattern of t-statistics reported in the literature.

*IV.2 Two-factor model.* A desire to reconcile the apparent success of the single-factor CIR model at replicating the smile and the smirk with its apparent failure along other dimensions led us to consider the two-factor CS model. Fitting the CS model to the average estimates for  $\beta$  reported in Figure 1 yielded the following values:

$$\begin{array}{ll}
 \kappa_1 = 1.8 & \kappa_2 = .05 \\
 \lambda_1 = -1.7 & \lambda_2 = .06 \\
 \sigma_1^2 = .024 & \sigma_2^2 = 1.0.
 \end{array}
 \tag{18}$$

In contrast, Chen and Scott report the following estimates, based on monthly term-structure data over the period 1960-1987:

$$\begin{array}{ll}
 \kappa_1 = 0.77 & \kappa_2 = 0.00090 \\
 \lambda_1 = -0.12 & \lambda_2 = -0.041 \\
 \sigma_1^2 = 0.017 & \sigma_2^2 = 0.0028.
 \end{array}
 \tag{19}$$

Like Chen and Scott's, our first factor is strongly mean-reverting, and the second factor is nearly a random walk, though in each case, our estimates suggest greater mean reversion than do Chen and Scott's. But there is an important difference between the volatilities: both of our factors are more volatile than Chen and Scott's. The implications of this difference are evident in Figures 5-7, which display population values of expectations hypothesis regression coefficients: the pattern of  $\beta$  values implied by these parameter values is displayed in Figure 5 for the case where  $n=2m$ ; Figure 6 displays the pattern of implied values of  $\delta$  for the case  $n=2m$ ; and Figure 7 shows the values of  $\delta$  when  $m$  is fixed at one month. The fit of these curves for our parameter values ("CS Best") is essentially the same as for the one-factor model of Figures 1-4. The Chen-Scott parameter values ("CS ML"), on the other hand, fail to produce the smile, and only produce the smirk at longer maturities.

The Chen-Scott parameter values were estimated to cause the model to fit a variety of characteristics of the data, so it is unsurprising that they fail to produce the smile and smirk as well as ours do. Put another way, with our parameter values, the CS model will perform less well in matching some other moments in the data. Yet our version of the two-factor model does,

in contrast to the one-factor model, manage to roughly match the unconditional moments of the short rate. For example, the parameter values in (18), in combination with the parameter values  $\theta_1 = .05$  and  $\theta_2 = .0005$ , imply the following unconditional moments for the short rate:

$$E(r) = \theta_1 + \theta_2 = 5.05\%$$

$$\text{var}(r)^{1/2} = \left[ \frac{\sigma_1^2 \theta_1}{2\kappa_1} + \frac{\sigma_2^2 \theta_2}{2\kappa_2} \right]^{1/2} = 7.3\%.$$

The implied unconditional standard deviation for the CS model is much closer to the actual value for postwar U.S. data on short-maturity interest rates ( $\approx 2.5\%$ ) than the one-factor CIR model.

Our version of the CS model can also deliver a pattern of declining estimated t-statistics in  $\beta$  for short maturities. This is readily apparent from Figure 8, which gives asymptotic t-statistics for  $\beta$  implied by parameter values (18) in the case that  $n=2m$  and  $T=400/n$ .

## V. Conclusion

In the context of a model which permits analytical derivation of the complete term structure, a smooth (persistent) short rate and time-varying term premia do seem necessary (Mankiw-Miron, 1986; Hardouvelis, 1994; McCallum, 1994; Bekaert, Hodrick, and Marshall, 1995; Dotsey and Otrok, 1995; Rudebusch, 1995) to explain existing results on implications of long-short spreads for subsequent movements in long and short interest rates. Considerable discipline is imposed by this framework – term premia are highly restricted, and there is no measurement error. It does fall short in some respects, particularly in producing sufficient volatility of long rates, but the framework goes a long way toward explaining term structure

facts, and signals dimensions along which modifications are necessary if models are to be designed which are consistent with a more complete set of such facts.

## References

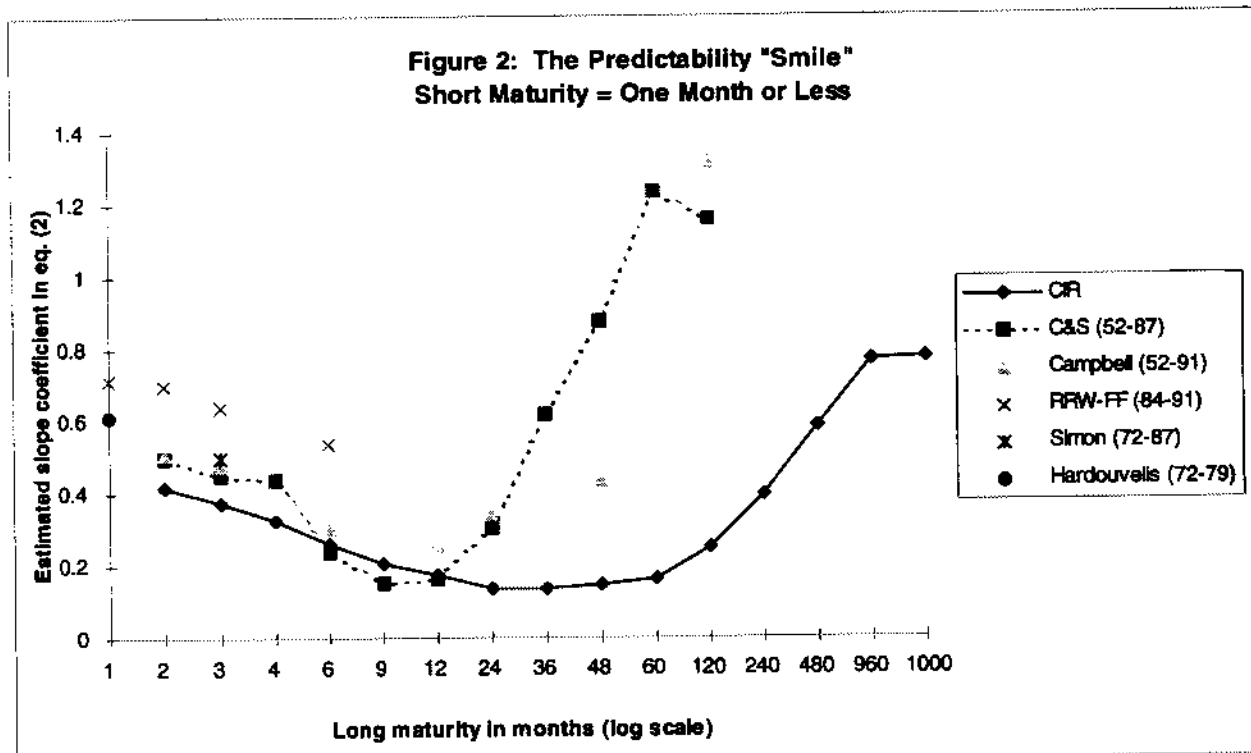
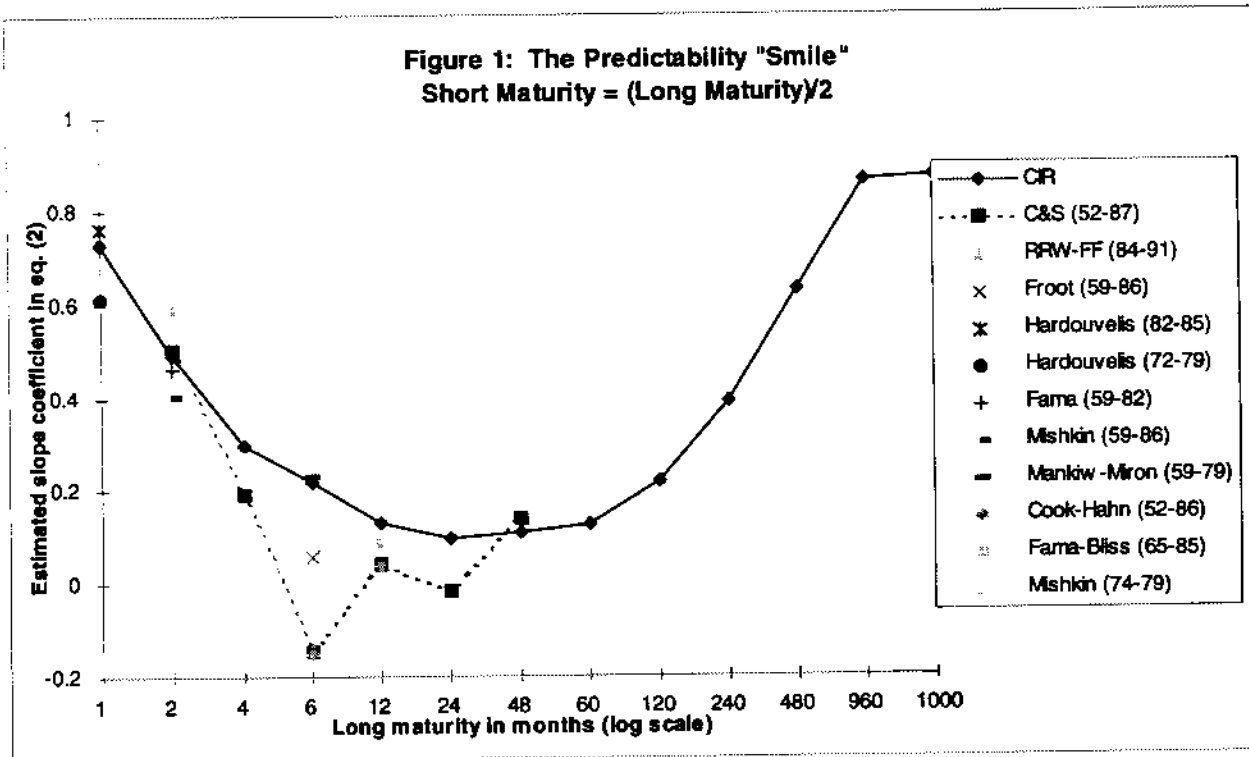
- Backus, D.K., Gregory, A.W., and S.E. Zin (1989), "Risk Premiums in the Term Structure: Evidence from Artificial Economies," *Journal of Monetary Economics* 24 (November):371-400.
- Balduzzi, P., Bertola, G., and S. Foresi (1993), "A Model of Target Changes and the Term Structure of Interest Rates," NBER Working Paper no. 4347.
- Bekaert, G., Hodrick, R.J., and D.A. Marshall, "'Peso Problem' Explanations for Term Structure Anomalies," manuscript, 1995.
- Bliss, R.R. and D.C. Smith (1994), "Stability of Interest Rate Processes," Unpublished Working Paper, Indiana University.
- Campbell, J.Y. (1995), "Some Lessons from the Yield Curve," *Journal of Economic Perspectives* 9:129-52.
- 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.
- Chan, K.C., G.A. Karolyi, F.A. Longstaff, and A.B. Saunders (1992), "An Empirical Comparison of Alternative Models of the Short-Term Interest Rate," *Journal of Finance* 47: 1209-27.
- Chen, R.-R., and L.O. Scott (1992), "Pricing Interest Rate Options in a Two Factor Cox-Ingersoll-Ross Model of the Term Structure," *Review of Financial Studies* 5: 613-36.
- Chen, R.-R., and L.O. Scott (1993), "Maximum Likelihood Estimation for a Multifactor Equilibrium Model of the Term Structure of Interest Rates," *Journal of Fixed Income* 3: 14-31.
- Cox, J.C., Ingersoll, J.E., and S.A. Ross (1985), "A Theory of the Term Structure of Interest Rates," *Econometrica* 53: 385-407.
- Dotsey, M. and C. Otrok (1995), "The Rational Expectations Hypothesis of the Term Structure, Monetary Policy, and Time-Varying Term Premia," Federal Reserve Bank of Richmond *Economic Quarterly*: 65-81.
- Fama, E.F. (1984), "The Information in the Term Structure," *Journal of Financial Economics* 13: 509-28.
- Fama, E.F., and Bliss, R.R. (1987), "The Information in Long-Maturity Forward Rates," *American Economic Review* 77: 680-92.



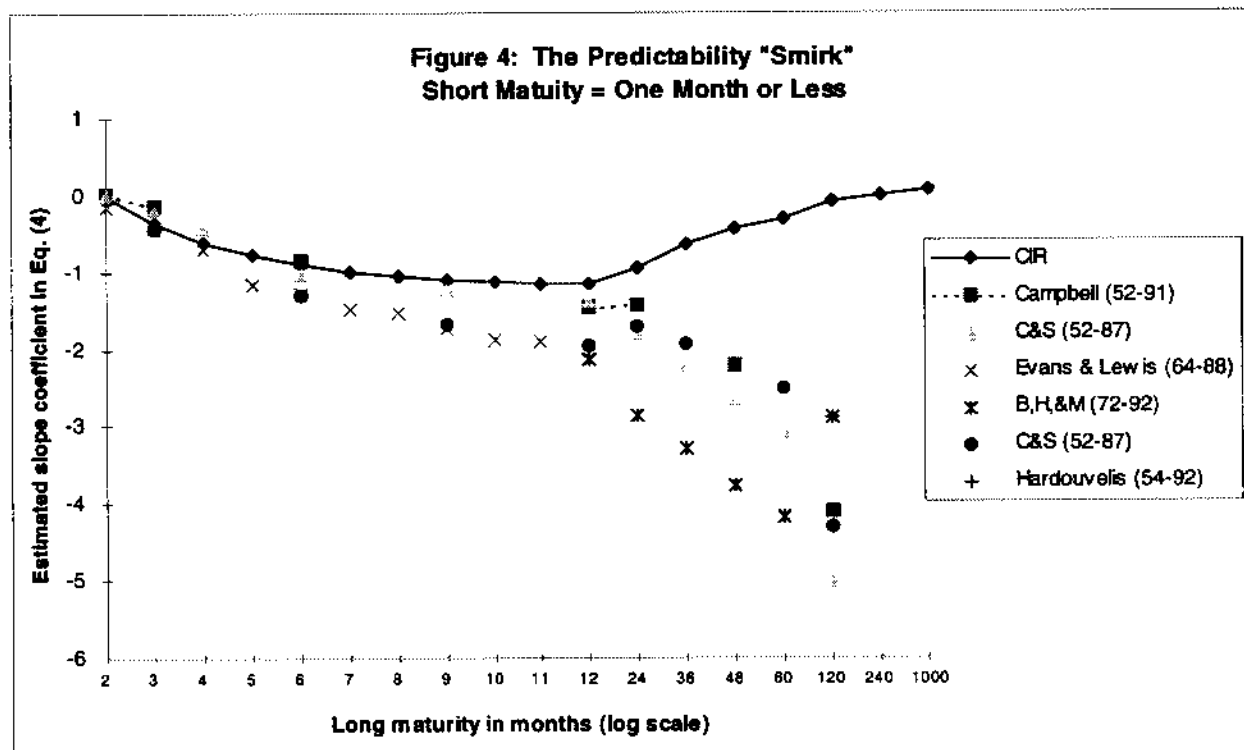
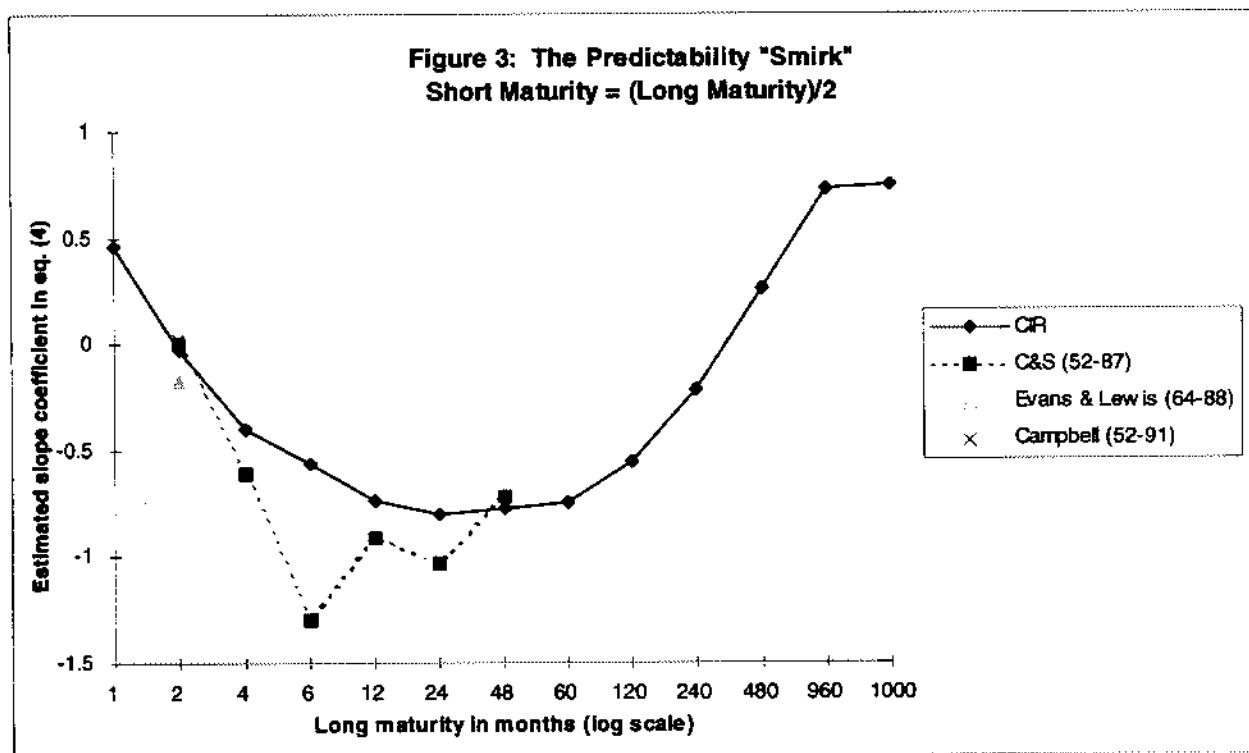
- Hardouvelis, G.A. (1988), "The Predictive Power of the Term Structure During Recent Monetary Regimes," *Journal of Finance* 43: 339-56.
- Hardouvelis, G.A. (1994), "The Term Structure Spread and Future Changes in Long and Short Rates in the G7 Countries," *Journal of Monetary Economics* 33: 255-83.
- Kugler, P. (1988), "An Empirical Note on the Term Structure and Interest Rate Stabilization Policies," *Quarterly Journal of Economics* 103: 789-92.
- Kugler, P. (1990), "The Term Structure of Euro Interest Rates and Rational Expectations," *Journal of International Money and Finance* 9: 234-44.
- Longstaff, F.A. and E.S. Schwartz (1992), "Interest Rate Volatility and the Term Structure: A Two-Factor General Equilibrium Model," *Journal of Finance* 47: 1259-82.
- Macaulay, F. R. (1938), *Some Theoretical Problems Suggested by the Movements of Interest Rates, Bond Yields, and Stock Prices in the United States Since 1856*. New York: National Bureau of Economic Research.
- McCallum, B. (1994), "Monetary Policy and the Term Structure of Interest Rates," National Bureau of Economic Research working paper no. 4938.
- Mankiw, N.G., and J.A. Miron (1986), "The Changing Behavior of the Term Structure of Interest Rates," *Quarterly Journal of Economics* 101: 211-28.
- Mankiw, N.G., and L.H. Summers (1984), "Do Long-Term Interest Rates Overreact to Short-Term Interest Rates?" *Brookings Papers on Economic Activity* (1984:1): 223-42.
- Marty, R. (1990), "The Term Structure of Euro Interest Rates: an Empirical Investigation," *Schweizerische Zeitschrift für Volkswirtschaft und Statistik* 1: 51-61.
- Mishkin, F.S. (1988), "The Information in the Term Structure: Some Further Results," *Journal of Applied Econometrics* 3: 307-14.
- Pagan, A.R., A.D. Hall, and V. Martin (1994), "Exploring the Relations Between the Finance and Econometrics Literatures on the Term Structure," Unpublished Working Paper, Australian National University.
- Roberds, W., Runkle, D., and C.H. Whiteman (1996), "Resurrecting the Term Structure: A Daily View of Yield Spreads and Short-Term Interest Rate Movements," *Journal of Money, Credit, and Banking* 28: 34-53.
- Rudebusch, G.D. (1995), "Federal Reserve Interest Rate Targeting, Rational Expectations, and the Term Structure," *Journal of Monetary Economics* 35: 245-74.

Simon, D.P. (1990), "Expectations and the Treasury Bill-Federal Rate Spread over Recent Monetary Policy Regimes," *Journal of Finance* 45: 467-77.

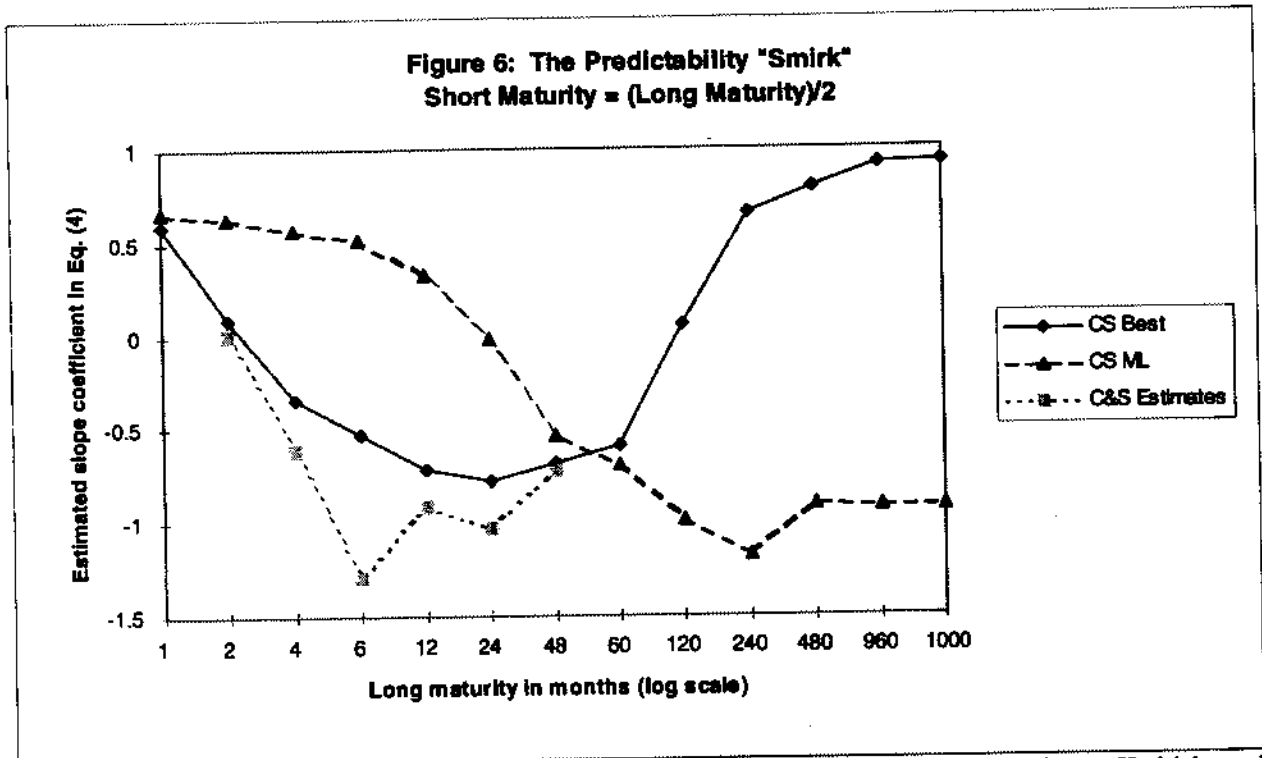
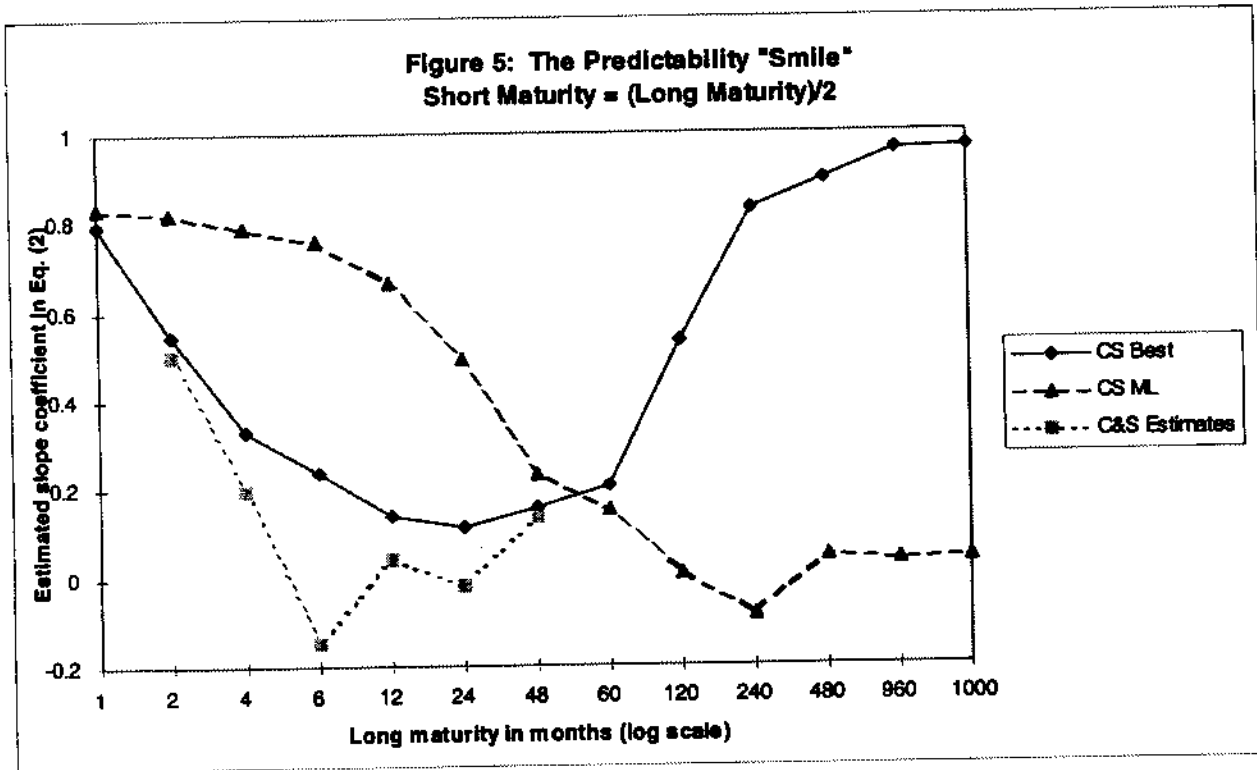
Wood J.H. (1964), "The Expectations Hypothesis, the Yield Curve and Monetary Policy," *Quarterly Journal of Economics* 78: 457-70.



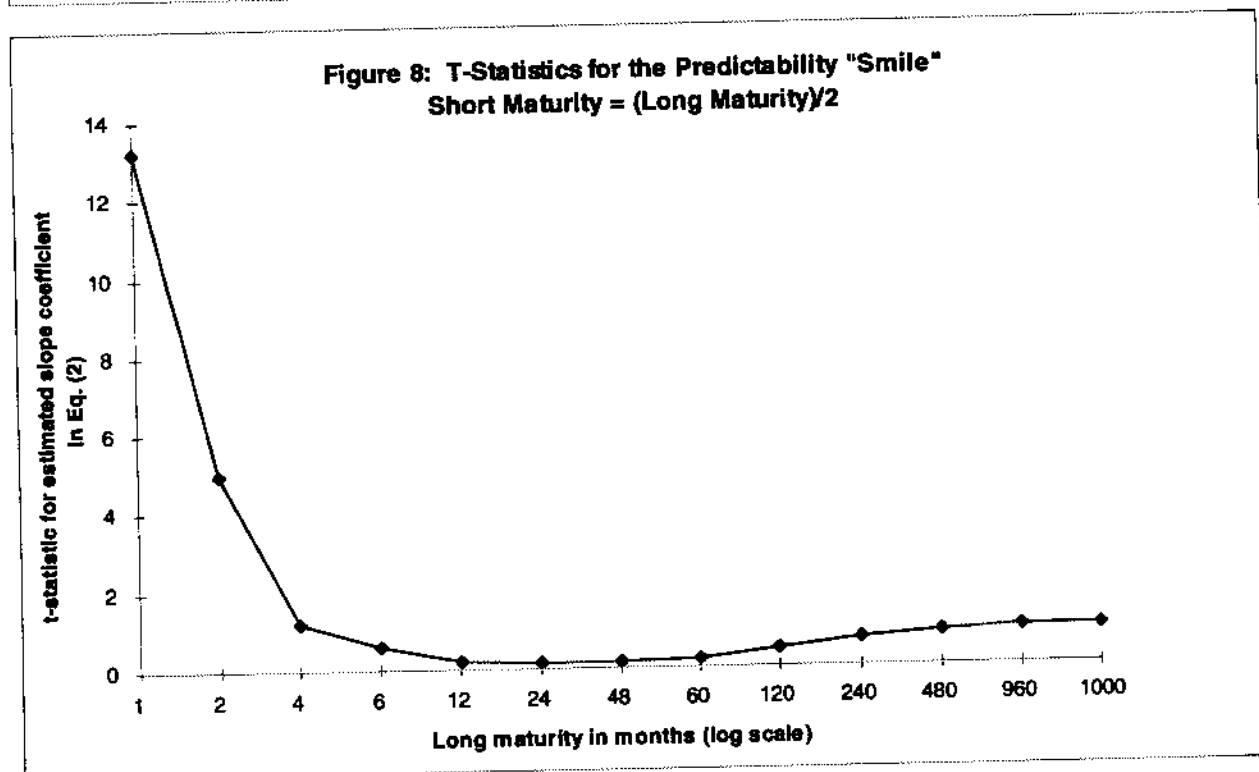
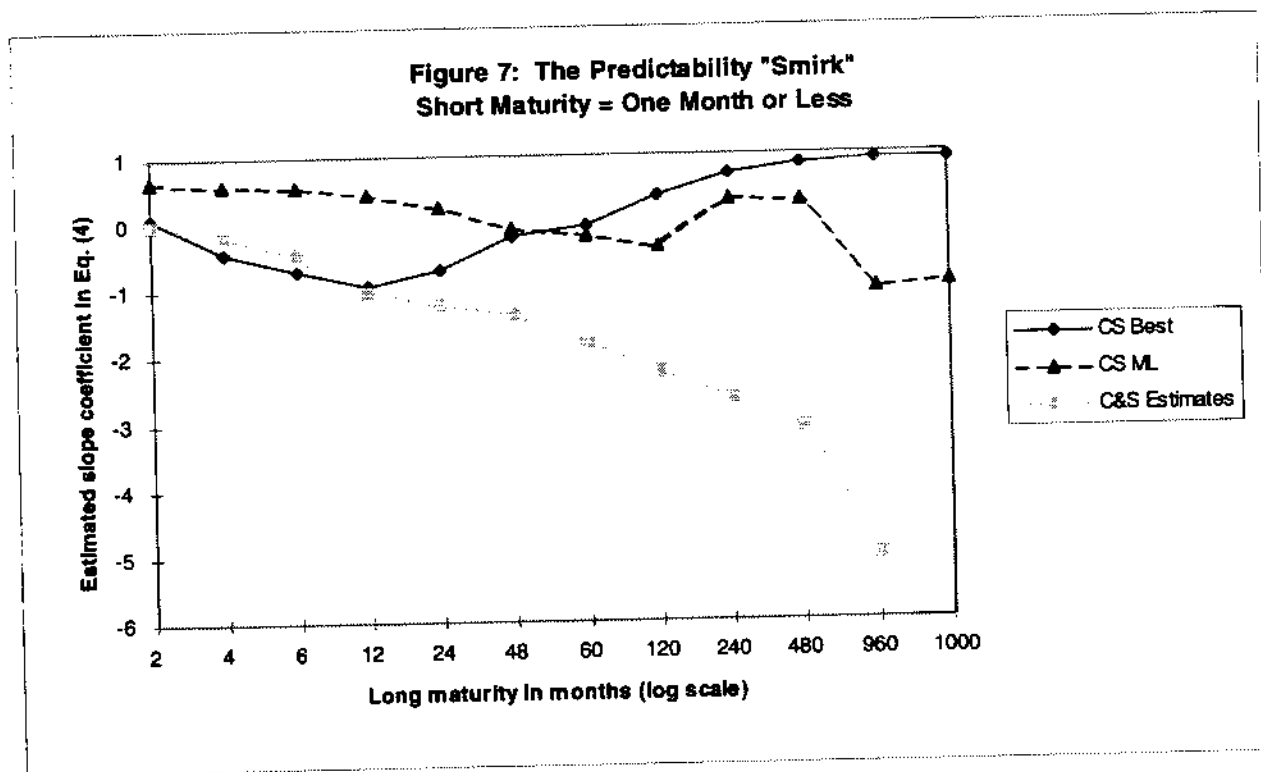
Sources: Fama (1984), Mankiw and Miron (1988), Fama and Bliss (1987), Hardouvelis (1988), Mishkin (1988), Froot (1989), Cook and Hahn (1990), Simon (1990), Campbell and Shiller (1991), Campbell (1995), Roberds, Runkle and Whiteman (1996).



Sources: Campbell and Shiller (1991), Evans and Lewis (1994), Hardouvelis (1994), Bekaert, Hodrick, and Marshall (1995), Campbell (1995).



Sources: Campbell and Shiller (1991), Evans and Lewis (1994), Hardouvelis (1994), Bekaert, Hodrick, and Marshall (1995), Campbell (1995).



Sources: Campbell and Shiller (1991) and calculations by the authors.

**Appendix A: Some Analytical Results concerning the Slope Coefficient in Equation (2)**

We consider the limiting behavior of  $\beta$  in the regression (2)

$$\frac{1}{J} \sum_{i=0}^{J-1} R_{t+mi,m} - R_{t,m} = \alpha + \beta(R_{t,n} - R_{t,m}) + e_{t+mJ,m}$$

where  $n=Jm$  and the term structure follows the CIR model. Two special cases considered in the paper are the first case where  $2m=n$ , for which

$$\beta = \beta\left(n, \frac{n}{2}\right) \equiv \frac{B\left(\frac{n}{2}\right)}{B(n) - 2B\left(\frac{n}{2}\right)} \left(e^{-\frac{\kappa n}{2}} - 1\right) \quad (\text{A1})$$

and the second case for which  $m=\text{one "small" time unit and } n>1$ , for which

$$\beta = \hat{\beta}(n, m) \equiv \frac{\frac{1}{n\kappa}(1 - e^{-\kappa n}) - \frac{B(m)}{m}}{\frac{B(n)}{n} - \frac{B(m)}{m}} \quad (\text{A2})$$

where  $\kappa$  is the mean reversion parameter from the CIR model and  $B(n)$  is given by the standard formula

$$B(n) = \frac{2(e^{\kappa n} - 1)}{(\gamma + \lambda + \kappa)(e^{\kappa n} - 1) + 2\gamma} \quad (\text{A3})$$

The following facts concerning  $B(n)$  are immediate and useful in deriving the results that follow:

$$B(\infty) \equiv \lim_{n \rightarrow \infty} B(n) = \frac{2}{\kappa + \gamma + \lambda} \quad (\text{A4})$$

$$B(0) = 0 \quad (\text{A5})$$

$$B'(n)|_{n=0} = 0 \quad (\text{A6})$$

$$\lim_{n \downarrow 0} \frac{B(n)}{n} = 1 \quad (\text{A7})$$

We can now state and prove the following results concerning  $\beta$ :

**Result 1:**

$$\lim_{m \downarrow 0} \hat{\beta}(n, m) = \frac{\frac{1}{n\kappa}(1 - e^{-\kappa n}) - 1}{\frac{B(n)}{n} - 1}$$

Proof: Follows immediately from (A6).

**Result 2:**

$$\lim_{n \rightarrow \infty} \hat{\beta}(n, m) = 1$$

$$\text{Proof: } \lim_{n \rightarrow \infty} \hat{\beta}(n, m) = \frac{\lim_{n \rightarrow \infty} \left[ \frac{1}{n\kappa}(1 - e^{-\kappa n}) \right] - \frac{B(m)}{m}}{\lim_{n \rightarrow \infty} \left[ \frac{B(n)}{n} \right] - \frac{B(m)}{m}} = \frac{\frac{B(m)}{m}}{\frac{B(m)}{m}} = 1.$$

**Result 3:**

$$\lim_{n \rightarrow \infty} \beta\left(n, \frac{n}{2}\right) = 1$$

$$\text{Proof: } \lim_{n \rightarrow \infty} \beta\left(n, \frac{n}{2}\right) = \frac{B(\infty)}{B(\infty) - 2B(\infty)}(-1) = 1.$$

**Result 4:**

$$\lim_{n \downarrow 0} \lim_{m \downarrow 0} \hat{\beta}(n, m) = \frac{\kappa}{\kappa + \lambda}$$

Proof: From result 1,

$$\lim_{n \downarrow 0} \lim_{m \downarrow 0} \hat{\beta}(n, m) = \lim_{n \downarrow 0} \frac{\frac{1}{\kappa}(1 - e^{-\kappa n}) - n}{B(n) - n}$$

Since both the numerator and denominator of this limit tend to zero, we evaluate the limit using L'Hôpital's rule (twice). First we note that for the numerator of the expression above



$$\frac{d^2}{dn^2}(\text{numerator}) = -\kappa e^{-m} \Rightarrow \frac{d^2}{dn^2}(\text{numerator})\Big|_{n=0} = -\kappa$$

Also, for the denominator

$$\frac{d^2}{dn^2}(\text{denominator}) = \frac{[(\gamma + k + \lambda)(e^m - 1) + 2\gamma]^2 4\gamma^3 e^m - 8\gamma^2 e^m [(\gamma + k + \lambda)(e^m - 1) + 2\gamma]}{[(\gamma + \kappa + \lambda)(e^m - 1) + 2\gamma]^4}$$

$$\frac{d^2}{dn^2}(\text{denominator})\Big|_{n=0} = \frac{16\gamma^4 [\gamma - (\gamma + \kappa + \lambda)]}{16\gamma^4} = -(\kappa + \lambda)$$

Taking the ratio of the two derivatives gives the desired limit.

*Result 5:*

$$\lim_{n \downarrow 0} \beta\left(n, \frac{n}{2}\right) = \frac{\kappa}{\kappa + \lambda}$$

Proof: By definition,  $\lim_{n \downarrow 0} \beta\left(n, \frac{n}{2}\right) = \lim_{n \downarrow 0} \left[ \frac{B\left(\frac{n}{2}\right) \left(e^{\frac{m}{2}} - 1\right)}{B(n) - 2B\left(\frac{n}{2}\right)} \right]$ . We again employ L'Hôpital's rule.

$$\frac{d^2}{dn^2}(\text{numerator})\Big|_{n=0} = \frac{1}{2} B'\left(\frac{n}{2}\right) \frac{d}{d(n/2)} \left( e^{\frac{m}{2}} - 1 \right) \Big|_{n=0} = \frac{1}{2} (1)(-\kappa) = -\frac{\kappa}{2}$$

Recall from Result 4 that

$$\lim_{n \downarrow 0} B''(n) = -(\kappa + \lambda)$$

Which implies that

$$\frac{d^2}{dn^2}(\text{denominator})\Big|_{n=0} = B''(n)\Big|_{n=0} - 2\left(\frac{1}{4}\right) B''\left(\frac{n}{2}\right)\Big|_{n=0} = -(\kappa + \lambda) + \frac{1}{2}(\kappa + \lambda) = -\frac{(\kappa + \lambda)}{2}$$

Taking the ratio of the two derivatives gives the desired limit.

**Appendix B: Some Analytical Results concerning the Slope Coefficient in Equation (4)**

We consider the limiting behavior of  $\delta$  in the regression (4)

$$R_{t+m,n-m} - R_{t,n} = \gamma + \delta \frac{m}{n-m} (R_{t,n} - R_{t,m}) + u_{t+m,n-m}, \quad n > m$$

When the term structure is determined by CIR, then

$$\delta = \delta(n, m) \equiv \frac{n-m}{m} \frac{e^{-\kappa m} \frac{B(n-m)}{n-m} - \frac{B(n)}{n}}{\frac{B(n)}{n} - \frac{B(m)}{m}} \quad (\text{B1})$$

where  $\kappa$  is the mean reversion parameter from the CIR model and  $B(n)$  is given in equation (A3).

*Result 1:*

$$\lim_{n \rightarrow \infty} \delta\left(n, \frac{n}{2}\right) = 1$$

*Proof:* By definition,

$$\delta\left(n, \frac{n}{2}\right) = \frac{2e^{-\frac{\kappa n}{2}} B\left(\frac{n}{2}\right) - B(n)}{B(n) - 2B\left(\frac{n}{2}\right)} \Rightarrow \lim_{n \rightarrow \infty} \delta\left(n, \frac{n}{2}\right) = \frac{2 \cdot 0 \cdot B(\infty) - B(\infty)}{B(\infty) - 2B(\infty)} = 1.$$

*Result 2:*

$$\lim_{m \downarrow 0} \delta(n, m) = \frac{(1 - \kappa n)B(n) - nB'(n)}{B(n) - n}$$

*Proof:* Using (A7) we obtain

$$\lim_{m \downarrow 0} \delta(n, m) = \frac{\lim_{m \downarrow 0} \left\{ \left( \frac{n-m}{m} \right) \left[ e^{-\kappa m} \frac{B(n-m)}{n-m} - \frac{B(n)}{n} \right] \right\}}{\lim_{m \downarrow 0} \left[ \frac{B(n)}{n} - \frac{B(m)}{m} \right]} = \frac{n \lim_{m \downarrow 0} \left\{ \frac{1}{m} \left[ e^{-\kappa m} \frac{B(n-m)}{m} - \frac{B(n)}{n} \right] \right\}}{\frac{B(n)}{n} - 1}$$

To evaluate the above limit, we evaluate the numerator using L'Hôpital's rule, i.e.,

$$\begin{aligned}
 \lim_{m \downarrow 0} \left\{ \frac{1}{m} \left[ e^{-\kappa m} \frac{B(n-m)}{n-m} - \frac{B(n)}{n} \right] \right\} &= \frac{\lim_{m \downarrow 0} \frac{d}{dm} \left[ e^{-\kappa m} \frac{B(n-m)}{n-m} - \frac{B(n)}{n} \right]}{\lim_{m \downarrow 0} \frac{dm}{dm}} \\
 &= \frac{d}{dm} \left[ e^{-\kappa m} \frac{B(n-m)}{n-m} - \frac{B(n)}{n} \right] \Big|_{m=0} = -\frac{B'(n)}{n} + B(n) \frac{d}{dm} \left( \frac{e^{-\kappa m}}{n-m} \right) \Big|_{m=0} \\
 &= -\frac{B'(n)}{n} + B(n) \left( \frac{1-\kappa n}{n^2} \right)
 \end{aligned}$$

Dividing the last term by  $[B(n)/n]-1$  gives the desired result.

*Result 3:*

$$\lim_{n \downarrow 0} \lim_{m \downarrow 0} \delta(n, m) = \frac{\kappa - \lambda}{\kappa + \lambda}$$

*Proof:* From Result 2,

$$\lim_{n \downarrow 0} \lim_{m \downarrow 0} \delta(n, m) = \lim_{n \downarrow 0} \frac{(1-\kappa n)B(n) - nB'(n)}{B(n) - n}$$

To evaluate this ratio we successively apply L'Hôpital's rule, yielding

$$\frac{d^2}{dn^2} (B(n) - n) = B''(n) \Rightarrow \frac{d^2}{dn^2} (B(n) - n) \Big|_{n=0} = -(\kappa + \lambda)$$

from Appendix A, Result 4, and

$$\begin{aligned}
 \frac{d^2}{dn^2} [(1-\kappa n)B(n) - nB'(n)] &= B''(n) - \kappa \frac{d}{dn} [B(n) + nB'(n)] - \frac{d}{dn} [B'(n) + nB''(n)] \\
 &= B''(n) - 2\kappa B'(n) - \kappa n B''(n) - 2B''(n) - nB'''(n) \\
 &\Rightarrow \frac{d^2}{dn^2} [(1-\kappa n)B(n) - nB'(n)] \Big|_{n=0} = -B''(n) \Big|_{n=0} - 2\kappa B'(n) \Big|_{n=0} = \kappa + \lambda - 2\kappa = -(\kappa - \lambda)
 \end{aligned}$$

Taking the ratio of the two derivatives gives the required result.

*Result 4:*

$$\lim_{n \rightarrow \infty} \lim_{m \downarrow 0} \delta(n, m) = \frac{2\kappa}{\kappa + \lambda + \gamma}$$

**Proof:** Using (A4), Result 2, and the easily verified fact that  $\lim_{n \rightarrow \infty} B'(n) = 0$ ,

$$\lim_{n \rightarrow \infty} \lim_{m \downarrow 0} \delta(n, m) = \lim_{n \rightarrow \infty} \frac{\frac{B(n)}{n} - \kappa B(n) - B'(n)}{\frac{B(n)}{n} - 1} = \frac{0 - \kappa B(\infty) - 0}{0 - 1} = \frac{\kappa}{\kappa + \lambda + \gamma}.$$

### Appendix C. Notes on the Mankiw-Miron (1986) Decomposition of the Slope Coefficients in Equations (2) and (4) when $n=2m$ .

In the case that the maturity of the long rate ( $n$ ) is twice that of the short rate ( $m$ ), equation (2) takes the form

$$\frac{1}{2} [R_{t+n/2, n/2} - R_{t, n/2}] = \alpha + \beta (R_{t, n} - R_{t, n/2}) + e_{t+n/2, n/2} \quad (C1)$$

As noted by Mankiw-Miron (1986, p.) and others, we can decompose the yield spread  $R_{t, n} - R_{t, n/2}$  as a weighted sum of an expected future change in the short rate plus a risk premium, i.e.,

$$R_{t, n} - R_{t, n/2} = \frac{1}{2} E_t (R_{t+n/2, n/2} - R_{t, n/2}) + \theta_{t, n}$$

where the equation above defines the risk premium  $\theta_{t, n}$ . We can similarly decompose the LHS of equation (C1) as

$$\frac{1}{2} (R_{t+n/2, n/2} - R_{t, n/2}) = \frac{1}{2} E_t \Delta R_{t+n/2, n/2} + \frac{1}{2} \varepsilon_{t+n/2, n/2}$$

where  $\Delta$  is the gapped difference operator  $(1 - L^{n/2})$  and  $\varepsilon_{t+n/2, n/2}$  is the forecast error. Since the forecast error is uncorrelated with the yield spread  $(R_{t, n} - R_{t, n/2})$ , we can rewrite  $\beta$  in equation (C1) as

$$\beta = \frac{\text{cov}(LHS, RHS)}{\text{var}(RHS)}$$

where

$$\text{cov}(LHS, RHS) = \frac{1}{4} \text{var}(E_t \Delta R_{t+n/2, n/2}) + \frac{1}{2} \text{cov}(\Delta R_{t+n/2, n/2}, \theta_{t, n})$$

and

$$\text{var}(RHS) = \frac{1}{4} \text{var}(E_t \Delta R_{t+n/2, n/2}) + \frac{1}{2} \text{cov}(E_t \Delta R_{t+n/2, n/2}, \theta_{t, n}) + \text{var}(\theta_{t, n})$$

Which implies that  $\beta$  may be written as

$$\beta = \frac{1 + 2\rho q}{1 + 2\rho q + 4q^2} \quad (C2)$$

where

$$\rho \equiv \text{corr}(E_t \Delta R_{t+n/2, n/2}, \theta_{t, n})$$

$$q \equiv \left( \frac{\text{var}(\theta_{t, n})}{\text{var}(E_t \Delta R_{t+n/2, n/2})} \right)^{\frac{1}{2}}$$

If the term structure follows the CIR model, then the following must hold (ignoring constant terms)

$$E_t \Delta R_{t+n/2, n/2} = \frac{2}{n} \left( e^{-\frac{kn}{2}} - 1 \right) B\left(\frac{n}{2}\right) r(t)$$

$$\theta_{t, n/2} = R_{t, n} - R_{t, n/2} - \frac{1}{2} E_t \Delta R_{t+n/2, n/2} = \frac{1}{n} \left[ B(n) - 2B\left(\frac{n}{2}\right) - \left( e^{-\frac{kn}{2}} - 1 \right) B\left(\frac{n}{2}\right) \right] r(t)$$

$$= \frac{1}{n} \left[ B(n) - \left( e^{-\frac{kn}{2}} + 1 \right) B\left(\frac{n}{2}\right) \right] r(t)$$

where  $r(t)$  is the instantaneous rate and  $B(n)$  is given by (7) in the text. Hence the correlation between the expected change in the short rate and the term premium will be given by  $\rho = \pm 1$  according to the sign of

$$B(n) - \left( e^{-\frac{kn}{2}} + 1 \right) B\left(\frac{n}{2}\right) \tag{C3}$$

For the range of the parameters  $(\kappa, \lambda, \sigma^2)$  considered above,  $\text{sgn}(\text{C3}) = -1$  which implies  $\rho = 1$ . Substituting into (C2) we obtain

$$\beta = \frac{1+q}{1+2q+q^2} = \frac{1}{1+q} \tag{C4}$$

Hence a "smile" pattern in  $\beta$  is implied by a sharp peak in  $q$  at intermediate maturities.

We can also use  $q$  to characterize the slope coefficient  $\delta$  in equation (4) when  $n=2m$ , i.e., in

$$R_{t+n/2, n/2} - R_{t, n} = \gamma + \delta (R_{t, n} - R_{t, n/2}) + u_{t+n/2, n/2} \tag{C5}$$

Rewrite this equation as

$$\begin{aligned}
 R_{t+n/2,n/2} - R_{t,n/2} - R_{t,n} + R_{t,n/2} &= \gamma + \delta(R_{t,n/2} - R_{t,n/2}) + u_{t+n/2,n/2} \\
 \Leftrightarrow E_t \Delta R_{t+n/2,n/2} + \frac{1}{2} \varepsilon_{t+n/2,n/2} - \left( \theta_{t,n} + \frac{1}{2} E_t \Delta R_{t+n/2,n/2} \right) &= \gamma + \delta \left( \theta_{t,n} + \frac{1}{2} E_t \Delta R_{t+n/2,n/2} \right) + u_{t+n/2,n/2} \\
 \Leftrightarrow \frac{1}{2} E_t \Delta R_{t+n/2,n/2} - \theta_{t,n} &= \gamma + \delta \left( \theta_{t,n} + \frac{1}{2} E_t \Delta R_{t+n/2,n/2} \right) + u_{t+n/2,n/2} - \frac{1}{2} \varepsilon_{t+n/2,n/2} \quad (C6)
 \end{aligned}$$

Using (C6) we can write  $\delta$  as

$$\delta = \frac{\text{cov}(LHS, RHS)}{\text{var}(RHS)}$$

where

$$\begin{aligned}
 \text{cov}(LHS, RHS) &= \frac{1}{4} \text{var}(E_t \Delta R_{t+n/2,n/2}) - \text{var}(\theta_{t,n}) \\
 \text{var}(RHS) &= \frac{1}{4} \text{var}(E_t \Delta R_{t+n/2,n/2}) + \text{var}(\theta_{t,n}) + \text{cov}(E_t \Delta R_{t+n/2,n/2}, \theta_{t,n})
 \end{aligned}$$

As was the case with  $\beta$ , we can further simplify the expression for  $\delta$  to

$$\delta = \frac{1-4q^2}{1+4q^2+4\rho q} = \frac{1-2q}{1+2q} \quad \text{when } \rho = 1 \quad (C7)$$

Equation (C7) implies that  $\delta$  is decreasing in  $q$ . Hence, when  $q$  is peaked at a maturity around  $n=1$  year,  $\delta$  will have a minimum at this maturity.

**Appendix D. Notes on the Mankiw-Miron (1986) Decomposition of the Slope Coefficients in Equations (2) and (4) when  $m \approx 0$  and  $n$  is Large.**

We first consider the decomposition of  $\beta$  in equation (2), i.e., in the regression

$$\frac{1}{J} \sum_{i=0}^{J-1} R_{t+mi,m} - R_{t,m} = \alpha + \beta(R_{t,n} - R_{t,m}) + e_{t+mJ,m}$$

where  $n=Jm$ . Assume for simplicity that  $m \approx 0$  so that we can approximate (2) by its limit as  $m \downarrow 0$ , i.e.,

$$\frac{1}{n} \int_0^n r(t+\tau) d\tau - r(t) = \alpha + \beta[R_{t,n} - r(t)] + e(t,n) \quad (\text{D1})$$

Define the risk premium  $\theta(t,n)$  as

$$\theta(t,n) \equiv R_{t,n} - \frac{1}{n} \int_0^n E_t r(t+\tau) d\tau$$

It is convenient to rewrite (D1) as

$$\Delta(t,n) + \varepsilon(t,n) = \alpha + \beta[\theta(t,n) + \Delta(t,n)] + e(t,n) \quad (\text{D2})$$

where

$$\Delta(t,n) \equiv \frac{1}{n} \int_0^n E_t r(t+\tau) d\tau - r(t)$$

and  $\varepsilon(t,n)$  is the forecast error associated with  $\Delta$ . The form of (D2) makes it evident that we can write  $\beta$  as

$$\beta = \frac{\text{var}(\Delta) + \text{cov}(\theta, \Delta)}{\text{var}(\Delta) + 2 \text{cov}(\theta, \Delta) + \text{var}(\theta)} = \frac{1 + \rho q}{1 + 2\rho q + q^2} \quad (\text{D3})$$

where

$$\rho = \text{corr}(\theta, \Delta)$$

$$q = \left[ \frac{\text{var}(\theta)}{\text{var}(\Delta)} \right]^{\frac{1}{2}}$$



In the CIR model (ignoring constant terms)

$$\Delta(t, n) = \frac{1}{n} \int_0^n E_t r(t + \tau) d\tau - r(t) = \frac{1}{n} \int_0^n e^{-\kappa\tau} r(t) d\tau - r(t) = \left[ \frac{1}{n\kappa} (1 - e^{-n\kappa}) - 1 \right] r(t) \quad (\text{D4})$$

$$\theta(t, n) = \left[ \frac{B(n)}{n} - \frac{1}{n\kappa} (1 - e^{-n\kappa}) \right] r(t) \quad (\text{D5})$$

For the relevant range of parameter values, (D4) and (D5) imply that  $\rho=1$ , which implies that (D3) simplifies to

$$\beta = \frac{1}{1+q} \quad (\text{D6})$$

where

$$q = \frac{\frac{B(n)}{n} - \left( \frac{1}{n\kappa} (1 - e^{-n\kappa}) \right)}{\frac{1}{n\kappa} (1 - e^{-n\kappa}) - 1}$$

As was the case where  $n=2m$  [cf. equation (C4)], a sharp peak in  $q$  at  $n=1$  year implies a predictability "smile."

Next consider the decomposition of  $\delta$  in equation (4), i.e., in

$$R_{t+m,n-m} - R_{t,n} = \gamma + \delta \frac{m}{n-m} (R_{t,n} - R_{t,m}) + u_{t+m,n-m}, \quad n > m$$

Begin by rewriting the scaled spread  $(R_{t,n} - R_{t,m})$  as

$$\frac{m}{n-m} (R_{t,n} - R_{t,m}) = E_t R_{t+m,n-m} - R_{t,n} + \Psi_t(n, m) \quad (\text{D7})$$

where (D7) defines the risk premium term  $\Psi_t(n, m)$ . Substituting (D7) into (4), we obtain

$$\Gamma_t(n, m) = \gamma + \delta (\Gamma_t(n, m) + \Psi_t(n, m)) + u_{t+m,n-m} + (\text{forecast error})$$

where

$$\Gamma_t(n, m) \equiv E_t R_{t+m,n-m} - R_{t,n}$$

Using the same strategy as with  $\beta$ , from (D7) we can write  $\delta$  as

$$\delta = \frac{\text{var}(\Gamma) + \text{cov}(\Psi, \Gamma)}{\text{var}(\Gamma) + 2\text{cov}(\Psi, \Gamma) + \text{var}(\Psi)} = \frac{1 + \vartheta Q}{1 + 2\vartheta Q + Q^2} \quad (\text{D8})$$

where

$$\vartheta = \text{corr}(\Gamma, \Psi)$$

$$Q = \left[ \frac{\text{var}(\Psi)}{\text{var}(\Gamma)} \right]^{\frac{1}{2}}$$

In determining the values of  $\vartheta$  and  $Q$ , it is convenient to scale both  $\Gamma$  and  $\Psi$  by  $(n-m)/m$  and drive  $m \downarrow 0$ , yielding the following expressions for an infinitesimal short maturity  $m$  (cf. Appendix B, result 2)

$$\frac{n-m}{m} \Gamma_t(n, m) = \left[ \left( \frac{1}{n} - \kappa \right) B(n) - B'(n) \right] r(t)$$

$$\frac{n-m}{m} \Psi_t(n, m) = \left( \frac{B(n)}{n} - 1 \right) r(t) - \left( \frac{B(n)}{n} - \kappa B(n) - B'(n) \right) r(t) = [B'(n) + \kappa B(n) - 1] r(t)$$

For the relevant parameter values  $\vartheta = -1$ , implying that (D8) reduces to

$$\delta = \frac{1}{1 - Q} \quad (\text{D9})$$

where

$$Q = \left| \frac{B'(n) + \kappa B(n) - 1}{\frac{B(n)}{n} - \kappa B(n) - B'(n)} \right|$$

Hence in order to obtain strongly negative values of  $\delta$ , we need to obtain  $Q$  slightly greater than unity.