Journal of Financial Economics 13 (1984) 529–546. North-Holland

TERM PREMIUMS IN BOND RETURNS

Eugene F. FAMA*

University of Chicago, Chicago, IL 60637, USA

Received November 1983, final version received June 1984

This paper examines expected returns on U.S. Treasury bills and on U.S. Government bond portfolios. Expected bill returns are estimated from forward rates and from sample average returns. Both estimation methods indicate that expected returns on bills tend to peak at eight or nine months and never increase monotonically out to twelve months. Reliable inferences are limited to Treasury bills and thus to maturities up to a year. The variability of longer-term bond returns preempts precise conclusions about their expected returns.

1. Introduction

In the early literature on the term structure of interest rates there is controversy about the existence of risk premiums in the expected returns on longer-term bonds. [See, for example Meiselman (1962) and Kessel (1965).] Recent empirical work documents reliable premiums in the expected returns on longer- versus shorter-maturity instruments. [See, for example, Roll (1970), McCulloch (1975), Fama (1976a, 1976b, 1984) and Startz (1982).] Except for McCulloch (1975), however, the recent work focuses on short-maturity instruments, in particular, U.S. Treasury bills.

This paper examines returns on bills and on portfolios of U.S. Government bonds that cover the maturity spectrum. Consistent with the liquidity preference hypothesis advanced by Kessel (1965) and others, we find statistically reliable evidence that expected returns on longer-term bills exceed the returns on one-month bills. We also find reliable evidence that expected returns on bills do not increase monotonically with maturity and generally tend to peak at eight or nine months. This conclusion is first obtained in tests on average returns, and it is then reinforced by estimates of expected returns from forward rates. The conclusion is inconsistent with the liquidity preference hypothesis which predicts that expected returns increase monotonically with maturity

*Theodore O. Yntema Professor of Finance, Graduate School of Business, University of Chicago. The comments of David Booth, Stephen Buser, Nai-Fu Chen, George Constantinides, Wayne Ferson, Robert Holthausen, David Hsieh, G. William Schwert, Rex Sinquefield, Robert Stambaugh, and the referee, Michael Gibbons, are gratefully acknowledged. This research is supported by the National Science Foundation.

because return variability increases monotonically. Moreover, since the *ex ante* estimates of expected returns from forward rates peak at the same maturities as *ex post* average returns, the non-monotonicity of average returns cannot be attributed to unexpected increases in the level of interest rates during the sample period.

During the five-year subperiods of 1953–82 covered by the bond portfolios, bonds with maturities greater than four years never have the highest average returns. During periods where the bond file overlaps with the bill file, the highest average return on a bond portfolio never exceeds the highest average return on a bill. We cannot conclude, however, that longer-term bonds have lower expected returns than short-term instruments. Like McCulloch (1975), but with the advantage of an exhaustive data base, we find that the high variability of longer-term bond returns preempts precise conclusions about their expected returns. The bond data are consistent with maturity structures of expected returns that are flat, upward sloping or downward sloping beyond a year.

The non-monotonicity of expected bill returns documented here is inconsistent with either the pure expectations hypothesis or the liquidity preference hypothesis of the classical term structure literature. It is not necessarily inconsistent with the more sophisticated models of capital market equilibrium of Sharpe (1964), Lintner (1965), Merton (1973) or Breeden (1979). I offer no direct tests. Rather, the descriptive evidence presented here, and the evidence of rich patterns of variation through time in expected bill returns presented by Fama (1984), Startz (1982) and others, stand as challenges or 'stylized facts' to be explained by candidate models.

2. The data

The U.S. Government bond file of the Center for Research in Security Prices (CRSP) of the University of Chicago contains monthly price and return information on all outstanding publicly traded U.S. Treasury securities. Two return files are created from these data, covering (1) bills with maturities up to twelve months, and (2) portfolios of bonds and notes covering all maturities.

Treasury bills with twelve months to maturity are consistently available beginning in 1964. On the last trading day of each month the bill with maturity closest to twelve months is chosen. At the end of the next month, this bill is chosen as the eleven-month bill, and at the end of the following month it becomes the ten-month bill, etc. In this way monthly returns on bills with one to twelve months to maturity are obtained at the end of each month.

Because the Treasury issues bonds and notes on an irregular basis, they must be grouped into portfolios to get unbroken time series of monthly returns. For maturities up to three years, grouping within six-month maturity intervals produces six unbroken time series of monthly returns for the 1953–82 period.

Continuous time series are also obtained for portfolios that contain maturities between three and four years, four and five years, and five to ten years. These portfolios contain only ordinary bonds and notes. Bills, 'flower' bonds (bonds redeemable at par to satisfy Federal estate taxes) and other bonds with special tax features are excluded.

To represent the behavior of the longest-term Government bonds, the Government bond return series of Ibbotson and Sinquefield (1983) is used. They choose the ordinary bond closest to twenty years to maturity. No ordinary bond with more than ten years to maturity is available during the October 1962 to January 1972 period. During this period they choose the long-term 'flower' bond with the highest price relative to par, since the returns on such bonds are likely to behave most like those on ordinary bonds.

3. Statistical issues[1]

We test vectors of average returns on bills and bonds against expected return vectors implied by hypotheses about the term structure. We also test individual average returns against their hypothetical expected values. Consistent probability statements require a multiple-comparisons framework.

3.1. Hotelling T^2 tests

The two files of bill and bond returns are treated as separate data sets rather than as a single data set to be viewed in a unified multiple-comparisons framework. The justification is that the two data sets cover different (but overlapping) periods. For each file, Hotelling T^2 statistics are used to test the hypothesis that all expected premiums in bill or bond returns are zero. A premium is defined as the difference between the one-month return on an instrument with a given maturity and the return on a one-month bill. Premiums rather than returns are used as the basic data elements because premiums are less autocorrelated than returns [Fama (1976a, 1984)].[2]

The T^2 tests on the vectors of average premiums for the two data sets produce strong evidence against the hypothesis that all expected premiums are zero in favor of the hypothesis that expected premiums are positive. Next we test the hypothesis that the structure of expected returns is flat beyond two

[1] Later sections of this paper are comprehensible without a detailed understanding of this section.

[2] For example, for the 1964–82 time period covered by the twelve-month bill file, the premiums in two- to twelve-month bill returns show first-order autocorrelation in the neighborhood of 0.2 and no systematic higher-order autocorrelation. For the 1953–82 period covered by the bond portfolio file, the premiums in the returns on shorter-term bond portfolios also show first-order autocorrelation of about 0.2. Premiums in longer-term bond returns show no systematic autocorrelation. Contemporaneous differences or spreads between the returns on adjacent maturity bills or bonds are also important in the tests that follow. These return spreads show little autocorrelation.

months. This involves a T^2 test on the vector of average differences between returns on instruments with successively longer maturities. At least for bill returns, the hypothesis that the structure of expected returns is flat beyond two months is rejected. The final step is to use a multiple-comparisons framework to examine the term structure of expected returns in detail.

The T^2 statistic is the square of the maximum possible t statistic that can be generated from any linear combination of the elements of a vector of means. The multiple-comparisons logic underlying the test is that the distribution of T^2 provides type I error protection for all possible t tests on linear combinations of the means. Thus, the distributional umbrella of the T^2 statistic can be used in an unrestricted search for differences in the expected returns on bills or bonds with different maturities. The right to conduct such a search has a cost. Because an indefinite number of tests is allowed, the differences between expected returns must be large to be identified reliably. Morrison (1976, ch. 4) and Miller (1981, ch. 5) discuss this approach to multiple comparisons, which is due to Roy and Bose (1953).

3.2. *Bonferroni multiple comparisons*

In contrast to the Roy–Bose approach, Bonferroni multiple comparisons require a specification that limits the relevant number of t tests on a vector of means. In particular, the Bonferroni inequality says that when p univariate t statistics are compared to 0.0 and the null hypothesis about the means is true, the probability that one or more of the t statistics is greater than the $1-\alpha$ fractile of the t distribution with $N-1$ degrees of freedom is equal to or less than $p\alpha$.

Table 1 shows large (∞) sample upper fractiles of the t statistic for Bonferroni comparisons of p means for $p = 1, 5, 10, 11$. As intuition suggests, the probability level for a given value of t is smaller the larger the number of means to be compared to zero. It is possible to specify the tests so that the relevant number of means compared to zero is the number of average bill or bond returns calculated. As a consequence, the critical values of t statistics for multiple comparisons carried out under the Bonferroni method are smaller than for the Roy–Bose approach. For example, for large-sample tests on 10 means, the 0.95 fractile of the Bonferroni t statistic is 2.58 (table 1), whereas the Roy–Bose t statistic corresponding to the 0.95 fractile of T^2 for 10 and ∞ degrees of freedom is 4.33. This higher t statistic under the Roy–Bose approach is the price one pays for the right to search over an arbitrary number of linear combinations of means.

To judge the behavior of expected returns for increasing maturities, Bonferroni fractiles are applied to the t statistics for the average differences or spreads between contemporaneous returns on adjacent maturity bills or adja-

Table 1
Upper fractiles of the t distribution for Bonferroni comparisons of p means in large (∞) samples.[a]

	t statistic value				
	1.65	1.96	2.33	2.58	3.29
			Probability level		
$p = 1$	0.95	0.975	0.99	0.995	0.9995
$p = 5$	0.75	0.875	0.95	0.975	0.9975
$p = 10$	0.50	0.75	0.90	0.95	0.995
$p = 11$	0.45	0.725	0.89	0.945	0.9945

[a]The Bonferroni inequality says that when p univariate t statistics are compared to 0.0 and the null hypothesis about the means is true, the probability that one or more of the t statistics is greater than the $1 - \alpha$ fractile of the t distribution with $N - 1$ degrees of freedom is equal to or less than $p\alpha$. Thus, the entries in the table for $p > 1$ follow directly from those for $p = 1$.

cent maturity bond portfolios. These tests are the basis of the conclusions summarized earlier about the behavior of expected returns as a function of maturity.

3.3. *Subperiod results*

For each of the return files, results for the overall period covered by the data and for subperiods are presented. One can argue that, if subperiod results are evaluated with the Bonferroni approach, the relevant number of comparisons is the number of means to be tested each period times the number of periods, and generally the power of the approach is lost. On the other hand, decisions about what constitute relevant 'families' of tests are to some extent at the option of the researcher. [See Miller (1981, pp. 31–35).] One can argue that subperiods can be treated individually under either the Bonferroni approach or the Roy–Bose approach to multiple comparisons.

I follow a middle-of-the-road approach to the subperiod results. Probability statements are limited to the tests for the overall sample period, and subperiod results are used for perspective and diagnostic checks on the tests for the overall period. Although the same summary statistics are used, the subperiod results are not interpreted with the same kinds of probability statements as the results for the overall period.

4. Evidence from average returns on the term structure of expected returns

Let $H\tau_{t+1}$ be the continuously compounded return from (the end of month) t to (the end of month) $t+1$ on a bill or bond portfolio with maturity τ at t. For bills, τ is a given number of months to maturity, but for bonds τ is the interval of maturities covered by a portfolio. The continuously compounded return on a bill is just the natural log of the ratio of its prices at $t+1$ and t, adjusted to a 30.4 day basis. [See Fama (1984) for details and rationale.] For a bond portfolio, the simple monthly returns on the CRSP file (which properly take account of coupons and accumulated interest) are averaged across bonds to get an equal weighted simple return. The continuously compounded portfolio return is the natural log of one plus the simple return. The premium in the return on a bill or bond portfolio with maturity τ is defined as

$$P\tau_{t+1} \equiv H\tau_{t+1} - H1_{t+1}, \tag{1}$$

where $H1_{t+1}$ is the continuously compounded return or rate of interest (later denoted R_{t+1}) calculated from the price of a bill with one month to maturity at t.

4.1. T^2 tests

Tables 2 and 3 summarize the evidence on average returns for the bills (table 2) and the bond portfolios (table 3).

The T^2 tests on the vectors of average premiums for the overall periods of available data provide no support for the hypothesis that all expected premiums are zero. The sample F statistics for the two T^2 statistics correspond to fractiles of the F distribution almost indistinguishable from 1.0. Likewise, the hypothesis that the structure of expected returns (or premiums) is flat beyond two months gets no support in the bill return data. The T^2 statistic for the vector of average values of the contemporaneous return spreads, $H\tau_{t+1} - H(\tau-1)_{t+1}$, $\tau = 3,\ldots,12$, in table 2 produces an F statistic so large that the computer program, which calculates probability levels to six decimals, rounds the probability level to 1.00.

At this point, however, the bills and the bond portfolios part company. The T^2 test on the vector of average differences between returns on adjacent maturity bond portfolios in table 3 is consistent with the hypothesis that the expected premiums in bond returns do not differ across portfolio maturities. Indeed, this T^2 test produces an F statistic just about at the median of the distribution of the F statistic under the hypothesis that expected premiums do not differ by maturity.

Most of the bond portfolios cover longer maturities than the bills. Thus, the inference from the T^2 tests that there are systematic differences in the expected returns on multi-month bills but not on bond portfolios suggests some systematic behavior of expected returns, or return variability, as a function of maturity. The results for individual maturities in tables 2 and 3 are relevant evidence.

Table 2

Average premiums, t statistics and T^2 tests for bills with up to twelve months to maturity.

Bill	$N = 211$[a] 8/64–12/82	$N = 101$ 8/64–12/72	$N = 110$[a] 1/73–12/82	$N = 56$[a] 1/73–12/77	$N = 54$[a] 1/78–12/82
			Average premiums		
P2	0.032	0.028	0.035	0.016	0.056
P3	0.057	0.045	0.067	0.042	0.094
P4	0.063	0.046	0.078	0.056	0.101
P5	0.074	0.061	0.086	0.065	0.108[b]
P6	0.073	0.066	0.079	0.062	0.097
P7	0.069	0.071	0.067	0.060	0.074
P8	0.088	0.084	0.091	0.083[b]	0.100
P9	0.089[b]	0.086	0.092[b]	0.082	0.102
P10	0.057	0.025	0.086	0.077	0.096
P11	0.064	0.066	0.063	0.066	0.059
P12	0.074	0.103[b]	0.047	0.040	0.054
			t statistics for average premiums		
P2	6.40	6.97	4.03	2.70	3.38
P3	6.40	7.17	4.21	3.42	3.15
P4	4.70	5.18	3.23	2.78	2.26
P5	4.14	5.12	2.64	2.43	1.78
P6	3.34	4.35	2.01	1.92	1.32
P7	2.75	3.68	1.50	1.58	0.90
P8	3.04	3.86	1.76	1.95	1.04
P9	2.59	3.27	1.49	1.69	0.88
P10	1.49	0.83	1.26	1.41	0.75
P11	1.54	2.09	0.83	1.10	0.42
P12	1.61	2.87	0.57	0.59	0.36
			t statistics for average values of $H\tau - H(\tau-1)$		
$H3 - H2$	4.68	4.40	3.33	3.27	2.14
$H4 - H3$	1.05	0.08	1.10	1.58	0.42
$H5 - H4$	1.93	3.01	0.76	0.97	0.35
$H6 - H5$	−0.27	0.74	−0.79	−0.38	−0.69
$H7 - H6$	−0.72	0.81	−1.41	−0.33	−1.40
$H8 - H7$	3.20	2.32	2.40	2.35	1.43
$H9 - H8$	0.10	0.19	0.01	−0.13	0.07
$H10 - H9$	−4.15	−5.66	−0.50	−0.42	−0.33
$H11 - H10$	1.06	5.00	−2.41	−1.20	−2.10
$H12 - H11$	1.28	4.27	−1.41	−1.60	−0.34
			T^2 *tests for average premiums*		
T^2	130.41	152.59	69.33	49.64	56.37
F	11.29	12.48	5.72	3.69	4.18
P-level	1.00	1.00	1.00	0.9994	0.9998
			T^2 *tests for average values of* $H\tau - H(\tau-1)$		
T^2	80.23	95.90	52.11	46.87	35.39
F	6.95	7.85	4.30	3.49	2.61
P-level	1.00	1.00	1.00	0.999	0.990

[a]Twelve-month bills are not available for eight months of the 1973–82 period, and ten- and eleven-month bills are each missing for one month. These months are deleted for all maturities.

[b]Largest average premium. Average premiums are multiplied by 100. Thus, they are percents per month.

Table 3

Average premiums, t statistics and T^2 tests for bond portfolios.

Portfolio number	Maturity range (months)	$N=360$ 1953–82	$N=60$ 1953–57	$N=60$ 1958–62	$N=60$ 1963–67	$N=60$ 1968–72	$N=60$ 1973–77	$N=60$ 1978–82
		Average premiums						
1	$M<6$	0.036	0.017	0.045	0.002[a]	0.040	0.051	0.063[a]
2	$6\leq M<12$	0.042	0.032	0.082	−0.008	0.062	0.049	0.034
3	$12\leq M<18$	0.048[a]	0.039	0.095	−0.013	0.069	0.062[a]	0.034
4	$18\leq M<24$	0.037	0.042	0.099	−0.036	0.072	0.040	0.002
5	$24\leq M<30$	0.026	0.066[a]	0.086	−0.053	0.060	0.050	−0.053
6	$30\leq M<36$	0.034	0.050	0.117	−0.066	0.093[a]	0.053	−0.044
7	$36\leq M<48$	0.012	0.046	0.120[a]	−0.082	0.041	0.028	−0.080
8	$48\leq M<60$	−0.024	0.043	0.059	−0.133	0.030	−0.029	−0.112
9	$60\leq M<120$	−0.012	0.052	0.085	−0.118	0.083	0.017	−0.190
10	$M\cong 240$	−0.128	0.015	0.001	−0.331	−0.032	−0.054	−0.068
		t statistics for average premiums						
1	$M<6$	3.86	1.52	3.42	0.18	2.34	2.62	1.38
2	$6\leq M<12$	1.97	1.73	3.23	−0.50	1.54	1.07	0.32
3	$12\leq M<18$	1.47	1.29	1.94	−0.46	1.09	0.85	0.22
4	$18\leq M<24$	0.88	0.96	1.60	−0.89	0.78	0.45	0.01
5	$24\leq M<30$	0.53	1.19	1.20	−1.10	0.56	0.49	−0.22
6	$30\leq M<36$	0.60	0.78	1.33	−1.25	0.74	0.46	−0.16
7	$36\leq M<48$	0.18	0.65	1.14	−1.20	0.28	0.22	−0.26
8	$48\leq M<60$	−0.32	0.45	0.46	−1.61	0.18	−0.20	−0.32
9	$60\leq M<120$	−0.13	0.43	0.62	−1.11	0.39	0.10	−0.46
10	$M\cong 240$	−0.99	0.08	0.01	−1.84	−0.09	−0.20	−0.66
		t statistics for average differences between adjacent-maturity portfolio returns						
2 − 1		0.42	1.29	2.37	−0.80	0.78	−0.04	−0.43
3 − 2		0.43	0.41	0.43	−0.32	0.26	0.40	−0.00
4 − 3		−0.90	0.16	0.17	−1.46	0.07	−0.91	−0.69
5 − 4		−0.79	0.93	−0.64	−1.15	−0.42	0.33	−0.93
6 − 5		0.78	−0.68	1.19	−0.54	0.75	0.12	0.16
7 − 6		−1.33	−0.20	0.09	−0.60	−1.04	−0.64	−0.61
8 − 7		−1.52	−0.06	−1.58	−1.87	−0.18	−0.94	−0.33
9 − 8		0.34	0.17	0.51	0.33	0.46	0.46	−0.65
10 − 9		−1.72	−0.40	−0.98	−2.20	−0.47	−0.51	−0.72
		T^2 tests for average premiums						
T^2		36.80	8.96	22.45	14.79	13.99	17.50	13.03
F		3.59	0.76	1.90	1.25	1.19	1.48	1.10
P-level		0.9998	0.33	0.94	0.72	0.68	0.83	0.626
		T^2 tests for average differences between adjacent-maturity portfolio returns						
T^2		9.75	4.29	11.93	14.71	3.71	4.88	4.39
F		0.95	0.36	1.01	1.25	0.31	0.41	0.37
P-level		0.51	0.04	0.55	0.72	0.03	0.07	0.05

[a]Largest average premium. Average premiums are multiplied by 100. Thus, they are percents per month.

4.2. The behavior of average returns by maturity

4.2.1. Bills

It is impressive that the average premiums in the returns on two- to twelve-month bills in table 2 are all positive in all periods. Moreover, the maximum average return never occurs at a maturity less than five months.

The t statistics for the average values of $P2 = H2 - H1$ and $H\tau - H(\tau - 1)$, $\tau = 3, \ldots, 12$, allow us to make probability statements about the structure of expected bill returns as a function of maturity. The t statistics for the positive average values of $P2$, $H3 - H2$ and $H8 - H7$ for the overall sample period are well beyond the 0.95 fractile for t for Bonferroni probability statements about 11 means. The t statistic (1.93) for the average value of $H5 - H4$ is less impressive but nevertheless in the upper part of the right tail of the null distribution. On the other hand, the t statistic for the average value of $H10 - H9$ is negative (-4.15) and far into the left tail (below the 0.005 fractile) of the null distribution.[3]

The straightforward inference from the t statistics for the overall period is that there are expected premiums in the returns on multi-month bills, and they increase with maturity out to eight or nine months. The expected premium drops at ten months but shows no reliable movement thereafter.

The subperiod results support the conclusion that expected premiums increase out to eight or nine months, but the conclusion that expected returns on ten-month bills are lower than on nine-month bills becomes more anomalous. In the post-1972 subperiods the average premiums decline monotonically with maturity after nine months, but the average value of $P10$ is never much less than the average value of $P9$. The only period when the ten-month bill has a much lower average return than the nine-month bill is August 1964 to December 1972, but during this period the twelve-month bill provides the highest average return! In other words, most of the reliably negative average value of $H10 - H9$ observed for the overall sample period is due to a subperiod during which average returns show a 'bow' between nine and twelve months.

The behavior of ten-month bill returns during the pre-1973 period is not due to a few extreme monthly returns. A screen of the month-by-month returns indicates that during this period ten-month bills often have lower returns than nine- and twelve-month bills. Later we examine estimates of expected returns extracted from forward rates. These *ex ante* estimates confirm (but likewise do not explain) the bow observed between nine and twelve months in the *ex post* average returns of the pre-1973 period. The *ex ante* estimates of expected

[3] Since we also make probability statements about T^2 statistics for the vector of average premiums and the vector of average differences between adjacent maturity bill returns, strictly speaking we should use Bonferroni fractiles for tests on thirteen rather than eleven t statistics. This refinement has no effect on our inferences. A similar comment is relevant in the analysis of table 3.

returns from forward rates also confirm the downward slope of longer-maturity average bill returns observed during the 1973–82 period.

4.2.2. *The bond portfolios*

The bond portfolios cover longer maturities and a longer time period (1953–82) than the bill returns. The additional results from the bond portfolios in table 3 complement those for bills at the short end of the maturity spectrum, but the bond portfolios provide ambiguous evidence about the behavior of expected returns on longer-maturity instruments.

As noted earlier, the T^2 statistic for the vector of average bond return premiums rejects the hypothesis that expected premiums are all zero. Unlike bills, however, the T^2 test on the vector of average differences or spreads between returns on adjacent maturity bond portfolios is consistent with the hypothesis that expected premiums do not differ by maturity. These results are corroborated by the detailed evidence from the t statistics for the average return spreads. Using the Bonferroni fractiles in table 1 which are relevant when ten means are tested against zero, the average premium for the bond portfolio covering maturities up to six months is reliably greater than zero, but none of the other average differences between returns on adjacent maturity portfolios are reliably different from zero. Indeed, except for the two shortest maturity portfolios, none of the t statistics for average premiums for the 1953–82 period are large.[4]

The bond data are consistent with the hypothesis that the structure of expected premiums in bond portfolio returns is flat. The data are also consistent with a wide range of alternative hypotheses. For example, during the 1953–82 period the 20 year bond portfolio has an average premium of -0.00128 per month or about -1.5 percent per year. In univariate terms, this average premium is less than one standard error from zero. A positive average premium of the same magnitude would likewise be less than one standard error from zero – but it would be almost three times the maximum average premium observed among shorter-maturity portfolios.

Though the evidence lacks statistical precision, it is interesting that there is no five-year subperiod of the 1953–82 sample period during which average bond returns increase systematically with maturity. The shortest-maturity portfolio (< 6 months) produces the largest average return in two of the five-year subperiods. Average returns never peak in maturity intervals beyond four years. At least on an *ex post* basis, the thirty-year period 1953–82 was not propitious for long-term bonds.

[4]Since they are based on different assumptions about the relevant number of tests on means (unspecified for the T^2 statistic but specified under the Bonferroni approach), the T^2 and Bonferroni tests need not lead to the same inferences. This problem does not arise in our data.

4.2.3. The variability of returns by maturity

The obvious source of the imprecision of inferences about expected returns on longer-term bonds – high return variability – is documented in table 4. The standard deviation of the longest-term bond return premiums, 2.47 percent per month, is about fourteen times the standard deviation of the premiums for the shortest-maturity (up to six months) bond portfolio, 0.18 percent per month. Even the bond portfolio covering the eighteen- to twenty-four-month maturity range produces premiums more than four times as variable as the shortest-maturity portfolio.

The standard deviations of the differences or spreads between adjacent-maturity bill returns and adjacent-maturity bond portfolio returns show even more clearly why inferences about expected bill returns are more precise than

Table 4

Standard deviations of premiums and differences between adjacent-maturity bill and bond portfolio returns.[a]

8/64–12/82		1953–82		
Bill (1)	Std. dev. bills (2)	Bond portfolio number (3)	Maturity range (months) (4)	Std. dev. bonds (5)
		Premiums		
P2	0.07	1	$M < 6$	0.18
P3	0.13	2	$6 \le M < 12$	0.40
P4	0.19	3	$12 \le M < 18$	0.62
P5	0.26	4	$18 \le M < 24$	0.79
P6	0.32	5	$24 \le M < 30$	0.93
P7	0.36	6	$30 \le M < 36$	1.08
P8	0.42	7	$36 \le M < 48$	1.23
P9	0.50	8	$48 \le M < 60$	1.42
P10	0.56	9	$60 \le M < 120$	1.70
P11	0.61	10	$M \cong 240$	2.47
P12	0.67			
		Differences between adjacent-maturity returns		
$H3 - H2$	0.08	2 – 1		0.26
$H4 - H3$	0.08	3 – 2		0.26
$H5 - H4$	0.09	4 – 3		0.24
$H6 - H5$	0.08	5 – 4		0.25
$H7 - H6$	0.08	6 – 5		0.27
$H8 - H7$	0.09	7 – 6		0.31
$H9 - H8$	0.10	8 – 7		0.44
$H10 - H9$	0.11	9 – 8		0.67
$H11 - H10$	0.10	10 – 9		1.29
$H12 - H11$	0.11			

[a] The standard deviations for bills in column (2) and those for the bond portfolios in column (5) should be read as percents per month.

inferences about expected bond returns. Bill return premiums have low absolute variability, but the premium on a twelve-month bill is nevertheless about nine times more variable than the premium on a two-month bill. Because of the high correlation of adjacent maturity bill returns, however, $H12 - H11$ is only about forty percent more variable than $P2 = H2 - H1$. Positive correlation between adjacent maturity portfolio returns also causes the standard deviations of bond portfolio return spreads to increase less rapidly with maturity than the standard deviations of premiums, but the standard deviations of bond portfolio return spreads are nevertheless large relative to those for adjacent-maturity bills. For example, the standard deviation of the difference between the return on the longest-term bonds and the bond portfolio that includes five- to ten-year maturities is 1.29 percent per month, whereas the standard deviation of the difference between the returns on twelve- and eleven-month bills is only 0.11 percent per month.

4.3. *Simple versus continuously compounded returns*

The continuously compounded monthly returns used in the preceding tests are always less than simple returns. Since the variances of returns increase with maturity, we can predict that the differences between average simple and continuously compounded returns increase with maturity. Thus, the tendency for average premiums in bill returns to peak at eight or nine months, and the low average returns on longer-term bonds, may be due to the use of continuously compounded returns. Finally, one can interpret the arguments in Cox, Ingersoll and Ross (1981) as calling for simple returns in tests for the existence of premiums.

Table 5 shows average premiums calculated from simple returns for the overall sample periods covered by the bills and by the bond portfolios. Average premiums calculated from continuously compounded returns are also shown. The simple average premiums are always larger than the continuously compounded average premiums, and the differences indeed increase with maturity. However, for bills the differences are trivial even at the longest maturities (the largest is 0.3 basis points per month), and they are small even for the longest-maturity bond portfolios. The use of simple returns does not change the maturities that produce the largest average returns (9 months for bills and 12–18 months for the bond portfolios). The use of simple returns also has no effect of consequence on the t statistics for average premiums and average spreads between adjacent maturity returns.

5. Estimates of expected bill premiums from forward rates

Prices of longer-maturity bills and bonds move opposite to interest rates, and changes in interest rates are on average positive during the sample period.

Table 5

Comparisons of average premiums, t statistics and T^2 tests for continuously compounded and simple returns.

Bill	$N = 211$[a] 8/64–12/82 Continuous	$N = 211$[a] 8/64–12/82 Simple	Portfolio number	Maturity range (months)	$N = 360$ 1953–82 Continuous	$N = 360$ 1953–82 Simple
			Average premiums			
P2	0.032	0.032	1	$M < 6$	0.036	0.036
P3	0.057	0.058	2	$6 \leq M < 12$	0.042	0.043
P4	0.063	0.063	3	$12 \leq M < 18$	0.048[b]	0.050[b]
P5	0.074	0.075	4	$18 \leq M < 24$	0.037	0.040
P6	0.073	0.074	5	$24 \leq M < 30$	0.026	0.031
P7	0.069	0.070	6	$30 \leq M < 36$	0.034	0.040
P8	0.088	0.090	7	$36 \leq M < 48$	0.012	0.020
P9	0.089[b]	0.091[b]	8	$48 \leq M < 60$	−0.024	−0.014
P10	0.057	0.059	9	$60 \leq M < 120$	−0.012	0.003
P11	0.064	0.067	10	$M \cong 240$	−0.128	−0.098
P12	0.074	0.077				
			t statistics for average premiums			
P2	6.40	6.38	1	$M < 6$	3.86	3.87
P3	6.40	6.38	2	$6 \leq M < 12$	1.97	2.00
P4	4.70	4.69	3	$12 \leq M < 18$	1.47	1.52
P5	4.14	4.14	4	$18 \leq M < 24$	0.88	0.95
P6	3.34	3.34	5	$24 \leq M < 30$	0.53	0.61
P7	2.75	2.77	6	$30 \leq M < 36$	0.60	0.69
P8	3.04	3.05	7	$36 \leq M < 48$	0.18	0.30
P9	2.59	2.61	8	$48 \leq M < 60$	−0.32	−0.18
P10	1.49	1.52	9	$60 \leq M < 120$	−0.13	0.03
P11	1.54	1.57	10	$M \cong 240$	−0.99	−0.75
P12	1.61	1.65				
			t statistics for average differences between adjacent-maturity returns			
H3 − H2	4.68	4.68	2 − 1		0.42	0.47
H4 − H3	1.05	1.07	3 − 2		0.43	0.51
H5 − H4	1.93	1.95	4 − 3		−0.90	−0.80
H6 − H5	−0.27	−0.24	5 − 4		−0.79	−0.70
H7 − H6	−0.72	−0.69	6 − 5		0.78	0.65
H8 − H7	3.20	3.23	7 − 6		−1.33	−1.23
H9 − H8	0.10	0.15	8 − 7		−1.52	−1.39
H10 − H9	−4.15	−4.10	9 − 8		0.34	0.46
H11 − H10	1.06	1.09	10 − 9		−1.72	−1.48
H12 − H11	1.28	1.33				
			T^2 test for average premiums			
T^2	130.41	129.48			36.80	35.24
F	11.29	11.21			3.59	3.44
P-level	1.00	1.00			0.9998	0.9997
			T^2 tests for average differences between adjacent-maturity returns			
T^2	80.23	79.45			9.75	8.53
F	6.95	6.88			0.95	0.83
P-level	1.00	1.00			0.51	0.40

[a] Twelve-month bills are not available for eight months of the 1973–82 period, and ten- and eleven-month bills are each missing for one month. These months are deleted for all maturities.

[b] Largest average premium. Average premiums should be read as percents per month.

A common claim of readers faced with the results above is that the higher sensitivity of returns on longer-term instruments to unexpected upward shifts in the level of interest rates explains their lower average realized returns.

One way to purge the effects of unexpected shifts in the term structure from estimates of expected premiums is to estimate expected premiums from forward rates. This approach can only be applied to bills since good estimates of forward rates are not possible for our (mixed maturity) bond portfolios.

5.1. Spot and forward interest rates

Define $V\tau_t$ as the price at time t (the end of month t) of a bill that has τ months to maturity at t and pays \$1 for certain at the end of month $t+\tau$. Define R_{t+1}, the one-month spot rate of interest from t to $t+1$, observed in the market at t, as

$$V1_t = \exp(-R_{t+1}). \tag{2}$$

The price $V\tau_t$ can then be expressed as

$$V\tau_t = \exp(-R_{t+1} - F2_t - \cdots - F\tau_t), \tag{3}$$

where $F\tau_t$, the forward rate for month $t+\tau$ observed at t, is

$$F\tau_t \equiv \ln(V(\tau-1)_t / V\tau_t). \tag{4}$$

Note that the spot rate R_{t+1} and the forward rates $F2_t, \ldots, F\tau_t$ can be calculated from bill prices observed in the market at time t.

Fama (1976b) shows that forward rates can be expressed as

$$\begin{aligned} F\tau_t = {} & \mathrm{E}_t(P\tau_{t+1}) + [\mathrm{E}_t(P(\tau-1)_{t+2}) - \mathrm{E}_t(P(\tau-1)_{t+1})] + \cdots \\ & + [\mathrm{E}_t(P2_{t+\tau-1}) - \mathrm{E}_t(P2_{t+\tau-2})] + \mathrm{E}_t(R_{t+\tau}), \end{aligned} \tag{5}$$

where E_t indicates an expected value at time t. Thus, the forward rate for month $t+\tau$, observed at t, contains $\mathrm{E}_t(R_{t+\tau})$, the expected value of the future spot rate for month $t+\tau$. The forward rate also contains $\mathrm{E}_t(P\tau_{t+1})$, the expected premium in the return on a τ-month bill from t to $t+1$, and current expected changes in future premiums.

Table 6 shows average values of $F\tau_t - R_{t+1}$, $\tau = 2, \ldots, 12$, for various periods of the twelve-month bill file. Forward rates are calculated from (4) and are adjusted to a 30.4 day monthly basis, as described in Fama (1984). The time t spot rate is subtracted from the time t forward rate to focus better on the expected premium component of $F\tau_t$. That is, although the expected

Table 6

Comparisons of average premiums (P_T) and average differences between contemporaneous forward and spot rates ($F_T - R$).

Maturity	8/64–12/82		8/64–12/72		1/73–12/82		1/73–12/77		1/78–12/82	
τ	P_T	$F_T - R$	P_T	$F_T - R$	P_T	$F_T - R$	P_T	$F_T - R$	P_T	$F_T - R$
Average values										
2	0.032	0.033	0.028	0.029	0.035	0.036	0.016	0.019	0.056	0.054
3	0.057	0.058	0.045	0.048	0.067	0.068	0.042	0.050	0.094	0.087
4	0.063	0.066	0.046	0.050	0.078	0.081	0.056	0.067	0.101	0.096
5	0.074	0.081	0.061	0.067	0.086	0.094	0.065	0.076	0.108[a]	0.112[a]
6	0.073	0.082	0.066	0.073	0.079	0.090	0.062	0.074	0.097	0.106
7	0.069	0.081	0.071	0.079	0.067	0.082	0.060	0.074	0.074	0.090
8	0.088	0.100	0.084	0.095	0.091	0.104	0.083[a]	0.098	0.100	0.110
9	0.089[a]	0.101[a]	0.086	0.097	0.092[a]	0.105[a]	0.082	0.099[a]	0.102	0.110
10	0.057	0.073	0.025	0.038	0.086	0.105[a]	0.077	0.099[a]	0.096	0.112[a]
11	0.064	0.082	0.066	0.080	0.063	0.084	0.066	0.091	0.059	0.077
12	0.074	0.097	0.103[a]	0.118[a]	0.047	0.078	0.040	0.085	0.054	0.070
t statistics for average values of P_T and $F_T - R$										
2	6.40	8.60	6.97	6.66	4.03	5.93	2.70	5.66	3.38	4.69
3	6.40	13.77	7.17	12.34	4.21	9.46	3.42	9.45	3.15	6.59
4	4.70	14.50	5.18	12.23	3.23	10.63	2.78	15.00	2.26	6.56
5	4.14	14.40	5.12	10.72	2.64	10.44	2.43	9.14	1.78	7.06
6	3.34	14.29	4.35	11.42	2.01	9.72	1.92	8.81	1.32	6.44
7	2.75	12.44	3.68	12.07	1.50	7.51	1.58	9.48	0.90	4.33
8	3.04	17.92	3.86	15.22	1.76	11.55	1.95	10.43	1.04	7.06
9	2.59	16.41	3.27	14.58	1.49	10.35	1.69	11.27	0.88	5.95
10	1.49	10.07	0.83	4.73	1.26	9.59	1.41	7.99	0.75	6.08
11	1.54	12.92	2.09	10.88	0.83	8.26	1.10	7.95	0.42	4.52
12	1.61	12.63	2.87	12.86	0.57	6.57	0.59	5.49	0.36	3.88
t statistics for average values of $H_T - H(\tau - 1)$ and $F_T - F(\tau - 1)$										
3 − 2	4.68	7.12	4.40	3.77	3.33	6.23	3.27	5.32	2.14	3.85
4 − 3	1.05	2.05	0.08	0.39	1.10	2.28	1.58	3.13	0.42	0.89
5 − 4	1.93	3.18	3.01	2.71	0.76	1.85	0.97	1.26	0.35	1.36
6 − 5	−0.27	0.19	0.74	0.78	−0.79	−0.56	−0.38	−0.24	−0.69	−0.51
7 − 6	−0.72	−0.17	0.81	0.78	−1.41	−0.85	−0.33	0.01	−1.40	−0.94
8 − 7	3.20	3.55	2.32	2.36	2.40	2.67	2.35	2.92	1.43	1.38
9 − 8	0.10	0.34	0.19	0.39	0.01	0.12	−0.13	0.11	0.07	0.06
10 − 9	−4.15	−3.96	−5.66	−5.41	−0.50	0.05	−0.42	−0.01	−0.33	0.07
11 − 10	1.06	1.20	5.00	3.83	−2.41	−2.57	−1.20	−0.74	−2.10	−2.98
12 − 11	1.28	2.08	4.27	3.69	−1.41	−0.68	−1.60	−0.41	−0.34	−0.57

[a]Largest average value. Means should be read as percents per month.

changes in future premiums and the expected change in the spot rate $\mathrm{E}(R_{t+\tau}) - R_{t+1}$ in $F\tau_t - R_{t+1}$ can vary from month to month, over long sample periods, the average values of these expected changes should be close to zero. Thus, the average value of $F\tau_t - R_{t+1}$ should be close to the average value of the expected premium $\mathrm{E}_t(P\tau_{t+1})$ in $F\tau_t$.

In setting up the comparisons in table 6, my strong prior was that the averages of $F\tau_t - R_{t+1}$ would increase monotonically with τ in all periods and allow us to infer the effects of unexpected shifts in the term structure in explaining the typically non-monotonic behavior of the average values of the premiums. Table 6 indicates, however, that the market's expectations are realized while mine are not. The average values of $F\tau_t - R_{t+1}$ and $P\tau_{t+1}$ peak at the same or adjacent maturities. This is true for the overall sample period and, perhaps more impressive, it is also true for every subperiod. The average values of $F\tau_t - R_{t+1}$ also replicate most of the details of the behavior of the average values of $P\tau_{t+1}$, for example, the bow in the average values of $P\tau_{t+1}$ observed for the nine- to twelve-month maturities during the August 1964 to December 1972 period.

The estimates of expected premiums from forward rates do not alter the general view obtained from average premiums, but the picture from the forward rates is more precise. Because the forward rate spreads, $F\tau_t - R_{t+1}$ and $F\tau_t - F(\tau-1)_t$, are less variable than the realized premiums, $P\tau_{t+1}$, and return spreads, $H\tau_{t+1} - H(\tau-1)_{t+1}$, the t statistics for the average values of $F\tau_t - R_{t+1}$ and $F\tau_t - F(\tau-1)_t$ are generally larger in absolute value than those for the means of $P\tau_{t+1}$ and $H\tau_{t+1} - H(\tau-1)_{t+1}$. As a consequence, the t statistics from the forward rate estimates for the overall sample period in table 6 provide stronger indications that expected returns on bills increase with maturity up to eight or nine months. However, the t statistic for the average value of $F10 - F9$ also reinforces the conclusion that during the overall sample period, the expected return on a ten-month bill is reliably less than that for a nine-month bill.

The straightforward inference from table 6 is that expected returns on bills do not increase monotonically with maturity. The one period (August 1964 to December 1972) when twelve-month bills produce the largest average values of $P\tau_{t+1}$ and $F\tau_t - R_{t+1}$ nevertheless shows a curious dip in the two averages for the ten- and eleven-month maturities. For all other subperiods and for the overall sample period, the largest average values of $P\tau_{t+1}$ and $F\tau_t - R_{t+1}$ occur at ten months or less.

6. Conclusions

This paper examines expected returns on U.S. Treasury bills and U.S. Government bond portfolios that cover the maturity spectrum.

Expected bill returns are estimated from forward rates and from sample average returns. Both estimation methods indicate that expected returns on bills tend to peak at eight or nine months and never increase monotonically out to twelve months. Since the estimates of expected premiums from *ex ante* forward rates peak at the same maturities as the *ex post* average premiums, the non-monotonicity of the average premiums cannot be attributed to unexpected shifts in the term structure.

During the five-year subperiods of 1953–82 covered by our bond portfolios, bonds with maturities greater than four years never have the highest average returns. During periods where the bond file overlaps with the twelve-month bill file, the highest average return on a bond portfolio never exceeds the highest average return on a bill. We cannot conclude, however, that longer-term bonds have lower expected returns than short-term instruments. The high variability of longer-term bond returns preempts precise conclusions about their expected returns. The bond data are consistent with maturity structures of expected returns that are flat, upward sloping or downward sloping beyond a year. Thus, longer-maturity bond portfolios do not provide much evidence on the behavior of expected returns.

The non-monotonicity of expected bill returns documented here is inconsistent with either the pure expectations hypothesis or the liquidity preference hypothesis of the classical term structure literature. However, it is not necessarily inconsistent with the more sophisticated models of capital market equilibrium of Sharpe (1964), Lintner (1965), Merton (1973) or Breeden (1979). I offer no direct tests. Rather, the descriptive evidence presented here, and the evidence or rich patterns of variation through time in expected bill returns presented by Fama (1984), Startz (1982) and others, stand as challenges or 'stylized facts' to be explained by candidate models.

References

Breeden, Douglas T., 1979, An intertemporal asset pricing model with stochastic consumption and investment opportunities, Journal of Financial Economics 7, 265–296.

Cox, John C., Jonathan E. Ingersoll, Jr. and Stephen A. Ross, 1981, A reexamination of traditional hypotheses about the term structure of interest rates, Journal of Finance 36, 769–799.

Fama, Eugene F., 1976a, Forward rates as predictors of future spot rates, Journal of Financial Economics 3, 361–377.

Fama, Eugene F., 1976b, Inflation uncertainty and expected returns on treasury bills, Journal of Political Economy 84, 427–448.

Fama, Eugene F., 1984, The information in the term structure. Journal of Financial Economics, this issue.

Ibbotson, Roger G. and Rex A. Sinquefield, 1983, Stocks, bonds, bills and inflation: The past and the future (Financial Analysts Research Foundation, Charlottesville, VA).

Kessel, Reuben A., 1965, The cylical behavior of the term structure of interest rates, National Bureau of Economic Research occasional paper no. 91.

Lintner, John, 1965, The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets, Review of Economics and Statistics 47, 13–27.

McCulloch, J. Huston, 1975, An estimate of the liquidity premium, Journal of Political Economy 83, 95–119.
Meiselman, David, 1962, The term structure of interest rates (Prentice-Hall, Englewood Cliffs, NJ).
Merton, Robert C., Jr., 1973, An intertemporal capital asset pricing model, Econometrica 41, 867–887.
Miller, Rupert G., Jr. 1981, Simultaneous statistical inference (Springer-Verlag, New York).
Morrison, Donald F., 1976, Multivariate statistical methods (McGraw-Hill, New York).
Roll, Richard, 1970, The behavior of interest rates (Basic Books, New York).
Roy, S.N. and R.C. Bose, 1953, Simultaneous confidence interval estimation, Annals of Mathematical Statistics 24, 513–536.
Sharpe, William F., 1964, Capital asset prices: A theory of market equilibrium under conditions of risk, Journal of Finance 19, 425–442.
Startz, Richard, 1982, Do forecast errors or term premia really make the difference between long and short rates?, Journal of Financial Economics 10, 323–329.