N. Bulent Gultekin
University of Pennsylvania

Richard J. Rogalski
Dartmouth College

Alternative Duration Specifications and the Measurement of Basis Risk: Empirical Tests*

This paper empirically investigates seven duration measures and their role in explaining price volatility caused by interest rate movements, that is, basis risk. The data analysis does not support the important testable implications of the various durations as measures of basis risk. All seven durations fail linearity and single risk factor tests from which their advantages are supposed to derive. Fixed-income performance comparisons and immunization strategies based on these durations have not been worthwhile.

I. Introduction

Dramatic increases in interest rate levels and volatility since the early 1970s have renewed interest in fixed-income securities and have motivated much study of the appropriate measure of risk for bonds. Many authors have expanded on Macaulay's (1938) concept of duration as a surrogate for risk measurement and have developed a variety of new measures of duration. Numerous commercial duration-based immunization programs are now available purporting to explain returns, to provide good measures of risk, and to protect fixed-income portfolios from volatile interest rates (e.g., Leibowitz 1979; Leibowitz and Weinberger 1981).

This paper tests the explanatory power of a number of recently proposed duration measures. We show that despite the flood of articles and commercial programs claiming superiority for particular measures of duration, the measures studied were virtually indistinguishable empirically. In fact, none of them did much better than

* Helpful comments and suggestions were received from Eugene Fama, the referees of this journal, and especially Merton Miller. Dorothy Bower and Richard P. McNeil provided valuable assistance and computer programming skills. This study was initiated while we worked together at the Amos Tuck School, with continued funding from the Tuck Associates Research Program.

(*Journal of Business,* 1984, vol. 57, no. 2)

0021-9398/84/5702-0004$01.50

simple maturity in explaining bond returns, and all duration measures are inferior to simple factor models.

These results are in sharp contrast to the claims made by authors about various duration measures, and they indicate that duration is not an appropriate index to hold constant in evaluating fixed-income portfolio strategies. Interest rate movements have been such that immunization strategies based on duration would not have worked. Duration-based immunization programs do not appear to warrant the resources spent on them.

In the next section we define the various duration measures to be examined. Section III outlines testable implications of duration measures and contains details of the return and duration calculations. Section IV tests hypotheses about the ability of the various duration measures to explain Treasury security price volatility during the sample period. A summary with conclusions completes the paper.

II. Different Duration Measures of Price Volatility

Macaulay (1938) defines the simple duration D of a bond to be

$$D = \frac{1}{P(m)} \sum_{t=1}^{m} C(t)t(1 + y)^{-t}, \tag{1}$$

where $P(m)$ is the current price of a bond with maturity m, $C(t)$ is the income stream at time t (coupon payments and principal), and y is the yield to maturity of the bond. Macaulay's intention was to devise an adequate measure of "longness" in order to compare loans with different payment schedules on the basis of the weighted average of the present value of their income streams.[1]

Hicks (1939) proposed, as Hopewell and Kaufman (1973) did more recently, to use duration as a proxy for price volatility, or basis risk, as Cox, Ingersoll, and Ross (1979) call it.[2] The first derivative of the bond price in equation (1) with respect to an interest rate change gives

$$\frac{dP(m)}{P(m)} = -D \frac{dy}{(1 + y)}. \tag{2}$$

1. Macaulay (1938) actually defined duration as $[1/P(m)] \Sigma_{t=1}^{m} C(t)t[1 + R(t)]^{-t}$ where $R(t)$ is the spot rate for cash flow $C(t)$. Most applications of duration, however, use the definition in eq. (1), where duration is simply computed by using yield as a proxy for spot rates. We make this distinction explicit by referring to the definition in eq. (1) as simple duration.

2. See Hicks (1939) and Ingersoll, Skelton, and Weil (1978) for an extensive review of the duration literature.

Equation (2) has been the justification for using duration as a volatility measure because it implies that a bond's price change is linearly related to its duration for small changes in yields. The simple linear relation of equation (2) suggested to others that duration could be used for cross-sectional comparisons of the riskiness of bonds. As Cooper (1977) and Ingersoll, Skelton, and Weil (1978) point out, however, this comparative static property of Macaulay's duration is useful only to the extent that yield-curve shifts are parallel over time.

The current literature suggests two possible ways to overcome this limitation in Macaulay's duration measure. One way is to extend the comparative static analysis to allow for other and presumably less restrictive kinds of shifts of the yield curve,[3] in other words, to custom-make duration measures for specific shocks to the yield curve. The second way is to develop a theoretically plausible characterization of the term structure and use it to derive a natural duration measure.

A. *Alternative Duration Measures*

The first approach is taken by Cooper (1977), Bierwag (1977), Bierwag and Kaufman 1978), and Khang (1979).[4] The six duration measures that emerge from their work are as follows:

3. Most attempts to devise new duration measures have come out of research dealing with immunization of fixed-income portfolios for different yield-curve movements. Of course, if duration is not a reliable measure of basis risk, its usefulness for immunization purposes is questionable.

4. To see this, note that price changes caused by a change in the term structure at an instant in time t can be expressed as

$$\frac{dP(m)}{P(m)} = \frac{1}{P(m)} \int_0^m C(t) \exp[-R(t)t]\, dR(t)\, dt, \tag{a}$$

where $R(t)$ is the spot rate associated with each cash flow $C(t)$. Suppose a polynomial of the following form adequately describes the term structure

$$R(t) = \alpha + \exp[(\gamma/t) \ln(\beta t + 1) + \phi t + \psi], \tag{b}$$

where α, γ, β, ϕ, and ψ are some finite number of parameters defining the level and shape of the term structure at time t. The motivation for such a polynomial representation of the term structure is based on empirically observed yield curves. However, theoretical representations of the term structure derived by Vasicek (1977) and by Cox, Ingersoll, and Ross (1978) are also families of polynomials if the parameters are assumed to be state variables. It is easily seen that the term structure changes with variations in α, γ, ϕ, and ψ:

$$dR(t) = \frac{\partial R(t)}{\partial \alpha}\, d\alpha + \frac{\partial R(t)}{\partial \gamma}\, d\gamma + \frac{\partial R(t)}{\partial \phi}\, d\phi + \frac{\partial R(t)}{\partial \psi}\, d\psi. \tag{c}$$

Substituting (c) into (a) gives the price change due to a change in the term structure. Different duration measures can be obtained by using a comparative static analysis after this substitution.

$$D1 = \left[\frac{1}{P(m)}\right]\int_0^m C(t)t \exp[-R(t)t]\, dt$$

$$D2 = \left[\frac{1}{P(m)}\right]\int_0^m C(t)tR(t) \exp[-R(t)t]\, dt$$

$$D3 = \left[\frac{1}{P(m)}\right]\int_0^m C(t)t^2R(t) \exp[-R(t)t]\, dt \tag{3}$$

$$D4 = \left[\frac{1}{P(m)}\right]\int_0^m C(t)t \ln(t)R(t) \exp[-R(t)t]\, dt$$

$$D5 = \left[\frac{1}{P(m)}\right]\int_0^m C(t)t^2 \exp[-R(t)t]\, dt$$

$$D6 = \left(\exp\left[\frac{1}{P(m)}\right]\int_0^m C(t) \ln(1 + \alpha t) \exp[-R(t)t]\, dt\right) - \frac{1}{\alpha}$$

where $P(m)$ is the market value of a bond with maturity m at some instant of time, $C(t)$ is the cash flow received at t, ln is the natural logarithm, $R(t)$ is the spot rate associated with each cash flow, and α in $D6$ is a measure of the variability of long-term yields relative to short-term yields. The greater α, the greater are changes in short-term rates relative to long-term rates.

$D1$ is Macaulay's (1938) duration. $D2$ is suggested by Bierwag (1977), $D3$–$D5$ are proposed by Cooper (1977), and $D6$ is Khang's (1979) duration. Each duration presumes a specific characteristic movement of the term structure. $D1$ permits only additive yield-curve movements; thus, $D1$ is useful only if the level of yields changes while the slope of the yield curve does not. $D2$–$D6$ allow additive and multiplicative yield-curve movements, that is, changes in both the level and the slope of yield curves. $D2$–$D6$ differ only in the implicit degree of slope and curvature of yield curves; that is, the duration measures $D2$–$D6$ differ in the assumed degree of variability of long-term yields relative to short-term yields. $D2$ assumes the steepest yield curves, while $D5$ is the flattest. $D3$ curves are between $D2$ and $D5$, while $D4$ curves are between $D2$ and $D3$.

B. *Stochastic Duration*

Cox et al. (1979) take the second approach and propose a duration measure they call stochastic duration. They take as their starting point the proposition that the term structure cannot be expected to change according to any known and fixed pattern. If it did, arbitrage opportunities would arise. Cox et al. develop instead a general equilibrium model of the term structure and then define a measure of basis risk that is relevant to their resultant bond pricing equation. Like the other duration measures, this measure accounts for the timing and size of coupons, but it also includes all information on the equilibrium term structure.

Cox et al. develop their stochastic duration measure by assuming that the instantaneous nominal spot rate *dr* follows a first-order autoregressive process

$$dr = \beta(\mu - r)dt + \sigma\sqrt{r}\,dt, \tag{4}$$

where the instantaneous drift term $\beta(\mu - r)$ allows interest rates to move back to the long-run (or steady-state) mean μ. The instantaneous variance $\sigma^2 r$ in the second term causes the process to vary around the mean level μ. It varies most at times of high interest rates and least when interest rates are lower. The stochastic interest rate process in equation (4) permits long rates to shift as much as the short rates. Stochastic duration based on the spot rate process in equation (4) can be expressed as

$$D7 = G^{-1}\left[\frac{\Sigma\, C(t)P(t)G(t)}{C(t)P(t)}\right] \tag{5}$$

where

$$\begin{aligned}
G^{-1}[x] &= \frac{2}{\gamma}\coth^{-1}\left(\frac{2}{\gamma x} + \frac{\pi - \beta}{\gamma}\right)\\
P(t) &= F(t)\exp[-rG(t)]\\
F(t) &= \left\{\frac{2\gamma\exp[\gamma + \beta - \pi)t/2]}{(\gamma + \beta - \pi)[\exp(\gamma t) - 1] + 2\gamma}\right\}^{2\beta\mu/\sigma^2}\\
G(t) &= \frac{2}{[\beta - \pi + \gamma\coth(\gamma t/2)]}\\
\gamma &= [(\beta - \pi)^2 + 2\sigma^2]^{1/2}\\
\pi &= \text{liquidity premium.}
\end{aligned}$$

Stochastic duration for pure discount bonds is $G(t)$, or simply maturity.

Even a casual examination of equation (5) reveals that stochastic duration involves a considerable amount of computation. Statistical procedures are also required to estimate the parameters μ, σ^2, and β for the assumed spot process. Liquidity premiums, if they exist, must be estimated. More complex instantaneous spot rate processes would result in even more complicated duration measures.

III. Testable Implications

A. A Stochastic Model

A careful reading of the duration literature yields explicit hypotheses about duration's properties as a proxy for price volatility, that is, as an index for cross-sectional comparisons of bond risk. Most of the implications of duration can be seen from inspection after generalizing equation (2) as

$$r(m) = -(\Delta y)Dk(m). \tag{6}$$

In other words, for a given change in the assumed term structure, $-[\Delta y]$, the price change in a security with maturity m, $r(m)$, is directly related to its duration $Dk(m)$, with $k = 1, 2, \ldots, 7$ denoting the duration measure under consideration. For example, if $y = [dy/(1 + y)]$, equations (6) and (2) are identical for duration measure $D1(m)$. The term becomes more complicated for the other duration measures.

Equation (6) provides three testable hypotheses for each of the duration measures $D1$–$D7$. First, the relation between security price changes and duration is linear.[5] Second, duration is a complete measure of risk; that is, duration incorporates the effect of maturity and coupon differences on price volatility. Implicit in this condition is that the yield curve on average demonstrates the functional form assumed by the duration measure.[6] Recall that $D1$ assumes that yield curves on average experience parallel shifts, whereas $D2$–$D6$ assume that the short maturity end of the term structure on average moves more than the longer end. The last hypothesis is that capital markets for bonds are efficient.

The linearity, completeness, and efficiency hypotheses can be tested with actual market data for many time periods with the use of securities and portfolios of securities. A model of period-by-period changes that allows us to use observed average price changes to test the three hy-

5. Durations $D1$–$D7$ are derived from different assumptions about the behavior of yields (see n. 6). Given an assumption about yield-curve shifts, the derived duration measure is linearly related to security prices changes $r(m)$. In addition, a linear relation between a duration measure and security price changes means immunization of bond portfolios can be achieved using that duration measure.

6. Lanstein and Sharpe (1978) observed that there are strong suggestions in the duration literature that measures of duration calculated ex ante predict the expected reactions of security prices to unexpected changes in interest rates. This essentially implies that duration is a linear risk measure similar to beta for common stocks. Lanstein and Sharpe also derive conditions for which duration is a risk measure similar to beta. Such interpretations could be tested by rewriting eq. (6) with expectations as

$$E[r(m)] = E[-\Delta y]\, Dk(m). \tag{6a}$$

The linearity hypothesis is that the relation between *expected* security price changes and duration is linear.

potheses is

$$\tilde{r}_\tau(m) = \tilde{\gamma}1_\tau + \tilde{\gamma}2_\tau\, Dk_{\tau-1}(m) + \tilde{\gamma}3_\tau\, Dk^2_{\tau-1}(m) + \tilde{\gamma}4_\tau\, C_\tau(m) + \tilde{e}_\tau(m). \tag{7}$$

The subscript τ refers to the period τ, so that $\tilde{r}_\tau(m)$ is the one-period continuously compounded price change on security m. $C_\tau(m)$ is the coupon on security m, and $Dk_{\tau-1}(m)$ denotes the duration measure under consideration with $k = 1, 2, \ldots, 7$. Note that $\tilde{\gamma}1_\tau$, $\tilde{\gamma}2_\tau$, $\tilde{\gamma}3_\tau$, and $\tilde{\gamma}4_\tau$ vary stochastically from period to period.

The completeness hypothesis is that the expected value of the interest rate change $\tilde{\gamma}2_\tau$ is statistically significant. The variable $Dk^2_{\tau-1}(m)$ is included to test linearity. Linearity presumes that $E(\tilde{\gamma}3_\tau) = 0$. The term involving $C_\tau(m)$ in equation (7) is meant to measure whether duration normalizes coupon differences. Completeness assumes that $E(\tilde{\gamma}4_\tau) = 0$. The intercept term $\tilde{\gamma}_\tau$ is included to measure the level of interest rates. Capital market efficiency implies that $\tilde{\gamma}1_\tau$, $\tilde{\gamma}2_\tau$, $\tilde{\gamma}3_\tau$, $\tilde{\gamma}4_\tau$, and $e_\tau(m)$ should be uncorrelated through time.

The disturbances are assumed to have zero mean and to be independent of all other variables in equation (7). The variables $\tilde{e}_\tau(m)$, $\tilde{\gamma}1_\tau$, $\tilde{\gamma}2_\tau$, $\tilde{\gamma}3_\tau$, and $\tilde{\gamma}4_\tau$ are assumed to follow approximately a multivariate normal distribution.

B. *Return Computations*

For our tests we used data from the Government Bond File of the Center for Research in Security Prices (CRSP), which consist of price, coupon, and maturity information on the last trading day of each month for all U.S. Treasury bills, notes, and bonds during the sample period January 1947–December 1976. We excluded callable and deep discount securities.

Realized holding period returns for all securities during period τ are calculated as

$$r_\tau(m) = \ln\left\{\left[P_\tau(m-1) + \left(\frac{n}{m}\right)C + C\right] \Big/ \left[P_{\tau-1}(m) + \left(\frac{n^*}{m}\right)C\right]\right\}, \tag{8}$$

where $r_\tau(m)$ is the total return on a security with maturity m during holding period τ; P_τ and $P_{\tau-1}$ are averages of bid-ask prices at the end of periods τ and $\tau - 1$; ln is the natural logarithm; C is the semiannual coupon; n is the number of days accrued toward the next coupon payment as of the end of period τ; and n is the number of days accrued toward the next coupon payment at the end of period $\tau - 1$. This is the measure of return included in the CRSP Government Bond File.

Table 1 presents summary statistics for returns over the period 1947–76. Returns are given for nonoverlapping 1-month and 3-month holding periods in panels A and B, respectively. The stub column lists the maturity categories, while column 1 indicates the number of price

TABLE 1 **Summary Statistics for Treasury Security Returns**

Days to Maturity	Number of Securities (1)	Mean (2)	Variance (3)	Skewness (4)	Kurtosis (5)	Studentized Range (6)
A. One-Month Holding Period						
30–59	1,847	.2749	.0275	.6746	2.93	5.11
60–89	1,794	.2947	.0330	.7617	3.56	6.42
90–119	1,428	.3161	.0452	.5007	3.91	6.81
120–149	1,261	.4027	.1287	−.5599	1.60	5.19
150–179	1,228	.4217	.1424	−.5376	1.49	4.10
180–269	1,191	.3932	.1286	−.1657	2.19	6.20
270–359	974	.4087	.1919	.0557	2.23	5.56
360–719	1,641	.3590	.2598	−.0331	6.17	8.84
720–1,079	1,004	.3615	.5213	.3433	5.09	7.77
1,080–1,439	950	.3661	.7856	.3973	5.49	7.23
1,440–1,799	746	.3518	1.1122	.1641	5.34	7.46
1,800–2,159	530	.3648	1.5437	.1119	5.19	6.78
2,160–	943	.3349	1.5346	.2485	5.11	6.93
B. Three-Month Holding Period						
90–119	416	.9387	.3849	−.3430	1.77	3.48
120–149	462	1.2130	1.0770	−.9188	1.20	2.18
150–179	388	1.1872	.9053	−.8630	1.33	2.56
180–269	383	1.1708	.8692	−.7516	1.42	3.25
270–359	323	1.2154	1.1061	−.5675	1.26	2.99
360–719	535	1.0760	1.1294	−.1109	2.85	5.54
720–1,079	352	1.1141	2.1765	.1976	3.21	5.34
1,080–1,439	308	1.0886	2.8242	.1976	3.52	5.70
1,440–1,799	244	1.0353	3.9121	.2292	3.57	5.60
1,800–2,159	173	1.0523	5.3179	.1673	3.23	5.11
2,160–	298	.9213	5.5151	.0827	4.21	5.93

NOTE.—Means and variances are multiplied by 100. Skewness is measured by the third moment from the mean divided by the second moment to the ½ power. Kurtosis is computed by dividing the fourth moment about the mean by the square of the second moment about the mean. The studentized range is the range of observations in the sample divided by the square root of the sample variance.

changes per category. Mean, variance, skewness, and kurtosis are shown in columns 2–5. The studentized range values in column 6 are an overall measure of normality. These statistics, which are quite similar to those reported by Bildersee (1975), suggest that Treasury security returns are skewed and leptokurtic.

IV. Tests of Hypotheses

We performed several tests of the implications for duration measures $D1$ to $D7$ for holding periods of 1, 3, 6, and 12 months.[7] Results are presented for eight periods: the overall period January 1947 to Decem-

7. The six duration measures $D1$–$D6$ depend on the cash flow stream $C(t)$, the current market price $P(m)$, and the spot rate $R(t)$ associated with each cash flow. The duration measures are calculated at the beginning of a holding period. Yield to maturity of each security is based on a proxy for the spot rate $R(t)$ and actual market prices for $P(m)$.

ber 1976; two 10-year subperiods, January 1957 to December 1966 and January 1967 to December 1976; and five subperiods starting in January 1952 and covering 5 years each. Concise summary tables are reported with other results highlighted in the discussion.[8]

A. Simple Duration as Measure of Basis Risk

The major tests of duration $D1$ are tabulated in table 2. Results are presented in four different versions of the return-duration regression model in equation (7) to facilitate comparisons with results presented later in the paper. Panel D is based on equation (7) exactly, but in panels A–C one or more of the variables in equation (7) is suppressed. For each period and model table 2 shows the average $\bar{\gamma}j$ of the 1-month holding period regression coefficient estimate and first order autocorrelations $\rho(\gamma j)$ of the various monthly γj_τ.

The table also shows t-statistics for testing the hypothesis that $\bar{\gamma}j = 0$. These are computed by taking the ratio of the average $\bar{\gamma}j$ times the square root of the number of months in the sample period considered over the standard deviation of the monthly estimate. This test procedure is essentially the one used by Fama and MacBeth (1973). In essence, a time series for $\bar{\gamma}j$ is created for each period τ in order to correct for any statistical biases in obtaining the period-by-period estimates of $\bar{\gamma}j$. The time series allows us to obtain confidence intervals for the significance tests. Finally, $\bar{R}^2$ and $S(\bar{R}^2)$ values are presented. These values are the mean and standard deviation of the monthly R^2 values adjusted for degrees of freedom.

The results in panels B and D do not allow us to accept the linearity hypothesis, that is, that the relation between returns and simple duration $D1$ is linear. In panel B, the value of $t(\bar{\gamma}3)$ for the overall period 1/47–12/76 is -3.82. The 5- and 10-year subperiod results also show large and negative t-values for $\bar{\gamma}3$. This persistent negative sign for $\bar{\gamma}3$, along with the high positive values for $\bar{\gamma}1$, indicates that simple duration overstates the basis risk for longer-term Treasury securities.

Evidence in panels C and D shows that duration normalizes securities of different coupon sizes.[9] In panel C the value of $t(\bar{\gamma}4)$ for the

8. All tables that are not presented here are available from the authors upon request.

9. Including coupon securities with zero coupon securities, as in table 2, may cause a bias. Moreover, some might argue that our duration calculations involve measurement error. Since the duration of Treasury bills is maturity, the measurement problem is circumvented. Therefore we divided the data in table 2 into two segments: Treasury bills and Treasury notes and bonds. We found results for notes and bonds to be essentially the same as those in table 2. Treasury bill results require further clarification. Returns for Treasury bills are positively related to maturity in every period. A reason for the positive coefficients $\bar{\gamma}2$ is that yield curves at the short end were predominantly upward sloping during the period sampled and that interest rates were rising. Consequently, realized returns on Treasury bills are on average greater than promised returns, so Treasury bill returns are proportional to maturity. Despite the positive sign, however, the results still do not allow us to accept the hypothesis that the relation between Treasury bill returns and maturity is linear.

TABLE 2 **Cross-sectional Regression Results for Simple Duration $D1$**

Period	$\bar{\gamma}1$	$\bar{\gamma}2$	$\bar{\gamma}3$	$\bar{\gamma}4$	$\hat{\rho}(\gamma1)$	$\hat{\rho}(\gamma2)$	$\hat{\rho}(\gamma3)$	$\hat{\rho}(\gamma4)$	$t(\bar{\gamma}1)$	$t(\bar{\gamma}2)$	$t(\bar{\gamma}3)$	$t(\bar{\gamma}4)$	$\bar{R}^2$	$S(\bar{R}^2)$
A. $r_\tau(m) = \gamma1_\tau + \gamma2_\tau D1_{\tau-1}(m) + e_\tau(m)$														
1/47–12/76	36.20	−.15			.587	.057			27.65	−1.32			.494	.339
1/57–12/66	29.63	−.08			.421	.178			26.10	−.53			.506	.326
1/67–12/76	53.06	−.24			.232	−.005			24.74	−1.02			.435	.331
1/52–12/56	15.63	−.12			.368	.137			13.13	−.79			.580	.364
1/57–12/61	26.66	.06			.306	.204			16.17	.21			.560	.326
1/62–12/66	32.60	−.21			.500	.078			22.07	−1.45			.451	.319
1/67–12/71	50.35	−.32			.350	−.011			24.71	−.78			.454	.374
1/72–12/76	55.77	−.15			.182	.019			14.82	−.67			.416	.282
B. $r_\tau(m) = \gamma1_\tau + \gamma2_\tau D1_{\tau-1}(m) + \gamma3_t D1^2_{\tau-1}(m) + e_\tau(m)$														
1/47–12/76	31.85	.54	−1.06		.489	−.002	−.010		25.37	2.52	−3.82		.634	.296
1/57–12/66	24.70	.76	−1.35		−.047	−.952	−.070		21.92	2.76	−3.36		.629	.269
1/67–12/76	48.22	.42	−.95		.188	.000	.032		23.31	.96	−1.81		.570	.315
1/52–12/56	13.42	.32	−.69		.162	.024	−.152		13.49	1.24	−1.63		.770	.267
1/57–12/61	22.33	.78	−1.22		.043	.020	−.060		12.60	1.88	−2.21		.642	.286
1/62–12/66	27.08	.74	−1.44		−.327	.147	.079		20.22	2.02	−2.52		.615	.252
1/67–12/71	47.69	.23	−1.03		.280	−.024	.041		15.19	.31	−1.07		.554	.358
1/72–12/76	48.75	.60	−.87		.066	.053	−.011		17.91	1.25	−2.03		.586	.267

Statistic

C. $r_\tau(m) = \gamma 1_\tau + \gamma 2_\tau D1_{\tau-1}(m) + \gamma 4_\tau C(m) + e_\tau(m)$

1/47–12/76	35.12	−.26		1.55	.686	.026		.111	30.24	−2.19		3.25	.553	.317
1/57–12/66	28.92	−.18		1.57	.543	.144		−.090	29.07	−1.14		2.21	.560	.293
1/67–12/76	51.41	−.35		1.27	.284	−.020		−.037	29.50	−1.54		2.77	.469	.326
1/52–12/56	14.94	−.21		2.08	.620	−.046		−.150	17.36	−1.16		1.22	.708	.290
1/57–12/61	25.71	−.09		2.33	.442	.155		−.102	18.32	−.32		1.74	.612	.291
1/62–12/66	32.13	.27		.88	.580	.101		−.106	24.83	−1.78		1.75	.507	.288
1/67–12/71	49.37	.48		1.63	.377	−.019		.029	25.94	−1.16		3.29	.462	.374
1/72–12/76	53.45	−.23		.92	.231	−.021		−.071	18.32	−1.11		1.18	.477	.272

D. $r_\tau(m) = \gamma 1_\tau + \gamma 2_\tau D1_{\tau-1}(m) + \gamma 3_\tau D1^2_{\tau-1}(m) + \gamma 4_\tau C(m) + e_\tau(m)$

1/47–12/76	31.91	.58	−1.10	−.40	.488	−.018	−.005	−.211	24.83	2.38	−3.50	−1.04	.655	.279
1/57–12/66	24.38	.95	−1.54	−1.00	−.061	−.099	−.070	−.303	20.33	2.92	−3.35	−1.68	.652	.241
1/67–12/76	48.63	.32	−.82	.28	.203	.005	.042	.068	22.99	.64	−1.40	.68	.580	.311
1/52–12/56	13.53	.39	−.77	−.57	.209	−.091	−.139	−.269	14.35	1.14	−1.50	−.44	.809	.217
1/57–12/61	21.81	.97	−1.42	−1.05	.003	−.076	−.065	−.313	11.30	1.90	−2.11	−.91	.674	.244
1/62–12/66	26.96	.92	−1.67	−.95	−.331	−.113	−.076	−.169	19.80	2.29	−2.62	−3.09	.631	.239
1/67–12/71	48.32	−.17	−.63	1.40	.273	−.018	−.052	.053	14.46	−.20	−.38	2.18	.562	.355
1/72–12/76	48.93	.80	−1.00	−.84	.090	.042	−.019	−.099	18.64	1.61	−2.20	−1.01	.597	.262

NOTE.—$\bar{\gamma}1$, $\bar{\gamma}2$, $\bar{\gamma}3$, and $\bar{\gamma}4$ are the arithmetic means of the monthly cross-sectional estimates using the multiple regression in eq. (7). $\hat{\rho}(\gamma 1)$, $\hat{\rho}(\gamma 2)$, $\hat{\rho}(\gamma 3)$, and $\hat{\rho}(\gamma 4)$ are estimated autocorrelations of the estimates for lag one. The t-values $t(\bar{\gamma}1)$, $t(\bar{\gamma}2)$, $t(\bar{\gamma}3)$, and $t(\bar{\gamma}4)$ test the null hypothesis that the mean values are zero. $\bar{R}^2$, $S(\bar{R}^2)$ are the average adjusted R^2 and its standard deviation for the monthly regressions. The $\bar{\gamma}2$ numbers are multiplied by 100.

overall period 1/47–12/76 is −2.19. Once nonlinearity is accounted for in panel C, however, $t(\bar{\gamma}4)$ is only −1.04, probably because the coupon as an independent variable is explaining part of the nonlinearity. The negative sign of $t(\bar{\gamma}4)$ is consistent with the premise that all else being equal, price volatility is lower for higher coupon securities (see Malkiel 1962).

The critical completeness hypothesis is rejected. Panel A of the table indicates that over the 30-year period, or any subperiod, the value of $t(\bar{\gamma}2)$ is not large, which is the case for all models (see panels B, C, and D). On average, a negative trade-off existed between returns and $D1$. Except in the 4-year period 1/57–12/60 the values of $\bar{\gamma}2$ are systematically negative in panel A.

The intercept term measures the average level of interest rates during the various periods. Interest rates generally were rising over the 30-year period, as the increasing $\bar{\gamma}1$ values for the 5-year subperiods indicate.

The time series behavior of $\bar{\gamma}1$ deserves further clarification. The serial correlations $\rho(\bar{\gamma}1)$ in table 2 are predominantly greater than zero. Excess returns (over the risk-free rate) can be used to approximate unexpected returns. Working with excess returns as the dependent variable generally reduces the autocorrelation of the intercept term. The explanatory power of the regressions, however, is no different whether one is working with returns or with excess returns, because the regressions are cross-sectional.

The behavior through time of $\gamma2$, $\gamma3$, and $\gamma4$ is consistent with the efficiency hypothesis. The serial correlations $\hat{\rho}(\gamma2)$, $\hat{\rho}(\gamma3)$, and $\hat{\rho}(\gamma4)$ are always statistically close to zero. Although we have not reported them here, serial correlations of the residuals $e_\tau(m)$ are also close to zero.

We also carried out the analysis in table 2 for holding periods of 3, 6, and 12 months. Our motivations were twofold. First, the duration literature does not generally specify any particular holding period for the return-duration relation. Second, it is important to determine the robustness of duration over varying holding periods. The length of the holding periods turns out not to affect our basic conclusions. The linearity and completeness hypotheses are still not accepted, but the efficiency hypothesis is.[10]

In sum, the data are not consistent with the hypothesis that price

10. We repeated the analysis in table 2 for portfolios of Treasury securities. We ranked securities on the basis of their simple durations each month and assigned them to 10 portfolios. Returns and durations for the 10 portfolios are obtained by equally weighting the assigned securities. The duration of a portfolio is a weighted sum of the durations of the individual securities in a portfolio. Coupons for the portfolios constructed are difficult to ascertain, so we did not perform regressions on coupon terms. The results are again very similar: for portfolios of Treasury securities the linearity and completeness hypotheses are not accepted, but the efficiency hypothesis continues to be accepted.

volatility of Treasury securities is adequately measured by simple duration. Although simple duration normalizes coupon differences, the relation between security returns and simple duration is not linear. Different holding period lengths do not affect the conclusions.

B. *Other Durations as Measures of Basis Risk*

We repeated all the tests we performed for $D1$ for the five other duration measures defined in equation (3). The tests used various different values for the parameter α in duration $D6$. Our results for durations $D2$–$D6$ and maturity m are similar to those shown in table 2. All the conclusions we reached for $D1$ regarding the linearity, completeness, and efficiency hypotheses are exactly the same for these other duration measures and maturity. In other words, even though the $D2$–$D6$ measures purport to reflect yield-curve movements more accurately than simple duration, on average none of them performs better than simple duration; that is, at the margin none of these duration measures explains price volatility caused by shape changes any better than any other.

Results for these other duration measures are too voluminous to report, so we will highlight some of the main results. Of duration measures $D1$–$D6$ the explanatory power of duration $D1$ was at least as good as, or better than, that of the other duration measures. All the duration measures overstate the riskiness of longer-term Treasury securities, which is evident by the increase in the $\bar{R}^2$ due to a large and negative value for $t(\bar{\gamma}3)$ when a nonlinear variable is included. In sum, all fail as linear proxies for price volatility.

The most important observation about these additional tests is that the average explanatory power of the regressions with maturity is almost as good as that of any of the other duration measures $D1$–$D6$. Apparently, unpredictable movements of interest rates have been such that, on average, durations $D1$–$D6$ are no better measures of basis risk than maturity.

C. *Stochastic Duration as a Measure of Basis Risk*

Computation of the stochastic duration measure $D7$ in equation (5) is more involved than the other duration measures. Estimates of the parameters π, μ, β, and σ^2 must be obtained for the instantaneous spot-rate process in equation (4). Statistical estimation of the parameters of the spot rate process is beyond the scope of this paper. Instead, we use annualized estimates of the parameters reported by Cox et al. (1979): $\pi = 0$, $\mu = 5.623\%$, $\beta = .692$, and $\sigma^2 = .00608$. The parameters μ, β, and σ^2 of the interest rate process are based on a weighted least squares of a series of weekly 91-day Treasury bill auctions over the period 1967–76. Cox et al. use these estimates to illustrate the superiority of stochastic duration as a measure of basis risk over Macaulay's (1938) duration. Given these estimates, we compute stochastic duration as in equation (5).

Results for the period 1967–76 are presented in table 3 in a format like that of table 2. A review of table 3 reveals that stochastic duration explains price volatility about as well as duration $D1$ or any of the other duration measures.

The results in panels B and D refute the hypothesis that the relation between returns and stochastic duration $D7$ is linear. In panel B, the negative value of $\bar{\gamma}3$ for the period 1/67–12/76, along with the large positive value for $\bar{\gamma}1$, implies that stochastic duration overstates the riskiness of longer-term securities. However, the percentage increases in the average adjusted R^2 values beween panels A and B are not as great for stochastic duration as for the other duration measures, which would suggest that stochastic duration does not overestimate the riskiness of longer-term securities as much as the other duration measures do.

The results in panels C and D affirm that stochastic duration normalizes coupon differences. The completeness hypothesis cannot be accepted, because the small and negative value of $t(\bar{\gamma}2)$ in panel A indicates that stochastic duration is not a complete measure of risk. The time series behavior of $\gamma 1$, $\gamma 2$, $\gamma 3$, $\gamma 4$, and $e_\tau(m)$ in panels A–D is consistent with the hypothesis that the Treasury market is efficient.

In sum, stochastic duration as calculated with the parameter estimates reported by Cox, Ingersoll, and Ross is not substantially superior to other durations or even to maturity as a measure of basis risk. This conclusion is contrary to simulation results reported by Cox et al. to illustrate that stochastic duration is more valuable in risk measurement than Macaulay's duration. There are at least three reasons for the relatively unimpressive performance of stochastic duration. First, liquidity premiums are assumed to be zero in our calculations and the numbers reported by Cox et al. To the extent that liquidity premiums explain the term structure, they must be incorporated into stochastic duration; they will be particularly important at the short end of the yield curve. Second, the interest rate process may be misspecified. A correct measure of stochastic duration depends on the true spot rate process. It is possible that the process assumed by Cox et al. and that we used in our tests is not the right one. Third, even if an elastic random-walk model correctly describes the stochastic nature of spot rates, the statistical estimation procedures may not be appropriate. Solutions to any of these potential problems require substantial research efforts that are beyond the scope of this paper.

D. Correlation among Duration Measures

Our findings so far do not indicate a noticeable difference in the explanatory power of various duration measures. This result is contrary to

TABLE 3 Cross-sectional Regression Results for Stochastic Duration $D7$

Period	Statistic													
	$\bar\gamma 1$	$\bar\gamma 2$	$\bar\gamma 3$	$\bar\gamma 4$	$\hat\rho(\gamma 1)$	$\hat\rho(\gamma 2)$	$\hat\rho(\gamma 3)$	$\hat\rho(\gamma 4)$	$t(\bar\gamma 1)$	$t(\bar\gamma 2)$	$t(\bar\gamma 3)$	$t(\bar\gamma 4)$	$\bar R^2$	$S(\bar R^2)$
	A. $r_\tau(m) = \gamma 1_\tau + \gamma 2_\tau D7_{\tau-1}(m) + e_\tau(m)$													
1/67–12/76	52.16	−.24			.251	.007			27.98	−.69			.452	.351
1/67–12/71	50.69	−.38			.286	−.009			22.49	−.70			.456	.378
1/72–12/76	53.62	−.09			.226	.036			18.01	−.22			.447	.326
	B. $r_\tau(m) = \gamma 1_\tau + \gamma 2_\tau D7_{\tau-1}(m) + \gamma 3_\tau D7^2_{\tau-1}(m) + e_\tau(m)$													
1/67–12/76	46.57	.99	−.03		.309	.006	−.056		27.13	2.19	−3.11		.515	.338
1/67–12/71	43.88	1.19	−.04		.398	.063	.011		19.17	1.87	−2.49		.514	.373
1/72–12/76	49.26	.80	−.02		.210	−.052	−.167		19.46	1.22	−1.85		.516	.303
	C. $r_\tau(m) = \gamma 1_\tau + \gamma 2_\tau D7_{\tau-1}(m) + \gamma 4_\tau C_\tau(m) + e_\tau(m)$													
1/67–12/76	51.58	−.45		1.42	.260	−.017		−.094	28.87	−1.21		3.44	.461	.350
1/67–12/72	50.08	−.71		2.31	.317	−.025		−.044	23.21	−1.17		3.41	.464	.377
1/72–12/76	53.08	−.19		.53	.221	−.003		−.355	18.61	−.44		1.18	.459	.323
	D. $r_\tau(m) = \gamma 1_\tau + \gamma 2_\tau D7_{\tau-1}(m) + \gamma 3_\tau D7^2_{\tau-1}(m) + \gamma 4_\tau C_\tau(m) + e_\tau(m)$													
1/67–12/76	46.99	.81	−.03	.45	.320	−.001	−.035	.015	26.01	1.49	−2.59	1.16	.520	.337
1/67–12/71	45.23	.60	−.03	1.29	.353	.033	.020	.014	16.93	.74	−1.70	1.89	.523	.369
1/72–12/76	48.76	1.03	−.02	−.39	.268	−.048	−.120	−.219	20.03	1.39	−1.99	−1.18	.516	.305

NOTE.—See table 2, n.

TABLE 4 Mean and Standard Deviation of Monthly Cross-sectional Correlations between Duration Measures

Duration Measure	Correlations						
	$D1$	$D2$	$D3$	$D4$	$D5$	$D6$	$D7$
A. 1/47–12/76 (all securities):							
$D1$	1	.9986	.9422	.9912	.9471	.9004	
$D2$	(.0019)	1	.9495	.9941	.9530	.8898	
$D3$	(.0275)	(.0250)	1	.9741	.9995	.7386	
$D4$	(.0059)	(.0059)	(.0162)	1	.9767	.8449	
$D5$	(.0249)	(.0250)	(.0008)	(.0154)	1	.7465	
B. 1/47–12/76 (excluding Treasury bills):							
$D1$	1	.9988	.9539	.9343	.9574	.9085	
$D2$	(.0010)	1	.9591	.9968	.9614	.8988	
$D3$	(.0124)	(.0071)	1	.9780	.9996	.7614	
$D4$	(.0031)	(.0007)	(.0039)	1	.9795	.8637	
$D5$	(.0083)	(.0081)	(.0004)	(.0051)	1	.7680	
$D6$	(.0276)	(.0330)	(.0544)	(.0408)	(.0497)	1	
C. 1/67–12/76 (all securities):							
$D1$	1	.9991	.9233	.9921	.9281	.8737	.9494
$D2$	(.0008)	1	.9298	.9945	.9337	.8660	.9431
$D3$	(.0316)	(.0278)	1	.9610	.9996	.6696	.7673
$D4$	(.0035)	(.0014)	(.0176)	1	.9640	.8138	.9082
$D5$	(.0285)	(.0263)	(.0004)	(.0162)	1	.6762	.7745
$D6$	(.0245)	(.0284)	(.0846)	(.0380)	(.0816)	1	.9349
$D7$	(.0345)	(.0378)	(.1223)	(.0556)	(.1185)	(.0097)	1

NOTE.—Figures above the ones are average correlation coefficients, and those below the ones, in parentheses, are corresponding standard deviations.

assertions made in the literature and can easily be explained if we examine the degree of association among the duration measures. Estimated correlation coefficients are computed as the cross-sectional correlations between all duration measures for each month during the sample period January 1947 to December 1976. Panel A of table 4 presents averages and standard deviations of the period-by-period correlation coefficients using all securities. Panel B excludes Treasury bills that have no coupon. Numbers above the ones are average correlation coefficients, and numbers below the ones in parentheses are corresponding standard deviations. Correlations among duration measures $D1$–$D6$ in panels A and B are on average very large, meaning that duration measures $D1$–$D6$ are not independent of one another.

Panel C of table 4 tabulates the same statistics for the 10-year sample period 1/67–12/76 using all securities. Stochastic duration $D7$ can now be correlated with the other duration measures $D1$–$D6$. Stochastic duration has the highest average correlation with simple duration $D1$ (.9494) and the lowest with $D3$. In general, all the average correlations are large, a fact implying that stochastic duration is not independent of the other duration measures.

The average coefficients of correlation suggest that duration measures $D1$–$D7$ rank securities in a similar fashion, a conclusion that may be somewhat surprising to readers familiar with duration examples in the literature. The literature may include, for example, comparisons of duration measures for a hypothetical 10-year pure discount security with those for a 10-year high coupon security. In their attempt to support an argument, creators of such examples use very extreme cases. Coupon differences are generally within a narrower range, so on average all duration measures rank securities in the same order of riskiness. Unpredictable movements of the yield curve over time would have made this ranking virtually useless for explaining price volatility.

E. *The Usefulness of Durations for Immunization*

We conducted a final test of the usefulness of duration $D1$ as a measure of basis risk, creating fixed-duration portfolios for each period. The results we found have important implications for immunization strategies and fixed-income portfolio performance measurement.[11]

For a given duration, we construct a portfolio for each period by randomly selecting two securities and weighting the securities so that the duration of the portfolio is, for example, 5 months. We do this 30 times to create 30 different portfolios of two securities each with a duration of 5 months. Another 30 portfolios are constructed of three securities each with a duration of 5 months, and another 30 are constructed by randomly selecting four securities. This procedure is repeated for a 5-month duration nine times, so that the last 30 portfolios consist of 10 randomly selected securities. This portfolio selection procedure is continued 17 times each period for durations up to 85 months in increments of 5 months. Seventeen means and variances of total returns (based on the definition of eq. [8]) are calculated for each of nine fixed-duration portfolios of size 2, 3, . . . , 10 each period.

Results for nonoverlapping 3- and 6-month holding periods over the sample period January 1947 to December 1976 are summarized in table 5. Columns 1 and 4 show the number of 3- and 6-month periods for which portfolios can be formed. (In some holding periods it is not possible randomly to generate portfolios for all the fixed durations.)

If duration is an appropriate measure of basis risk, portfolios of varying size but of the same duration should have equal returns on average. To test the equality of average returns we performed an analysis of variance for each holding period. The null hypothesis to be tested is that all nine of the means are equal. Columns 2 and 5 of table 5 list

11. Our tests in this section supplement tests performed by Ingersoll (1981). In immunization tests using duration Ingersoll reports root mean square errors of returns for different portfolios formed systematically to have the same duration. He concludes that duration performed no better than maturity matching.

TABLE 5 Return Performance of Portfolios with Equal Duration $D1$

Portfolio Duration (in Months)	Three-Month Holding Period: Number of Periods (1)	Three-Month Holding Period: Percentage of Means Not Equal (2)	Three-Month Holding Period: Percentage of Variances Not Equal (3)	Six-Month Holding Period: Number of Periods (4)	Six-Month Holding Period: Percentage of Means Not Equal (5)	Six-Month Holding Period: Percentage of Variances Not Equal (6)
5	71	43.66	95.77	...	...	...
10	71	36.61	94.36	33	21.21	96.69
15	72	25.00	98.61	38	36.84	100.00
20	73	43.83	91.78	42	26.19	100.00
25	75	42.66	88.00	44	29.54	100.00
30	75	41.33	88.00	44	25.00	97.77
35	75	38.66	82.66	44	36.36	94.45
40	75	45.33	84.00	44	34.09	100.00
45	75	53.33	81.33	44	40.90	90.90
50	74	47.29	87.00	44	43.18	93.18
55	68	55.88	86.76	44	56.81	95.45
60	48	66.66	89.58	44	63.63	95.45
65	37	64.86	91.89	43	55.58	88.37
70	31	61.29	100.00	43	76.74	90.69
75	29	62.06	100.00	42	71.42	92.85
80	26	50.00	100.00	42	69.04	88.09
85	21	42.85	100.00	42	73.80	88.00

NOTE.—F-tests were performed for each duration to test the equality of means over the number of periods. A Bartlett test based on χ^2 values was used for each duration to determine whether the variances are equal over the number of periods. The percentages reported represent the number of statistically significant F-values and χ^2 values at the 1% level divided by the number of holding periods examined.

the percentage of F-values statistically significant at the 1% level for 3- and 6-month holding periods, respectively. For example, given a duration of 5 months, the mean returns of portfolios of size 2–10 are not equal in 43.66% (or 31) of the 71 holding periods investigated. Table 5 reveals that the percentage of means not equal generally increases as duration length increases.

Another test involves the variances of the portfolio returns. If duration is an appropriate measure of basis risk, portfolios of varying size but of the same duration should have equal variances. Bartlett's test, which we used to determine whether the variances are equal, assumes that each of the populations is normal and that independent random samples are obtained for each population. Columns 3 and 6 of table 5 give the percentage of χ^2 values that are statistically significant at a 1% level. From 81.33% to 100.0% of the χ^2 values for a given duration are large enough to be significant.

To test whether any of the mean and variance percentages in columns 2, 3, 5, and 6 of table 5 are significantly different from zero, a two-tailed t-test was performed. Assuming a binomial sampling distribution for the percentage of occurrences, the null hypothesis is that the percentages are equal to zero. Using a 1% level of significance we cannot conclude that any of the percentages are zero.

In sum, the evidence seems to indicate that simple duration is not useful for cross-sectional comparisons of Treasury security risk. Our results also shed light on some issues related to fixed-income portfolio performance and immunization. Bierwag and Kaufman (1978) have argued that the duration of a portfolio is not explicitly taken into account in simulations of bond portfolio performance; that is, portfolios with varying compositions may actually be of similar duration, which creates confusion in comparing different portfolio strategies. The results shown in table 5, however, indicate that the concern of Bierwag and Kaufman is unfounded. The variability of constant duration portfolios in both mean and variance of returns suggests that duration is not the appropriate index to hold constant in comparing fixed-income portfolio strategies.

As for immunization of bond portfolios, Fisher and Weil (1971) argue that it is possible to achieve the promised yield of a bond by buying a portfolio with duration equal to the assumed investment horizon and appropriately rebalancing as coupons are received. As duration is always less than term to maturity, a security with a duration equal to the investment horizon actually has a maturity beyond the investment horizon. Capital gains or losses at the end of the investment horizon compensate for losses or gains or reinvestment of coupons during the holding period. Our results in table 5, however, demonstrate that portfolios with the same initial duration can have completely different returns at the end of the holding period. In other words, immunizing with five securities may not be the same as immunizing with 10 securities even though both portfolios have the same initial duration.

F. Simple-minded Factor Models of Price Volatility

To develop additional alternative hypotheses (maturity m is an alternative), we performed a factor analysis on bond portfolio returns to see what percentage of variation is explained by up to four factors compared with how much variation is explained by single specific factors, such as duration measures $D1$–$D7$. We are not suggesting that a four-factor model is the best one, but we do believe that it is a useful starting point for the examination of multiple-factor duration models. In other words, we simply want to test the explanatory power of different duration measures for holding period returns against simple one-, two-, three-, and four-factor models of interest rate behavior.

For our tests of the multiple-factor model it is necessary to construct Treasury portfolios holding maturity approximately constant each month because the maturity of a bond changes with the passage of time. For example, 1 year after issue a 3-year Treasury bond becomes a 2-year Treasury bond, necessitating a portfolio change. Constant maturity portfolios help to alleviate this nonstationarity problem and enable us to create the time series of Treasury portfolio returns we need in order to perform a factor analysis.

We systematically construct constant maturity portfolios to consist of all securities within a maturity range. Return and duration measures over any maturity range are computed as a weighted average of all the individual returns and durations that fall within the range.

We select specific maturity ranges to avoid any empty portfolios, constructing 12 portfolios during the 18-year period 1959–76.[12] Thirty-day increments are taken from 60 days up to a half year, 90-day increments up to 1 year, 360-day increments thereafter to about 6 years, and all securities with maturities beyond 2,160 days. Smaller increments are used for short maturities because the government has more short-term than long-term debt outstanding. For the same reason the 2,160-day cutoff ensures that each maturity range has at least one security in it.

We factor analyze the 12 Treasury portfolio return series to obtain factor loadings for up to four factors.[13] A factor model of period-by-period returns that allows us to use observed average returns to test the

12. Fewer portfolios are possible starting before 1959 if none is to be empty. In addition, the portfolios would have fewer securities in them, especially in the 360-day or longer maturity ranges, because the Treasury issued fewer long-term securities.

13. See Lawley and Maxwell (1963) for a complete and easy-to-read description of factor analysis.

alternative hypotheses is

$$\tilde{r}_\tau(p) = \tilde{\gamma}1_\tau\ \tilde{\gamma}2_\tau\lambda1(p) + \tilde{\gamma}3_\tau\lambda2(p) + \tilde{\gamma}4_\tau\lambda3(p) + \tilde{\gamma}5_\tau\lambda4(p) + \tilde{e}_\tau(p), \tag{11}$$

where $\tilde{r}_\tau(p)$ is the one-period continuously compounded return on portfolio $p = 1, 2, \ldots, 12$; and $\lambda k(p)$ denotes the factor loadings with $k = 1, 2, 3, 4$. The $\tilde{\gamma}_1$, $\tilde{\gamma}_2$, $\tilde{\gamma}_3$, and $\tilde{\gamma}_5$ vary stochastically from period to period.

Means and standard deviations of R^2 values adjusted for degrees of freedom obtained from period-by-period regressions using the multiple-factor model in equation (11) are given in table 6. A 1-month holding period is used, and we again follow the test procedure of Fama and MacBeth (1973). Adjusted R^2 values are reported over the period 1959–76 and several subperiods. The model in equation (11) is a four-factor model. A one-factor model assumes $\tilde{\gamma}3_\tau$, $\tilde{\gamma}4_\tau$, and $\tilde{\gamma}5_\tau$ in equation (11) are zero; a two-factor model sets $\tilde{\gamma}4_\tau$ and $\tilde{\gamma}5_\tau$ equal to zero; and $\tilde{\gamma}5_\tau$ is zero in a three-factor model. Comparable means and standard deviations of R^2 values adjusted for degrees of freedom are presented in table 6 using model (7) for duration $D1$, with $\tilde{\gamma}3_\tau$ and $\tilde{\gamma}4_\tau$ assumed equal to zero.

Over the period 1959–76 the average $\bar{R}^2$ for duration $D1$ was .577. One, two, three, and four factors had an average $\bar{R}^2$ of .550, .588, .645, and .759, respectively. The standard deviations of the $\bar{R}^2$ values are systematically smaller as the number of factors is increased. Thus a four-factor model explains about 18% more return variability than duration $D1$.

Subperiod results reported in table 6 exhibit characteristics that are identical to those for the longer period. As the number of factors is increased from one to four the explanatory power of the factor model increases relative to duration $D1$. Similar results are also obtained for duration measures $D2$–$D7$. This is to be expected, given the high correlations among different duration measures reported in table 4.[14]

These results indicate that a multiple-factor model explains more variability of monthly holding-period returns than different duration measures. An immunization strategy can be implemented on the basis of the factor approach for measuring economic significance rather than statistical significance. Ingersoll (1981) tests a multiple-factor model during the period 1950–79 and concludes that a simple two-factor model immunizes better than durations $D1$ or $D6$. Our findings, in

14. Cooper (1977) suggests that a linear combination of $D1$, $D2$, $D3$, $D4$, and $D5$ might explain bond price changes better than any single measure, but he does not pursue the suggestion. The multicollinearity among duration measures $D1$ to $D5$ reported in table 5 indicates the futility of implementing Cooper's suggestion in a meaningful way.

TABLE 6 **Mean and Standard Deviation of Monthly Adjusted R^2 Values from Cross-sectional Regressions for Duration $D1$ and Multiple Factors**

	Sample Period											
	1/59–12/76		1/59–12/66		1/67–12/76		1/62–12/66		1/67–12/71		1/72–12/76	
Variable	$\bar{R}^2$	$S(\bar{R}^2)$	$\bar{R}^2$	$S(\bar{R}^2)$	$\bar{R}^2$	$S(\bar{R}^2)$	$\bar{R}^2$	$S(\bar{R}^2)$	$\bar{R}^2$	$S(\bar{R}^2)$	$\bar{R}^2$	$S(\bar{R}^2)$
$D1$	.577	(.357)	.577	(.365)	.577	(.350)	.550	(.371)	.602	(.372)	.553	(.325)
1 factor	.555	(.360)	.536	(.349)	.569	(.367)	.505	(.353)	.570	(.373)	.568	(.361)
2 factors	.588	(.355)	.565	(.333)	.606	(.371)	.542	(.340)	.582	(.396)	.629	(.344)
3 factors	.645	(.315)	.592	(.312)	.687	(.311)	.570	(.321)	.633	(.375)	.740	(.216)
4 factors	.759	(.283)	.736	(.303)	.777	(.265)	.735	(.320)	.744	(.318)	.810	(.194)

NOTE.—The $D1$ cross-sectional regressions are based on eq. (7) with $\tilde{\gamma}3_\tau = \tilde{\gamma}4_\tau = 0$. The multiple cross-sectional regression models are represented by eq. (11) with the appropriate coefficients set equal to zero.

conjunction with those of Ingersoll, suggest that multiple-factor durations may provide a practical solution to immunization.

V. Summary and Conclusions

In this paper we have used actual market data to test Macaulay's (1938) duration, Bierwag's (1977) duration, three durations derived by Cooper (1977), Khang's (1979) duration, and the stochastic duration proposed by Cox et al. (1979). Our aim has been to determine whether any of these durations is an adequate measure of basis risk for Treasury securities during the sample period January 1947–December 1976.

Our results do not support the important testable implications of any of these durations as measures of basis risk. Specifically, we cannot accept the hypothesis that the relation between returns and duration is linear. On average there seem to be persistent stochastic nonlinearities for all seven durations. This means that all durations substantially overstate the volatility of longer term-to-maturity securities. Thus we cannot reject the hypothesis that additional measures of risk systematically affect average price changes. In fact, we introduced and tested a four-factor model that shows some promise for explaining price variability. On the positive side, all duration measures account for coupon differences, with volatility lower for high coupon securities. Finally, our results suggest that none of the duration measures is useful for fixed-income performance comparisons and that immunization strategies based on duration may not work.

Our empirical results contrast sharply with the analytical claims made by authors of the duration measures. Even the theoretically more appealing stochastic duration, which is based on a generalized equilibrium model of the term structure, overstates the riskiness of longer-term securities. Stochastic duration results, however, may be affected by the absence of liquidity premiums plus misspecification and/or misestimation of the underlying true spot rate process.

To sum up, there are periods in which some durations are adequate measures of basis risk, as well as periods when some are obviously superior because of the particular yield-curve movement. Over long periods of time, however, uncertain movements of the term structure do not permit us to conclude that any of the durations tested is a satisfactory surrogate for basis risk.

References

Bierwag, G. O. 1977. Immunization, duration and the term structure of interest rates. *Journal of Financial and Quantitative Analysis* 12 (December): 725–42.

Bierwag, G. O., and Kaufman, George. 1978. Bond portfolio strategy simulations: A critique. *Journal of Financial and Quantitative Analysis* 13 (September): 519–26.

Bildersee, John. 1975. Some new bond indexes. *Journal of Business* 48 (October): 506–25.

Cooper, Ian A. 1977. Asset values, interest-rate changes, and duration. *Journal of Financial and Quantitative Analysis* 12 (December): 701–24.

Cox, John C.; Ingersoll, Jonathan E.; and Ross, Stephen A. 1978. A theory of term structure of interest rates. Working paper. Chicago: University of Chicago.

Cox, John C.; Ingersoll, Jonathan E.; and Ross, Stephen A. 1979. Duration and the measurement of basis risk. *Journal of Business* 52 (January): 51–62.

Fama, Eugene, and MacBeth, James. 1973. Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy* (May/June): 607–36.

Fisher, Lawrence, and Weil, Roman L. 1971. Coping with the risk of interest rate fluctuations: Returns to bondholders from naïve and optimal strategies. *Journal of Business* 44 (October): 408–31.

Hicks, John R. 1939. *Value and Capital*. Oxford: Clarendon Press.

Hopewell, Michael, and Kaufman, George. 1973. Bond price volatility and term to maturity: A generalized respecification. *American Economic Review* 63 (September): 749–53.

Ingersoll, Jonathan. 1981. Is immunization feasible? Evidence from the CRSP data. CRSP Working Paper no. 58. Chicago: University of Chicago, July.

Ingersoll, Jonathan; Skelton, Jeffrey; and Weil, Roman L. 1978. Duration forty years later. *Journal of Financial and Quantitative Analysis* 13 (November): 627–50.

Khang, Chulsoon. 1979. Bond immunization when short-term rates fluctuate more than long-term rates. *Journal of Financial and Quantitative Analysis* 13 (November): 1085–90.

Lanstein, Ronald, and Sharpe, William F. 1978. Duration and security risk. *Journal of Financial and Quantitative Analysis* 13 (November): 653–68.

Lawley, D. N., and Maxwell, A. E. 1963. *Factor Analysis as a Statistical Method*. London: Butterworth's.

Leibowitz, Martin L. 1979. *Bond Immunization: A Procedure for Realizing Target Levels of Return*. New York: Salomon Brothers, October 10.

Leibowitz, Martin L., and Weinberger, Alfred. 1981. *Contingent Immunization: A Risk Control Procedure for Structured Active Management*. New York: Salomon Brothers, January 28.

Macaulay, F. R. 1938. *Some Theoretical Problems Suggested by the Movements of Interest Rates, Bond Yields, and Stock Prices in the U.S. Since 1856*. New York: National Bureau of Economic Research.

Malkiel, Burton. 1962. Expectations, bond prices, and the term structure of interest rates. *Quarterly Journal of Economics*, no. 303 (May), pp. 197–218.

Vasicek, Oldrich. 1977. An equilibrium characterization of the term structure. *Journal of Financial Economics* 5 (November): 177–88.