MEASURING THE TERM STRUCTURE OF INTEREST RATES

J. HUSTON MC CULLOCH*

INTRODUCTION

This paper develops a technique of fitting a smooth curve, called the "discount function," to observations on prices of securities with varying maturities and coupon rates. The yield curve, instantaneous forward interest rates, mean forward interest rates, and consistent values for securities are derived from this discount function. Formulas for estimating the variances of these derived statistics are given. All formulas are worked out for a broad family of discount functions amenable to linear regression. A preferred form for the generalized discount function is described which focuses resolution in the vicinity of concentrations of observations. It is used to compare regression yield curves with those obtained by Durand[1] and those shown in the *Treasury Bulletin*.[2]

THE DISCOUNT FUNCTION

The most fundamental curve describing the term structure of interest rates, the one from which all others must be derived, is the *discount function* $\delta(m)$. It describes the present value of \$1.00 repayable in m years. It is natural to suppose that the discount function is continuously differentiable. We may expect it to be monotonically decreasing.

Except for a few short-term securities called "bills," none of the zero coupon "bonds" whose prices are determined directly from the discount function exists. However, given the maturity m_0 and the coupon rate c of a security, its value p can be computed as the sum of the values of the payments that comprise it:

$$p = 100\delta(m_0) + c\int_0^{m_0}\delta(m)dm\,. \quad (1)$$

For simplicity, we have assumed that the coupons arrive in a continuous stream instead of semiannually. This enables us to use "and interest" prices as quoted.[3]

In order to fit a curve to the discount function by linear regression, we must postulate k continuously differentiable functions $f_j(m)$, and then express it as a constant term plus a linear combination of these functions:

$$\delta(m) = a_0 + \sum_{j=1}^{k} a_j f_j(m)\,.$$

Since the present value of present money is unity, we must have $\delta(0) = 1$. The only way to force the curve through this point is to set $a_0 = 1$ and

$$f_j(0) = 0\,. \quad (2)$$

Therefore the discount function takes the form

$$\delta(m) = 1 + \sum_{j=1}^{k} a_j f_j(m)\,. \quad (3)$$

* Department of Economics, University of Chicago. I am indebted to Professors Reuben A. Kessel, Merton H. Miller, Lester G. Telser, and Henri Theil, and Mr. Michael Mussa for helpful suggestions. The comments and suggestions of the referees were especially fruitful, and resulted in a thorough revision of the paper.

[1] David Durand, *Basic Yields of Corporate Bonds, 1900–1942* (New York: National Bureau of Economic Research, 1942).

[2] *Treasury Bulletin* (Washington, D.C.: Government Printing Office, March 1966), p. 78.

[3] A prorated share of the next coupon is added to the quoted "and interest" price to arrive at the "flat" price at which the security actually changes hands. Today, only bonds in default are quoted

The form of the functions $f_j(m)$ and the value of k are very important to the quality of our fit of the discount function. However, opinions may differ on their specification and there is no indisputably best method. Therefore we will develop all formulas at the present level of generality. Possible forms of these functions and rules for selecting k will be considered at the end of this paper.

Combining (1) and (3) we obtain

$$\begin{aligned} p &= 100\left[1+\sum_{j=1}^{k} a_j f_j(m_0)\right] \\ &\quad + c\int_0^{m_0}\left[1+\sum_{j=1}^{k} a_j f_j(m)\right]dm \\ &= 100\left[1+\sum_{j=1}^{k} a_j f_j(m_0)\right] \\ &\quad + c\left[m_0+\sum_{j=1}^{k} a_j \int_0^{m_0} f_j(m)dm\right] \\ &= 100 + cm_0 \\ &\quad + \sum_{j=1}^{k} a_j\left[100 f_j(m_0) + c\int_0^{m_0} f_j(m)dm\right]. \end{aligned} \tag{4}$$

Setting

$$y = p - 100 - cm_0 \tag{5a}$$

and

$$x_j = 100 f_j(m_0) + c\int_0^{m_0} f_j(m)dm\,, \tag{5b}$$

equation (4) becomes

$$y = \sum_{j=1}^{k} a_j x_j\,. \tag{5c}$$

Because c, m_0, and the postulated functions $f_j(m)$ are given, the right-hand side of (5c) is a linear combination, in unknown constants a_j, of known constants x_j.

flat. However, a few prewar bonds have to be converted to the "and interest" basis before equation (1) can be used.

We chose to start with the discount function, expressing it as a linear combination in unknowns as in (3), because we knew that the linearity of the integration operator in (1) would then also make p a linear combination in these unknowns, permitting estimation of the a_j by linear regression. Previous workers, notably Cohen, Kramer, and Waugh,[4] have instead started with the yield curve $\eta(m)$, a nonlinear transform of $\delta(m)$: $\eta(m) = -(1/m)\ln\delta(m)$. If one were to begin with this yield curve instead, so that

$$\eta(m) = a_0 + \sum_{j=1}^{k} a_j f_j(m)\,,$$

when the value of the coupons was added to that of the principal using (1), he would obtain

$$p = 100\exp\left\{-m_0\left[a_0+\sum_{j=1}^{k} a_j f_j(m_0)\right]\right\}$$
$$+ c\int_0^{m_0}\exp\left\{-m\left[a_0+\sum_{j=1}^{k} a_j f_j(m)\right]\right\}dm\,.$$

This expression is not linear in the a_j and therefore the a cannot be estimated by linear regression without the use of crude approximations. Consequently, we will use the approach of equation (3) and will not develop the yield curve until later.

ESTIMATION OF THE UNKNOWN PARAMETERS a_j

At any moment in time there will not be simultaneous actual sale prices for every security. This is especially true of slow-moving corporate issues. However, there often are enough securities with simultaneously standing bid and asked offers to make inferences about the term

[4] Kalman J. Cohen, Robert L. Kramer, and W. Howard Waugh, "Regression Yield Curves for U.S. Government Securities," *Management Science* 13, no. 14 (December 1966): B168–75.

structure. If we have such observations p_i^b and p_i^a on n securities, define mean prices $\bar{p}_i$ as $\bar{p}_i = (p_i^b + p_i^a)/2$. Let c_i and m_i be the coupon rate and term to final maturity of the ith security.

Instead of (1) holding exactly for the bid-asked mean price, we will find that

$$\bar{p}_i = 100\delta(m) + c_i \int_0^{m_i} \delta(m)dm + \epsilon_i\,, \quad (6)$$

where ϵ_i is an error term with positive variance. These errors can be caused by transactions costs, tax exemption, the capital gains tax treatment of deep discount bonds, callability, convertibility, ineligibility for commercial bank purchase, ability to be surrendered at par in payment of estate taxes (true of so-called flower bonds), risk of default, imperfect arbitrage, and the rigidity which will be introduced by postulating any specific form for $f_j(m)$. Thanks to transactions costs alone, the absolute value of ϵ_i could be as high as $v_i = (p_i^a - p_i^b)/2 + b$, where b is the brokerage fee of 0.5 parts per 100 for the broker-traded corporate issues and zero for dealer-quoted U.S. Governments. The difference between the maximum price to a buyer and the minimum price to the seller is $2v_i$. Because of the other sources of error, the error term will often be larger than v_i. Nevertheless, it will have a variance that is related to v_i. Since the other sources of error are more difficult to quantify, it is convenient to assume that the standard error of ϵ_i is simply proportional to v_i: S.E. $(\epsilon_i) = \sigma v_i$. The value of σ, which is to be measured, gives us an indicator of how well arbitrage is working and of the size of the factors other than coupon and maturity which enter into the value of the securities. If it is as low as 1.0, the bid-asked mean price of most of the securities observed will be within the transactions costs tolerance v_i of a value consistent with the observations on the other securities used. We probably cannot expect the fit to be any better than 1.0.[5]

Adapting (5) to the error term assumption of (6), the regression equation is:

$$y_i = \sum_{j=1}^{k} a_j x_{ij} + \epsilon_i\,, \quad i = 1, 2, \ldots n\,, \quad (7a)$$

and

$$\text{var}\,(\epsilon_i) = \sigma^2 v_i^2\,, \quad (7b)$$

where

$$y_i = \bar{p}_i - 100 - c_i m_i\,, \quad (7c)$$

$$x_{ij} = 100 f_j(m_i) + c_i \int_0^{m_i} f_j(m)dm\,, \quad (7d)$$

and

$$v_i = (p_i^a - p_i^b)/2 + b\,. \quad (7e)$$

We run a weighted least-squares regression on (7) to obtain estimates $\hat{a}_1$, $\hat{a}_2$, $\ldots$, $\hat{a}_k$ and $\hat{\sigma}$ of the parameters a_1, a_2, $\ldots$, a_k and σ. The discount function is then estimated by

$$\hat{\delta}(m) = 1 + \sum_{j=1}^{k} \hat{a}_j f_j(m)\,. \quad (8)$$

We are not justified in extrapolating $\hat{\delta}(m)$ or any of its derived functions beyond the longest maturity of the securities observed. Notice that we are able to fit the discount function with a smooth curve, even though we do not have direct observations on it. We could never have done this by hand, or even by ordinary curve-fitting techniques.

I have actually fit (7) to observations on railroad bonds for fifteen selected dates from 1920 to 1938 and on U.S. Government securities for the close of every month from December 1946 to

[5] In the context of this application of linear regression, R^2 is a bad indicator of goodness of fit. It has no obvious intuitive interpretation and almost always is over .999. On the other hand, $\hat{\sigma}$ is meaningful and sensitive.

March 1966. The discount functions for two of these dates are shown in figures 1 and 2. They are displayed plus and minus their estimated standard errors of measurement. The estimated curve itself is not shown in order to avoid clutter. It lies halfway between the upper and lower edges of the band shown. Notice that the error, relative to the value of the curve, increases with time to maturity because the market is less concerned with the distant future than with the

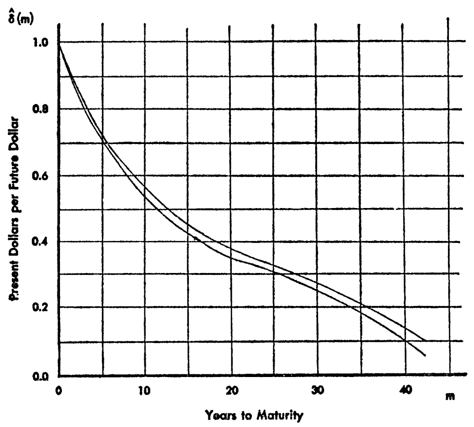

FIG. 1.—Discount function for the close of February 1922 based on bid-asked mean prices of high-grade (Moody's Aa and Aaa) railroad bonds. Convertibles and securities with any chance of being called before maturity were excluded. The band shows the best estimate plus and minus its standard error. In this regression, $n = 26$, $k = 5$, and $\hat{\sigma} = 2.67$.

near future, and therefore does not define the curve for the distant future with as great a precision. The calculation of these errors is discussed in a later section.

Mean values of $\hat{\sigma}$ for selected subperiods are given in table 1. These figures would seem to indicate that prior to the Treasury–Federal Reserve Accord of March 4, 1951, and again after the beginning of "Operation Twist" in mid-1961, some sort of "disarbitrageur" was active in the market for U.S. Government securities. The fall in $\hat{\sigma}$ from 13.9 at the close of February 1951 to 5.6 at the close of March 1951 was especially dramatic. Since most of the same securities were still present in the market, this fall could not have been entirely due to

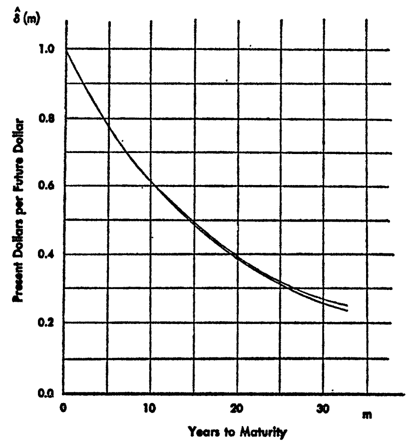

FIG. 2.—Discount function for the close of February 1966 based on bid-asked mean prices for taxable U.S. Government bills, notes, and bonds. Redemption of callable issues is assumed to be at earliest call date if price is above par and at maturity date when price is below par. In this regression, $n = 78$, $k = 9$, and $\hat{\sigma} = 7.81$. In spite of the higher value of $\hat{\sigma}$, this curve is better defined than that of fig. 1 because bid-asked spreads were smaller and because of the absence of brokerage fees.

TABLE 1

MEAN VALUES OF ERROR COEFFICIENT

Period	Type of Security	Mean Value of $\hat{\sigma}$
1920–1938.......	High-grade railroad bonds	2.6
1/1/47–3/1/51..	Taxable U.S. government securities	15.9
4/1/51–1/1/62..	Taxable U.S. government securities	4.6
2/1/62–4/1/66..	Taxable U.S. government securities	9.0

a change in the special features of the securities.

Having estimated the parameters a_j we can estimate the true values p_i of the n securities:

$$\hat{p}_i = 100 + c_i m_i + \sum_{j=1}^{k} \hat{a}_j \Big[100 f_j(m_i) + c_i \int_0^{m_i} f_j(m) dm \Big] . \tag{9}$$

This formula can even be used to estimate the value of securities that did not enter into the regression. It is of use to dealers, banks, insurance companies, and large borrowers who need to compare the values of securities differing in coupon rate and maturity. Sophisticated users may even want to adjust (1) for taxes and for the value of special provisions. Examination of the weighted residuals, $(\bar{p}_i - \hat{p}_i)/v_i$, shows that the ineligibility for commercial bank purchase of many bonds prior to the Accord (and to a lesser degree until 1954) tended to cause negative residuals and that the special tax status of deep discount bonds tended to cause positive residuals. These properties account in large measure for the disappointingly high values of $\hat{\sigma}$ for postwar Treasury securities. Compensating for such factors should give better fits and reduce the unaccounted-for error.

FORWARD INTEREST RATES

The discount function $\delta(m)$ is an exponential decay curve whose rate of decay need not be constant. Its rate of decay is the *instantaneous forward interest rate* $\rho(m)$:

$$\rho(m) = -\frac{\delta'(m)}{\delta(m)} . \tag{10}$$

Equivalently,

$$\delta(m) = \exp\Big[-\int_0^m \rho(x) dx \Big] , \tag{11}$$

and

$$\rho(m) = \lim_{h \to 0} \Big[\frac{\delta(m)/\delta(m+h) - 1}{h} \Big] . \tag{12}$$

By differentiating (3) we have

$$\delta'(m) = \sum_{j=1}^{k} a_j f_j'(m) . \tag{13}$$

Consequently we can estimate (10) with

$$\hat{\rho}(m) = \frac{-\Sigma \hat{a}_j f_j'(m)}{1 + \Sigma \hat{a}_j f_j(m)} . \tag{14}$$

Forward curves corresponding to the discount functions shown in figures 1 and 2 are depicted in figures 3 and 4, plus

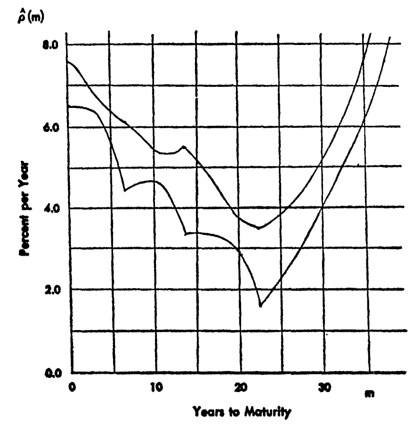

FIG. 3.—Instantaneous forward interest rate curve corresponding to the discount curve of fig. 1 for the close of February 1922.

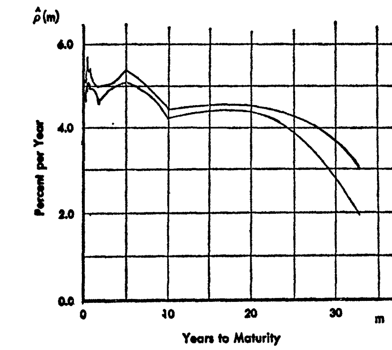

FIG. 4.—Instantaneous forward interest rate curve corresponding to the discount curve shown in fig. 2 for the close of February 1966. The high resolution at the short end is made possible by the concentration of bill observations.

and minus their standard errors of measurement. The calculation of these errors will be discussed in a later section. The "knuckles" in the bands are to be expected, unless we are willing to specify that $\delta(m)$ must be twice continuously differentiable instead of just once. As m goes to infinity, $\rho(m)$ does not necessarily approach an asymptote. Rather, its standard error of measurement will be found to grow without limit, so that its value simply fades away. This effect is more apparent in figure 4 than in figure 3, which goes off scale.

The instantaneous forward rate curve is a very important theoretical construct. However, its value for a single maturity m is of little practical concern, because it is prohibitively expensive in terms of transactions costs to make a forward contract between two points in the distant future if these points are only a small distance apart, as are m and $m + h$ in definition (12) of the instantaneous forward rate.

Only the average of $\rho(m)$ over a considerable interval in the future is of practical concern. Given any two values of m, say m_1 and m_2, the *mean forward interest rate* $r(m_1, m_2)$ is the average of $\rho(m)$ over the interval $[m_1, m_2]$:

$$r(m_1, m_2) = \frac{1}{m_2 - m_1} \int_{m_1}^{m_2} \rho(m)dm . \quad (15)$$

Equivalently,

$$r(m_1, m_2) = \frac{1}{m_2 - m_1} \ln \frac{\delta(m_1)}{\delta(m_2)} . \quad (16)$$

Computationally, (16) is more useful and can be estimated by

$$\hat{r}(m_1, m_2) = \frac{1}{m_2 - m_1} \ln \frac{\hat{\delta}(m_1)}{\hat{\delta}(m_2)} . \quad (17)$$

Notice that we have derived and estimated forward interest rates without use of the yield curve.

THE YIELD CURVE

The instantaneous forward interest rate curve $\rho(m)$ gives the rate of decay of the discount function $\delta(m)$ at each point m. The *yield curve* $\eta(m)$ is the average of that rate of decay over the interval from 0 to m. Thus,

$$\eta(m) = \frac{1}{m} \int_0^m \rho(x)dx . \quad (18)$$

Equivalent formulations are

$$\delta(m) = \exp[-m\eta(m)] , \quad (19)$$

$$\eta(m) = -\frac{1}{m} \ln \delta(m) , \quad (20)$$

and

$$\eta(m) = r(0, m) .$$

Equation (18) states that η stands in the relation of an average curve to the marginal curve ρ. Although η and ρ are not cost curves, they still bear the same mathematical interrelationships as do average and marginal cost curves:

i) $m\eta'(m) + \eta(m) = \rho(m) . \quad (21)$

ii) $\eta(0) = \rho(0) .$

iii) If η is $\left\{\begin{matrix}\text{rising} \\ \text{falling}\end{matrix}\right\}$ at m, then ρ is $\left\{\begin{matrix}\text{above} \\ \text{below}\end{matrix}\right\} \eta$ at m .

iv) If $\eta'(m) = 0$, then $\eta(m) = \rho(m)$.

The yield curve, as defined in equation (20), can be estimated by

$$\hat{\eta}(m) = -\frac{1}{m} \ln \hat{\delta}(m) . \quad (22)$$

Yield curves corresponding to figures 1 and 2 are given in figures 5A and 6A. The figures show the estimators plus and minus their standard errors of measurement. Notice how the standard error is large for both large and small m on the railroad curve for 1922 and is smaller for intermediate m. The computation of these errors will be discussed in a later section.

Other investigators have measured the term structure of interest rates by fitting

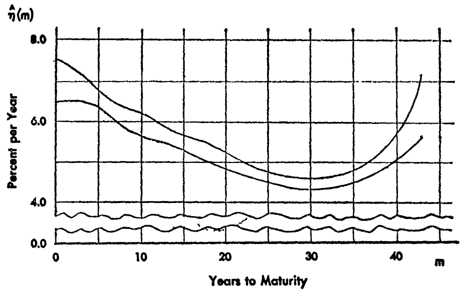

FIG. 5A.—Yield curve corresponding to the discount curve of fig. 1, for the close of February 1922.

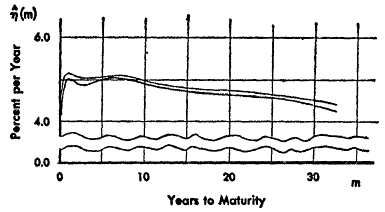

FIG. 6A.—Yield curve corresponding to the discount curve shown in fig. 2 for the close of February 1966.

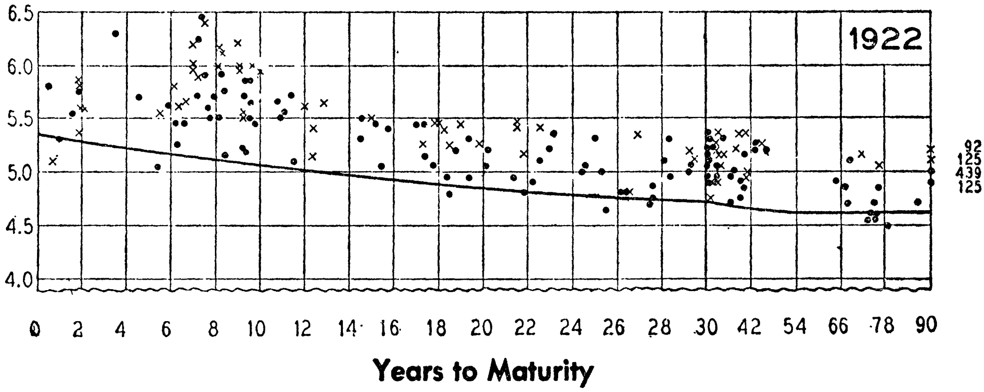

FIG. 5B.—A Durand yield curve for comparison with fig. 5A based on high-grade corporate bond transaction prices from the first quarter of 1922.

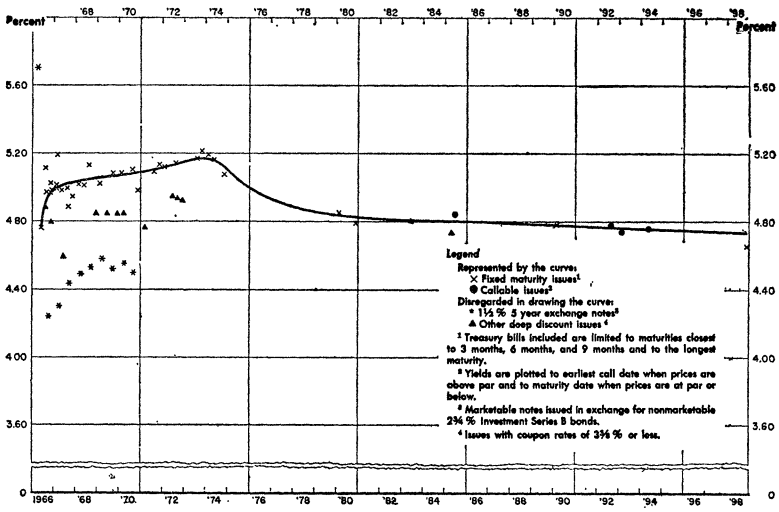

FIG. 6B.—A *Treasury Bulletin* yield curve for comparison with fig. 6A based on bid quotations for the close of February 1966. The smooth curve is fitted by eye. Market yields on coupon issues due in less than three months are excluded. Source: *Treasury Bulletin,* March 1966, p. 78.

a smooth curve to the average yields to maturity of the securities observed. Durand and the *Treasury Bulletin*[6] hand fitted the points with a French curve, while Cohen, Kramer, and Waugh[7] fit them with a linear regression. A Durand curve and a *Treasury Bulletin* curve are shown in figures 5B and 6B for comparison with figures 5A and 6A.

Both the hand and the regression approaches to directly fitting the yield curve are open to two serious objections. First, unless the yield curve $\eta(m)$ is flat, there is no reason to expect the average yield of a bond with a positive coupon rate to lie on it. The pure yield curve is defined for hypothetical bonds with zero coupon rates, so that the yield of an ordinary bond with maturity m_0 is a complicated average of $\eta(m)$ over the whole interval $[0, m_0]$, with only one of many weights at m_0, corresponding to the principal. For instance, if a bond has more than fifteen or twenty years to go before maturity, less than half of its value is due to the principal. The rest is embodied in the coupons. This averaging process washes out any shape the yield curve might have at the long end. The upward slope at the right end of the curve in figure 5A may be a pertinent example of the shape the yield curve may still have at the long end. Unfortunately, however, its measurement error becomes so large that this upward slope may or may not be statistically significant.

Second, any minor error incurred while directly fitting the yield curve will be magnified, especially for large m, if one tries to use formula (21) to calculate forward rates from the yield curve. Durand himself has insisted that his curves should not be used to derive forward rates.[8] For example, Durand's 1922 curve, shown in figure 5B, could not be used to infer the interestingly low forward rates shown in figure 3 which spanned the interval from 1937 to 1950 in the future (maturities fifteen to twenty-eight years). These remarkable rates foreshadowed the low rates for that period which were again to prevail during the later part of the Depression. When the data for 1922 were first fit, the author used a polynomial form for $\delta(m)$, and similar low forward rates resulted. In an effort to get rid of them, he devised the piecewise quadratic formulation which will be discussed in a later section, screened the data more carefully for bonds which were callable, convertible, or were not entirely risk free, and added more observations. In spite of these efforts the low forward rates persisted.

The tendency for the standard error of measurement of $\eta(m)$ to be large for small m means that Durand's "basic" yield curves are open to another objection: They are biased so that they tend to have an upward slope. In order to obtain rates for absolutely risk-free loans, he drew his yield curves to pass under the bulk of the plotted points. This would be a valid procedure if the width of the observed band of points were caused only by differences in the premiums for the risk of default. However, the observed bid-asked mean price of a virtually default-free security often differs from its predicted value by several times the transactions cost involved. In terms of yield, this difference is more important for short maturities than for long, causing Durand's plotted points to diverge for short maturities. By fitting his

[6] See nn. 1 and 2.

[7] See n. 4.

[8] David Durand, "A Quarterly Series of Corporate Basic Yields, 1952–57, and Some Attendant Reservations," *Journal of Finance* 13 (1958):348–56.

curves under the bulk of the points instead of through them, his curves tend to have an upward-sloping shape too often. Figures 5A and 5B illustrate one case when Durand appears to have obtained an insufficiently downward-sloping yield curve. The fact that Durand's curves are biased to slope upward could provide an alternative to the "liquidity preference" explanation of why the conglomerate yield curve derived by averaging his annual curves is upward sloping.[9]

ESTIMATION OF THE ERROR OF MEASUREMENT

The observed prices of securities are seldom even close to their true values as given by (1). In fact, the observed price itself is indeterminate, since the maximum price to a buyer (the asked price plus the brokerage fee, if any) is substantially higher than the minimum price to a seller (the bid price minus any brokerage fee). Consequently, the term structure cannot be measured exactly, and any estimator of a value derived from it is subject to random measurement errors. It is important to estimate the variance of these errors and to accompany any display of statistics derived from our regression with estimates of their standard errors. This has been done graphically in the diagrams accompanying this paper by displaying a band whose upper and lower edges are the actual estimate plus and minus its standard error.

The weighted least-squares regression on (7) produces a $k \times k$ matrix C which is the estimator of the covariance matrix of the estimators $\hat{a}_i$ of the parameters a_i. If z is a k-vector of known values and $\hat{a}$ is the vector $(\hat{a}_1, \ldots, \hat{a}_k)$, the estimator of the variance of the linear combination $z^T\hat{a}$, for example, will be the quadratic form z^TCz.

The estimator of the variance of $\hat{\delta}(m)$, defined by (8), is therefore

$$\widehat{\text{var}}\,[\hat{\delta}(m)] = z^TCz\,, \tag{23a}$$

where

$$z_j = f_j(m)\,. \tag{23b}$$

Similarly, the estimator of the variance of $\hat{p}_i$, as defined in (9), is

$$\widehat{\text{var}}\,(\hat{p}_i) = z^TCz\,, \tag{24a}$$

where

$$z_j = 100 f_j(m_i) + c_i \int_0^{m_i} f_j(m)dm\,. \tag{24b}$$

Just as (9) can help banks, dealers, and financial intermediaries estimate the proper bid-asked mean price to offer, (24) can help estimate the proper bid-asked spread with a little experience. For instance, they might make their offers differ from $\hat{p}_i$ by 1–3 S.E., depending upon how cautious they feel and the size of the offer.

The variance of the quotient of two random variables x and y with expected values Ex and Ey can be approximated by use of the formula

$$\frac{\text{var}\,(x/y)}{(Ex/Ey)^2} \approx \frac{\text{var}\,(x)}{(Ex)^2} + \frac{\text{var}\,(y)}{(Ey)^2} - 2\frac{\text{cov}\,(x, y)}{ExEy}\,, \tag{25}$$

provided that var $(x) \ll (Ex)^2$, var $(y) \ll (Ey)^2$, and that the distribution of y is positive.[10] Using (25), it can be shown

[9] See Reuben A. Kessel, *The Cyclical Behavior of the Term Structure of Interest Rates* (New York: National Bureau of Economic Research, 1965), p. 18.

[10] Formula (25) in the case of independently distributed variables and formula (27) are commonly used in experimental physics and chemistry. Both the variance of a quotient and that of a logarithm are related to the variance of a product, which is

that the variance of $\hat{p}(m)$, as given in (14), is approximated by

$$\widehat{\text{var}}\,[\hat{p}(m)] \approx \hat{p}(m)^2 z^T C z\,, \qquad (26a)$$

where

$$z_j = [f_j'(m)]/[\hat{\delta}'(m)] - [f_j(m)]/[\hat{\delta}(m)]\,. \qquad (26b)$$

The variance of the natural logarithm of a random variable x can be approximated by

$$\text{var}\,(\ln x) \approx \frac{\text{var}\,(x)}{(Ex)^2}\,, \qquad (27)$$

provided that var $(x) \ll (Ex)^2$, and that the distribution of x is positive. Using (27), it can be shown that the variance of the mean forward rate $\hat{r}(m_1, m_2)$, as defined in (17), is approximated by

$$\widehat{\text{var}}\,[\hat{r}(m_1, m_2)] \approx z^T C z\,, \qquad (28a)$$

where

$$z_j = \frac{1}{m_2 - m_1}\left[\frac{f_j(m_1)}{\hat{\delta}(m_1)} - \frac{f_j(m_2)}{\hat{\delta}(m_2)}\right]. \qquad (28b)$$

Again using (27), we see at once that the variance of the yield curve $\hat{\eta}(m)$, as given in (22), is approximated by

$$\widehat{\text{var}}\,[\hat{\eta}(m)] \approx \frac{\widehat{\text{var}}\,[\hat{\delta}(m)]}{[m\hat{\delta}(m)]^2}\,. \qquad (29)$$

THE FORM OF THE FUNCTIONS $f_j(m)$

The choice of the functions $f_j(m)$ is central to the quality of our fit of the term structure. However, the selection of a form will always be a matter of judgment. Only a few hard and fast rules hold. Two of these are that the $f_j(m)$ must be continuously differentiable and that $f_j(0)$ must be 0.

The maturities of the securities we observe will not be uniformly distributed over the interval from 0 to m_n, the longest maturity observed, except by accident. Where concentrations of observations occur, the shape of the discount function is relatively well defined. Where observations are sparse, we are not justified in distinguishing as much shape. Therefore it will be desirable to make $f_j(m)$ depend on the distribution of the m_i in such a way as to provide greater resolution wherever maturities are clustered. In the case of U.S. Treasury securities, following this rule will place the greatest resolution at the short end, where there are many bills outstanding. This is as it should be, since participants in the market are more concerned with small differences in time in the near future than in the far future. This greater concern means that the discount function $\delta(m)$ they define by the values they place on the outstanding securities will have the most detailed shape at the short end.

A relatively naïve approach is simply to set

$$f_j(m) = m^j\,, \quad j = 1, 2, \ldots, k. \qquad (30)$$

This assumption makes $\delta(m)$ a kth-degree polynomial with unity for its constant term. A polynomial is straightforward, but it has no theoretical motivation. Its formulation does not depend on the distribution of the m_i, nor does it have a greater capability for providing resolution for values of m where the m_i are more likely to occur. As a result of its uniform resolving power, when it is used to fit a discount function which has a finely defined shape in the first 1 or 2 percent of its length and is relatively

rigorously derived by Leo A. Goodman, "On the Exact Variance of Products," *Journal of the American Statistical Association* (December 1960), pp. 708–13. An approximation for products analogous to (25) can be derived from his exact formula when the variances are relatively small.

smooth thereafter, it will either ignore the short end and conform only to the remaining 98 or 99 percent, or else, if there are so many bill observations that they take over the regression, it will conform only to the short end and ignore the long end. It would take an extremely high-order polynomial to fit both the long and short ends of such a curve. Even so, this high-order polynomial would probably take on extreme values between observations at the long end, and would not be monotonically decreasing, as the discount function must be. On the other hand, a functional form which inherently permits greater resolution in the vicinity of data concentrations would be consistent with such a curve throughout its length, would require the estimation of only a few unknown parameters, and would be monotonic. The one portion would not have to be sacrificed to suit the other.

A better functional form for $\delta(m)$ than a polynomial is a continuously differentiable, piecewise quadratic function. To define such a curve we must divide the interval $(0, m_n)$ into $k - 1$ subintervals (d_j, d_{j+1}). We will have d_1 equal to 0 and d_k equal to m_n. Our $\delta(m)$ will follow a different quadratic function of m over each of the subintervals. In order for $\delta(m)$ to be continuously differentiable, the quadratics defined over adjacent subintervals (d_{j-1}, d_j) and (d_j, d_{j+1}) must have a common slope, as well as a common value, at d_j. The greater the number of subintervals covering any part of the interval $(0, m_n)$, the greater will be the resolving power of the discount function in that part of the interval. Therefore, by defining the subintervals to contain approximately equal numbers of the terminal maturities m_i, we will get greater potential resolution where the data observations are most numerous. Each of the quadratic segments will have an approximately equal number of observations to conform to. We can define the subintervals in this way by setting $d_j = m_l + \theta(m_{l+1} - m_l)$, where l = greatest integer in $[(j-1)n]/(k-1)$, and $\theta = [(j-1)n]/(k-1) - l$. (We have assumed the securities to have been arranged in order by increasing terminal maturity, so that $m_j \leq m_{j+1}$.)

The set of functions of the form

$$\delta(m) = 1 + \sum_{j=1}^{k} a_j f_j(m)$$

will comprise the entire family of continuously differentiable functions which satisfy $\delta(0) = 1$ and which are piecewise quadratic over the subintervals defined above if we define the $f_j(m)$ as shown in figure 7. The first one, $f_1(m)$, starts with value zero and with a positive slope at $m = 0$, flattens until it has a zero slope at $m = d_2$, and remains constant thereafter, as in figure 7*A*. Intermediate ones, $f_j(m)$, where $j = 2, 3, \ldots, k-1$, are zero up until d_{j-1}. There the slope begins to increase from zero up to some positive value at d_j. Then the slope falls from its d_j value to zero at d_{j+1}. Its value is constant thereafter. Figure 7*B* shows the particular case of $f_2(m)$. The last one, $f_k(m)$, is defined the same as the intermediate ones, except that it is undefined after $m = m_n$, as shown in figure 7*C*. Algebraically, the $f_j(m)$ are defined as follows:

$$f_1(m) = \begin{cases} m - \dfrac{1}{2d_2} m^2, & 0 \leq m \leq d_2 \\ \frac{1}{2} d_2, & d_2 < m \leq m_n \end{cases} \tag{31a}$$

$$f_j(m) = \begin{cases} 0\,, & 0 < m < d_{j-1} \\ \dfrac{(m-d_{j-1})^2}{2(d_j - d_{j-1})}\,, & d_{j-1} < m \leq d_j \\ \tfrac{1}{2}(d_j - d_{j-1}) + (m - d_j) - \dfrac{(m-d_j)^2}{2(d_{j+1}-d_j)}\,, & \\ \qquad d_j < m \leq d_{j+1} & \\ \tfrac{1}{2}(d_{j+1} - d_{j-1})\,, & d_{j+1} < m \leq m_n \end{cases} \quad j = 2, \ldots, k-1 \tag{31b}$$

$$f_k(m) = \begin{cases} 0\,, & 0 \leq m \leq d_{k-1} \\ \dfrac{(m-d_{k-1})^2}{2(m_n - d_{k-1})}\,, & d_{k-1} < m \leq m_n\,. \end{cases} \tag{31c}$$

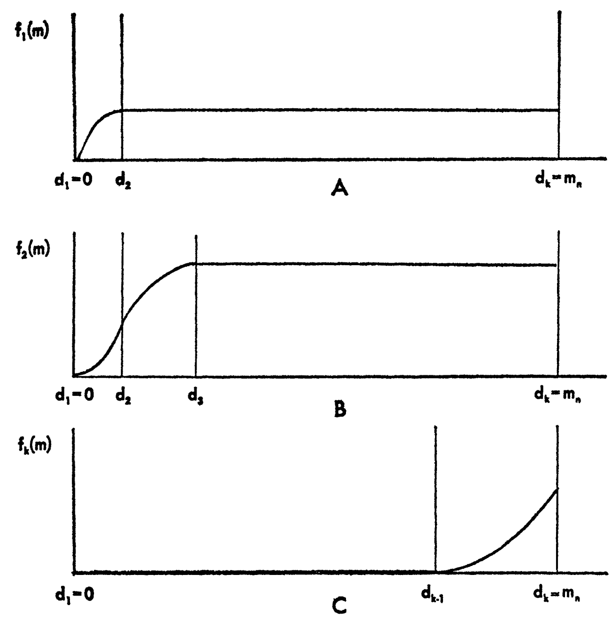

FIG. 7.—The preferred form of the $f_j(m)$. These $f_j(m)$ make (m) piecewise quadratic and continuously differentiable.

Since the vertical scales of the $f_j(m)$ are immaterial, we have arbitrarily chosen them so that

$$\begin{aligned} \delta'(d_j) &= a_j f_j'(d_j) \\ &= a_j\,. \end{aligned} \tag{32}$$

Integration of $f_j(m)$ in order to evaluate (7d) and (24b) and differentiation in order to evaluate (14) and (26b) are matters of elementary calculus, and will be omitted here.[11]

The specification of the $f_j(m)$ given in (31) was used for the regression fits of the

[11] Other specifications of the $f_j(m)$ will generate exactly the same family of piecewise quadratic functions. The one chosen was selected only because it can have the property (32) if the scales are chosen appropriately. The other specifications will give the same $\delta(m)$ if used with the same data.

discount function shown in figures 1 and 2. Because a piecewise quadratic function has a discontinuous second derivative, the instantaneous forward interest rate curves derived from these discount functions have discontinuous first derivatives, which explains the angular shape of the bands shown in figures 3 and 4. However, as mentioned earlier, the instantaneous forward rate $\rho(m)$ is interesting mainly as a theoretical construct. Its level at one isolated value of m has little practical significance. Consequently we are not worried by the outlying values of $\hat{\rho}(m)$ and its standard error that are sometimes implied by our specification of the $f_j(m)$. In fact, our specification is sufficient to imply that mean forward rates $r(m_1, m_2)$, which are of practical concern, are continuously differentiable with respect to both m_1 and m_2. The yield curve, a special case of $r(m_1, m_2)$ with $m_1 = 0$, is therefore also continuously differentiable, as may be seen in figures 5A and 6A.

THE VALUE OF k

The number k of parameters to be estimated is another area where judgment must be used. If k is too low, we will not be able to fit the discount function closely when it takes on difficult shapes. If it is too high, the discount function may conform too closely to outliers instead of being smooth. If k is as high as n, there will be no way to estimate σ^2. In the spirit of least squares, we might try all values of k inside a range we regard as reasonable, and select that value which minimizes the unbiased estimator $\hat{\sigma}^2$ of σ^2:

$$\hat{\sigma}^2 = \frac{1}{n-k} \sum_{i=1}^{n} \left(\frac{\bar{p}_i - \hat{p}_i}{v_i} \right)^2 .$$

As k increases, the residuals generally decrease, but then so do the degrees of freedom. The result is that $\hat{\sigma}^2$ declines sharply as k increases from 2 to 3 or 4, but thereafter fluctuates irregularly with a small amplitude, and often with more than one local minimum. Sometimes it shows no sign of permanently rising, even after k becomes so large that the discount function adheres to outliers.

A second approach is simply to make k a fixed function of n. We would like this function to have the following properties: First, in order to have resolution increase as the number of observations increases, our function $k(n)$ should increase with n. Second, in order to make the number of observations in the domain of each quadratic segment increase with the total number of observations, the ratio $n/k(n)$ should also increase with n. An elementary function with these properties is $k(n)$ = nearest integer to $n^{1/2}$. In practice, this formula gives approximately the same results as the first approach, without the expensive search.[12]

[12] Since the final revision of this paper, a precedent for the continuously differentiable, piecewise quadratic functional form has come to my attention (see Wayne A. Fuller, "Grafted Polynomials as Approximating Functions," *Australian Journal of Agricultural Economics* 13, no. 1 [June 1969]:35–46).