Volume Title: Economic Forecasts and Expectations: Analysis of Forecasting Behavior and Performance

Volume Author/Editor: Jacob A. Mincer, editor

Volume Publisher: NBER

Volume ISBN: 0-870-14202-X

Volume URL: http://www.nber.org/books/minc69-1

Publication Date: 1969

Chapter Title: Models of Adaptive Forecasting

Chapter Author: Jacob A. Mincer

Chapter URL: http://www.nber.org/chapters/c1216

Chapter pages in book: (p. 83 - 111)

THREE

Models of Adaptive Forecasting

JACOB MINCER

I. INTRODUCTION

Economic behavior is frequently a response to an anticipated future rather than to the past or present. Consequently, anticipated rather than actual values must be assigned to some of the variables in empirical economic models. Unfortunately, data on anticipatory magnitudes are scarce. Even when such data are obtainable, as in some surveys, they pose questions of reliability going beyond matters of sampling or measurement error. Reliable anticipatory values are those on which economic agents are, indeed, acting. *Ex ante* reports of such values are necessarily imperfect.

In the absence of reliable data or, more commonly, of any anticipations data, the economic analyst is forced to ascribe certain methods of formation of expectations to the subjects of his analysis. The issue cannot be ignored: The use of current rather than anticipated values is equivalent to a hypothesis that expectations are largely based on current magnitudes and do not differ from them in any systematic way.

A more sophisticated approach is to employ realized future values as proxies for anticipations of them. Such proxies are most appropriate when it is believed that economic agents forecast successfully, so that differences between anticipations and realizations are small.[1]

The usefulness of the implicit approach is obviously limited. Hence, the search for explicit models of expectations is a growing preoccupation of econometricians. In principle, the best model is one that most closely approximates the effective anticipatory values in economic behavior. However, to be useful the model must be relatively simple conceptually and statistically tractable.

In practice, most of the expectational models used in econometric analyses are extrapolations of current and past values of the time series, the future value of which is anticipated. In most cases the extrapolation function is linear: The forecast value of the variable Y is obtained by a weighted sum of past values of the series. Denoting forecast value by an asterisk, the date at which the forecast is formed (date of forecast base) by the right-hand subscript, and the date to which the forecast applies (date of forecast target) by the left-hand subscript, the extrapolated value for the next future period is:[2]

$$(1) \qquad {}_{t+1}Y_t^* = \sum_{j=0}^{\infty} \beta_j Y_{t-j}$$

The purpose of this paper is to inquire into the applicability and properties of several classes of forecasting models within the general class of linear extrapolations (1). The analysis suggests considerations

[1] This approach, known as the "implicit expectations" model was suggested by Mills [1]. If implicit forecast errors are not very small, the approach leads to unbiased estimation of parameters only when forecast errors are uncorrelated with realizations. It is worth noting, however, that the zero correlation assumption suggests *inefficient* forecasting:

If P = forecast, A = realization, and u = forecast error, zero correlation between u and A implies a nonzero correlation between u and P. Assuming $P = A + u$ is unbiased, a regression of A on P yields $A = \alpha + \beta P + v$, with $\beta < 1$. A corrected forecast $P' = \alpha + \beta P$ is clearly more efficient than P, as $\sigma^2(v) < \sigma^2(u)$. Proof: The coefficient of determination between A and P equals $1 - \frac{\sigma^2(u)}{\sigma^2(P)} = 1 - \frac{\sigma^2(v)}{\sigma^2(A)}$, and $\sigma^2(A) < \sigma^2(P)$.

[2] Functions of form (1) arise also in nonforecasting contexts. Thus, an observed value Z_t may constitute a response to the current and past values of the variable Y_{t-j}. While such distributed lag functions are clearly distinguishable in concept from extrapolation functions dealt with here, econometric practice has seldom differentiated among them. For a comprehensive survey of conceptual and practical issues in the application of both functions in econometric analyses, see Griliches [2].

which, to some extent, may guide the specification of such models. These considerations may help in determining when and where simplicity should be traded for greater realism. On the positive side, the analysis permits discrimination among several models of expectational behavior, when actual, albeit imperfect, forecasting data are available. Empirical illustrations are shown in the conclusion of the paper.

II. EXPONENTIAL COEFFICIENTS AND LINEAR EXTRAPOLATION

The specification of expectational models (1) is a specification of the coefficients β_j. The usual though not necessary restrictions are for the β_j to be nonnegative, less than unitary, and—particularly in the forecasting context—declining into the past. If Y is a trendless series, and (1) contains no intercept, the coefficients must sum to unity[3] to produce an unbiased forecast Y^*.

A widely used extrapolation function of form (1), which obeys the restrictions listed above, is the "geometrically declining weights," or exponential forecasting function:

$$ {}_{t+1}Y_t^* = \beta \sum_{j=0}^{\infty} (1-\beta)^j Y_{t-j}. \tag{1a} $$

Several reasons account for the popularity of forecasting formula (1a):

1. Only one coefficient β need be supplied or estimated.
2. The formula is theoretically appealing because it can be derived from a simple and rather plausible model of expectations adapting to unforeseen developments:

$$ {}_{t+1}Y_t^* - {}_tY_{t-1}^* = \beta(Y_t - {}_tY_{t-1}^*). \tag{2} $$

According to (2) expectations are formed by error learning; they are revised in consequence of (and in proportion to) currently experienced surprises. By successive substitution for the lagged term in (2), and

[3] If the series is trending with a rate of growth g, the coefficients must sum to $(1 + g)$.

by a reduction procedure on (1), due to Koyck [3], the two formulations can be shown to be equivalent.

3. More recently, strong support for exponential forecasting (1a) has been adduced from the statistical theory of optimal prediction.[4] It has been shown that such forecasts are optimal linear predictions, in the sense of minimizing the mean square error of forecast, for certain types of nonstationary time series.[5]

The knowledge that exponential extrapolation (1a) expresses a type of adaptive, error-learning behavior, and that it is an optimal forecasting scheme under certain conditions, lends a degree of confidence to this specification, which is often sound. It is not, of course, a claim on generality. As yet, the limitations on the appropriateness of exponential forecasting have not been fully explored.

One way to proceed is to consider the appropriateness of alternative formulations. A start can be made by inquiring into the applicability and interpretation of forecasting behavior (1) when the β_j coefficients *do not decline exponentially*. Can it still serve as an optimal predictor? Under what conditions?

The simplest class of time series for which the answer is readily available has been discussed by Muth [4].

Let the time series Y_t, which is being forecast, originate as a linear function of independent random shocks:

$$(3) \qquad Y_t = \epsilon_t + \sum_{i=1}^{\infty} w_i \epsilon_{t-i},$$

with zero mean and common variance for all ϵ. Exponential forecasting (1a) is optimal for such time series, if and only if all w_i are the same and equal β. And, whether the w_i are equal or not, (3) is a sufficient condition[6] for the optimal forecasting function to be linear in the past values of the series Y_t, as is (1). Here:

$$(4) \qquad {}_tY^*_{t-1} = \sum_{j=1}^{\infty} \beta_j Y_{t-j}.$$

[4] In particular, see the works of Muth [4], Nerlove and Wage [5], and Whittle [6].

[5] A series is stationary if its auto-covariance matrix is independent of calendar time.

[6] The linear extrapolation (1) is optimal for a wider class of processes than (3), as was indicated in note 4 for the exponential case.

The condition of optimality is that the β_j of (4) and the w_i of (3) are related by [7]:

$$w_i = \sum_{j=1}^{i} \beta_j w_{i-j}, \quad \text{with} \quad w_0 = 1. \tag{5}$$

III. THE GENERAL LINEAR ADAPTATION HYPOTHESIS

As assumption (3) illustrates, linear extrapolation (1) can be an optimal predictor without having exponentially declining weights. But does it, in that case, cease to represent adaptive, error-learning behavior? The answer is no. Indeed, all linear forecasts (1) which are optimal under (3) imply adaptive behavior as optimal forecasting for any number of spans. However, the revision function (2) describes adaptive behavior only in the exponential case. In the general case it is replaced by a set of revision functions, closely resembling (2), but not identical with it.

Consider optimal predictions for (3) for more than one period ahead. Start with forecasts made at $(t - 2)$ for target date (t): The smallest error such a forecast can have is $(\epsilon_t + w_1\epsilon_{t-1})$, as is apparent from assumption (3), because ϵ_t and ϵ_{t-1} are not known yet at time $(t - 2)$. Generally, the minimal forecast error for a k-span forecast is:

$$Y_t - {}_tY^*_{t-k} = \epsilon_t + \sum_{i=1}^{k} w_i\epsilon_{t-i}. \tag{6}$$

Now, we can show that a forecast at $(t - k)$ for k spans ahead will achieve this minimal error by substitution for the as yet unknown values in the extrapolation function (4) by their optimal forecast values:

[7] Substitute (3) into (4) to obtain:

$${}_tY^*_{t-1} = \sum_{i=1}^{\infty} M_i\epsilon_{t-i}, \quad \text{where} \quad M_i = \sum_{j=1}^{i} \beta_j w_{i-j}, \ (w_0 = 1). \tag{4a}$$

Since ϵ_t is not known at $(t - 1)$ when the forecast is made, the optimal forecast ${}_tY^*_{t-1} = Y_t - \epsilon_t = \sum_{i=1}^{\infty} w_i\epsilon_{t-i}$. Hence $M_i \equiv w_i$.

(7)
$$_{t}Y^*_{t-k} = \beta_1({}_{t-1}Y^*_{t-k}) + \beta_2({}_{t-2}Y^*_{t-k}) + \cdots + \beta_{k-1}({}_{t-k+1}Y^*_{t-k}) + \beta_k Y_{t-k} + \beta_{k+1}Y_{t-k-1} + \cdots .$$

A succession of substitutions, such as

$$_{t-k+1}Y^*_{t-k} = Y_{t-k+1} - \epsilon_{t-k+1}; \quad {}_{t-k+2}Y^*_{t-k} = Y_{t-k+2} - (\epsilon_{t-k+2} + \beta_1\epsilon_{t-k+1});$$

and so on, leads to the conclusion that optimality condition (6) is fulfilled.[8]

Compare now optimal forecasts made at different dates in the past for the same future target:

(7a)
$$_{t+1}Y^*_t = \beta_1 Y_t + \beta_2 Y_{t-1} + \cdots ,$$

and

(7b)
$$_{t+1}Y^*_{t-1} = \beta_1({}_tY^*_{t-1}) + \beta_2 Y_{t-1} + \cdots .$$

Subtracting (7b) from (7a) we obtain:

(8)
$$_{t+1}Y^*_t - {}_{t+1}Y^*_{t-1} = \beta_1(Y_t - {}_tY^*_{t-1}).$$

Equation (8), which we shall call a one-span revision function, can be viewed as a description of adaptive expectational behavior. It is, indeed, very similar to adaptive equation (2). But note the generality of (8): *It is implied by any optimal linear extrapolation* (4) under model (3), while (2) is implied only by the exponential extrapolation. Note also the difference: The forecast target in (2) shifts forward just as the forecast base does. In (8) the forecast target is fixed, and the forecast is revised only because the forecast base has moved forward.

In the exponential case, both (2) and (8) must hold, with $\beta_1 = \beta$. It follows that $_{t+1}Y^*_{t-1} = {}_tY^*_{t-1}$. More generally, exponential forecasting implies that the same expectations are held for any span in the future.[9]

Since, in the general case, adaptive hypotheses (2) and (8) are distinct, they can be applied to distinguish empirically between exponential and nonexponential expectational behavior.[10]

[8] In the two-span case: $_tY^*_{t-2} = \beta_1({}_{t-1}Y^*_{t-2}) + \beta_2 Y_{t-2} + \beta_3 Y_{t-3} + \cdots$. Since $_{t-1}Y^*_{t-2} = Y_{t-1} - \epsilon_{t-1}$, and $_tY^*_{t-1} = Y_t - \epsilon_t$, it follows that $_tY^*_{t-2} = {}_tY^*_{t-1} - \beta_1\epsilon_{t-1} = Y_t - (\epsilon_t + \beta_1\epsilon_{t-1})$. Condition (6) is fulfilled, as $\beta_1 = w_1$ by (5). Generalization to k spans is straightforward.

[9] This result is well known. For example, see Muth [4].

[10] See Section V, below.

Revision functions for k-span forecasts can be derived by the same procedure as that which produced (8). Subtracting

$$_{t+k}Y^*_{t-1} = \beta_1({}_{t+k-1}Y^*_{t-1}) + \beta_2({}_{t+k-2}Y^*_{t-1}) + \cdots + \beta_k(Y^*_{t-1}) + \cdots$$

from

$$_{t+k}Y^*_t = \beta_1({}_{t+k-1}Y^*_t) + \beta_2({}_{t+k-2}Y^*_t) + \cdots + \beta_k Y_t + \cdots$$

yields

$$(9) \quad {}_{t+k}Y^*_t - {}_{t+k}Y^*_{t-1} = \beta_1({}_{t+k-1}Y^*_t - {}_{t+k-1}Y^*_{t-1}) + \beta_2({}_{t+k-2}Y^*_t - {}_{t+k-2}Y^*_{t-1}) + \cdots + \beta_k(Y_t - {}_tY^*_{t-1}).$$

Since (9) is recursive, repeated substitutions produce:

$$(10) \qquad {}_{t+k}Y^*_t - {}_{t+k}Y^*_{t-1} = \gamma_k(Y_t - {}_tY^*_{t-1}).$$

The revision coefficient γ_k is a function of the weights $\beta_1, \beta_2, \ldots, \beta_k$ in the linear autoregressive forecasting function (4). In particular, we have seen in (8) that for $k = 1$, $\gamma_1 = \beta_1$. In general, the expression for the kth span revision coefficient is: [11]

$$(11) \qquad \gamma_k = \sum_{j=1}^{k} \beta_j \gamma_{k-j}, \quad (\gamma_0 = 1).$$

Now compare (11) with (5). Since $\gamma_1 = \beta_1 = w_1$, it follows that:

$$(12) \qquad \gamma_i \equiv w_i, \quad \text{for all } i = k.$$

We have reached the following conclusions:

1. The general autoregressive extrapolation (4) is consistent with an error-learning model of type (10) which uses different revision coefficients for different spans in the future.

2. Given observed revision coefficients γ_i, we can reconstruct the autoregressive extrapolation which generates the adaptive behavior by means of identities (5).

3. In the exponential case there is only one adaptation function, since all revision coefficients are the same. Forecasts and revisions for all future spans are the same. Adaptation function (2) is equivalent to (8).

4. If the time series can be described by the linear process (3) and the extrapolation is optimal for it, the revision coefficients γ_i in equa-

[11] This result is obtained by straightforward, though laborious, substitutions in (9).

tions (10) must be equal to the coefficients w_i in (3). Thus, there are as many distinct adaptive equations (10) as there are distinct parameters w_i in the linear process (3).

If extrapolations are available for more than one span, it is possible to test whether the extrapolations are optimal by comparing the mean square errors of the forecasts for each of the available spans with the revision functions corresponding to these spans.

The mean square error of an optimal forecast of Y_t in (3) for k periods ahead is:

$$\text{(13)} \qquad \begin{aligned} M_k &= \sigma^2(\epsilon_t + w_1\epsilon_{t-1} + \cdots + w_{k-1}\epsilon_{t-k+1}) \\ &= (1 + w_1^2 + w_2^2 + \cdots + w_{k-i}^2) \cdot \sigma^2(\epsilon), \end{aligned}$$

where

$$\sigma^2(\epsilon) = M_1,$$

from which

$$\text{(14)} \qquad w_i^2 = \frac{M_{i+1} - M_i}{M_1} = \gamma_i^2.$$

If M_i and γ_i are observable, a test of the right-hand equality in (14) is a test of optimality of the observed forecasts.[12]

IV. NONEXPONENTIAL FORECASTING AND STABILITY OF EXPECTATIONS

The pattern of coefficients γ_i in the adaptive equations (10) describes the pattern of revisions of future forecasts in response to current surprises. Implicit in these patterns are notions about stability of expectations. Thus, γ_i coefficients declining with span imply greater stability of long-term than of short-term expectations. That is, longer-term expectations remain relatively unaffected by unforeseen current developments. Conversely, γ_i coefficients increasing with span imply a greater sensitivity to such developments on the part of long-run

[12] Provided the forecasts are extrapolations, and the structure of Y is given by (3). Actual forecasts are seldom mere extrapolations (see the discussion in Section IX of this paper).

rather than short-run expectations. It would seem, perhaps, that in the trendless case, or in the case where expectations concerning trends do not change, the pattern of relatively greater stability of long-term expectations is more plausible.[13] However, the issue need not be decided a priori. Observed revision equations (10) can, in principle, provide insights into actual behavior.

The variation of γ_i with span i need not be monotonic. The implications about comparative stability of short- versus long-term expectations does, however, suggest a special interest in the monotonic cases over the relevant time span. Let us call *convex* forecasting behavior that which manifests itself in declining with span revision coefficients γ_i, and *concave* (or "explosive") that which shows the opposite pattern of coefficients in the revision equations (10).

Suppose we require an expectational model in which long-term expectations are comparatively more stable. This requirement rules out concave and exponential forecasting, that is, all linear extrapolation functions (4) which yield revision equations with fixed or increasing revision coefficients. What pattern must be imposed on the coefficients β_j of the extrapolation (4) to yield convex forecasts? Since the β_j coefficients must decline geometrically in order to yield fixed revision coefficients γ_i, must they, in some sense, decline more than exponentially, in order to generate declining γ_i? The answer is yes, if we adhere to the restrictions on β_j: that they must be positive, less than unitary, and strictly declining.

Convexity can be produced without these restrictions. That is to say, declining revision coefficients γ_i can be achieved with linear extrapolation functions (4) in which some of the β_j can be negative, exceed unity, and oscillate with j. Since the restrictions are often encountered in empirical work, and the pattern of β_j is easier to identify under such restrictions, we shall employ them for illustrative purposes.

It will be useful first to compare the coefficients β_j of single-span forecasting functions (4) with the corresponding coefficients of k-span forecasting functions (7), when the latter is reduced (by successive substitutions) to a function of past observed values only. Call the coefficients $\beta_j^{(k)}$ (in particular $\beta_j^{(1)} = \beta_j$), and consider $k = 2$. Then, by (7):

[13] The opposite pattern is certainly plausible when current developments lead to changed beliefs about future trends.

$$_{t+2}Y_t^* = \beta_1({}_{t+1}Y_t^*) + \beta_2 Y_t + \beta_3 Y_{t-1} + \cdots$$

which reduces to:

(15) $$_{t+2}Y_t^* = (\beta_1^2 + \beta_2)Y_t + (\beta_1\beta_2 + \beta_3)Y_{t-1} + (\beta_1\beta_3 + \beta_4)Y_{t-2} + \cdots.$$

Here, then:

(15a) $$\beta_1^{(2)} = \beta_1^2 + \beta_2, \quad \text{and} \quad \beta_j^{(2)} = \beta_1\beta_j + \beta_{j+1}.$$

Similarly, for $k = 3$:

$$\beta_1^{(3)} = \beta_1\beta_1^{(2)} + \beta_2^{(2)} = \beta_1^3 + 2\beta_1\beta_2 + \beta_3,$$

and

$$\beta_j^{(3)} = \beta_1\beta_j^{(2)} + \beta_2\beta_j + \beta_{j+2}.$$

And, more generally, for $k = i$:

(16) $$\beta_j^{(i)} = \beta_1\beta_j^{(i-1)} + \beta_2\beta_j^{(i-2)} + \cdots + \beta_{i-1}\beta_j + \beta_{i+j-1}.$$

Inspection reveals that, for $j = 1$, expression (16) is exactly the same as (5) and (11).

Hence:

(17) $$\beta_1^{(i)} \equiv \gamma_i \equiv w_i.$$

In words: *The coefficient attached to the forecast base value in the* ith *span forecasting function is equal to the* ith *span revision coefficient.* Note that this result follows without any restriction on β_j. Note also: [14]

(18) $$\sum_{j=1}^{\infty} \beta_j = 1 \quad \text{implies} \quad \sum_{j=1}^{\infty} \beta_j^{(i)} = 1, \quad \text{for all } k = i.$$

Expressing the i-span forecast by:

(19) $$_{t+i}Y_t^* = \beta_1^{(i)}Y_t + \beta_2^{(i)}Y_{t-1} + \beta_3^{(i)}Y_{t-2} + \cdots,$$

we conclude that, since $\beta_1^{(i)} = \gamma_i = w_i$ and $\sum_j \beta_j^{(i)} = 1$, declining revision coefficients γ_i imply that the further we forecast into the future the lesser the absolute and the relative weight attached to the most re-

[14] For $i = 2$: $\sum_j \beta_j^{(2)} = \beta_1 \sum_j \beta_j + \sum_j \beta_{j+1} = \beta_1 + (1 - \beta_1) = 1$. The proof for any i follows by mathematical induction.

cent observations $\beta_1^{(i)}$ and the greater the weights attached to the more distant observations.

The opposite is true when γ_i (and w_i) increase with i. It is easily seen that in the exponential case the weights $\beta_j^{(i)}$ are the same for all spans.

This is another way of explaining why, in this case, forecast values for all future spans are identical.

To repeat: When the γ_i (and w_i) decline, the $\beta_j^{(i+1)}$ coefficients for the $(i + 1)$-span forecasting function are at first smaller but eventually larger than the corresponding $\beta_j^{(i)}$ coefficients for the i-span forecasting function. If we assume strictly declining β_j with j this property formally means that:

$$\frac{\beta_j^{(i+1)}}{\beta_j^{(i)}} < \frac{\beta_{j+1}^{(i+1)}}{\beta_{j+1}^{(i)}} \quad \text{for all } j. \tag{20}$$

Using (15a) to substitute in the numerator we get, for $i = 1$:

$$\frac{\beta_1\beta_j + \beta_{j+1}}{\beta_j^{(i)}} < \frac{\beta_1\beta_{j+1} + \beta_{j+2}}{\beta_{j+1}^{(i)}}.$$

Hence:

$$\frac{\beta_{j+1}}{\beta_j} < \frac{\beta_{j+2}}{\beta_{j+1}}. \tag{21}$$

The meaning of (21) is that the forecasting functions, ${}_{t+1}Y_t^*$, have coefficients β_j whose rate of decline diminishes as j increases. When the inequality sign in (21) is reversed, the $\gamma_i^{(i)}$ increase; when (21) is an equality, the γ_i coefficients are constant.

The latter case is, of course, *exponential:* The rate of decline of β_j with j is fixed. The extrapolation function is *convex* when the rate of decline of β_j diminishes with j, and *concave* in the opposite case.

Note that inequality (21) is a sufficient condition of convexity. It is not required by the latter, except under the restriction of strictly declining β_j in the extrapolation function.[15]

Finally, inequality (21) suggests another way of describing the implications for stability of expectations in each of the three classes:

[15] The terms convexity and concavity are derived from the time shape of log β_j, as illustrated in Figure 3-1. But they apply to all forecasts which generate declining (or increasing) revision coefficients.

When the forecast is formed at relatively high values of the time series Y_t, convex forecasting implies that ${}_{t+i+1}Y_t^* < {}_{t+i}Y_t^*$. The opposite is true when the forecast base value Y_t is relatively low. More generally, convex forecasting implies that forecast values for successive spans trace out a monotonic path (upward or downward) from the base toward positions of "normalcy." No such "return to normalcy" is implied by exponential or concave forecasts. Movements away from normalcy are implied in the latter, and a horizontal path in the former. This relation between the patterns of β_j and of multispan predictions is shown in Figure 3-1.

The relation shown in this figure is illustrated for extrapolation functions with strictly declining coefficients β_j. But it is more general: Return to normalcy is a tendency in all convex forecasts. And the movement away from normalcy is a phenomenon in all concave forecasting.

Return to normalcy is best defined, in terms of our discussion, as a tendency to observe a negative correlation between current levels of Y_t and the direction of the predicted future flow, e.g., $({}_{t+k}Y_t^* - {}_{t+1}Y_t^*)$:

$$ {}_{t+k}Y_t^* - {}_{t+1}Y_t^* = b(Y_t - \tilde{Y}) + v_t = c + bY_t + v_t, \tag{22} $$

where v_t is a residual and $\tilde{Y}$ is a "normal" level of Y, which changes slowly. It is here impounded in the constant c.

Since ${}_{t+k}Y_t^* = \sum_{j=1}^{\infty} \beta_j^{(k)} Y_{t-j}$, (22) becomes

$$ {}_{t+k}Y_t^* - {}_{t+1}Y_t^* = (\beta_1^{(k)} - \beta_1)Y_t + v_t. \tag{23} $$

Recall that $\beta_1^{(k)} = \gamma_k$, by (17). Hence

$$ {}_{t+k}Y_t^* - {}_{t+1}Y_t^* = (\gamma_k - \gamma_1)Y_t + v_t, \tag{24} $$

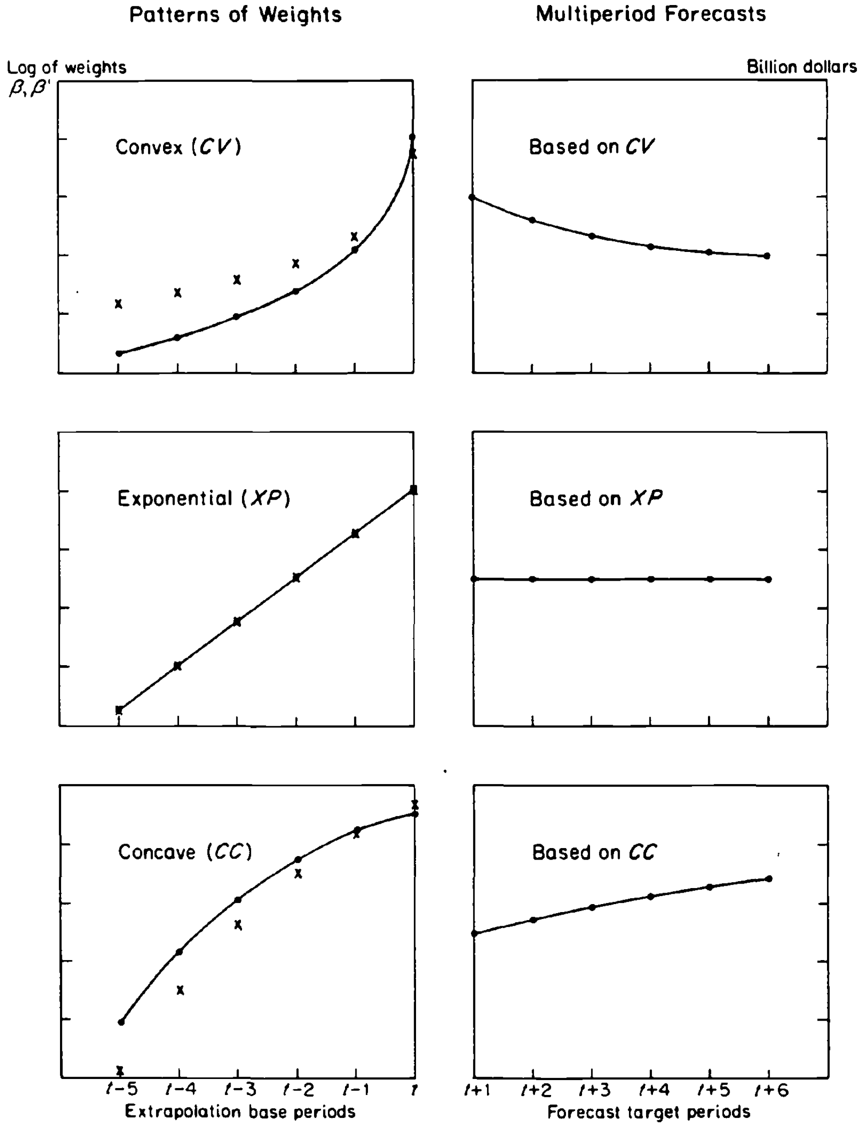
where

$$ v_t = \sum_{j=2}^{\infty} (\beta_j^{(k)} - \beta_j)Y_{t-j}. $$

Clearly, $b = (\gamma_k - \gamma_1)$ is negative in convex forecasts, positive in concave forecasts, and zero in exponential forecasts.[16]

[16] The sign of the correlation in (24) depends also on the sign, size, and correlation of the remainder term v_t with the independent variable Y_t. If the autocorrelation in Y is weak, the remainder term will have little effect. If the autocorrelation is substantial and positive, as it is more commonly, the relation can be expected to hold more dependably for low than for high values of k.

FIGURE 3-1. Exponential, Convex, and Concave Expectations Hypothetical Weight Patterns and Forecasts



Note: In the left panel, the solid line shows β, the weights in forecasting one period ahead; the crosses show β^1, the weights in forecasting two spans ahead.

The return to normalcy as an aspect of stability of expectations should be distinguished from the usual characterization of expectational stability by a less than unitary "elasticity coefficient of expectations" [β in adaptation function (2)]. Actually, so long as all β_j in (4) are less than 1 (a generalization of less than unitary elasticity), the forecast value for any span will be intermediate between high or low values of Y_t and normal levels. In this sense, all moving average extrapolations imply a return to normalcy. Convexity, however, adds a dynamic aspect to these stability characteristics: The path of expected movement persists in the direction of normal levels.[17]

A phenomenon closely related to the return to normalcy is known in the literature as regressivity in forecasting. This has been defined as a negative correlation between predicted change and past change in the time series.[18] If regressivity is basically a notion that future values of a series are expected to move in the direction of its mean, then a negative correlation in (22) is a better statement of this tendency.[19] If so, we may conclude that regressivity is an implication of convex forecasting.

V. STATIONARITY, OPTIMALITY, AND AGGREGATION

The time series Y_t arising from the linear process (3) is stationary, if $\sigma^2(Y) = \sigma^2(\epsilon) \sum_{i=1}^{\infty} w_i^2$ is finite. In that case, the w_i must converge to zero. Hence, the revision coefficients derived from an optimal ex-

[17] The need to introduce a return to normalcy feature into a basically exponential expectational model led Allais [7] to the addition of a separate term to the exponential extrapolation. Such a "splicing" is unnecessary in convex extrapolations.

[18] For references, see Bossons and Modigliani [8]. Bossons and Modigliani define regressivity specifically as a negative correlation between $({}_tY_{t-1}^* - Y_{t-1})$ and $(Y_{t-1} - Y_{t-k})$.

[19] Equation (22) avoids the overlapping term Y_{t-1} in the Bossons-Modigliani definition. This can produce the appearance of regressivity even when forecasting $({}_tY_{t-1}^*)$ is random. Equation (22) also generalizes predicted change to more than one span, and substitutes "deviation from normal" for "past change."

In proposing optimal predictions of future interest rates, Harberger [9] does, indeed, formulate regressivity by (22). His optimal forecasts are clearly convex: Compare his figure 1 [9, p. 137] with our Figure 3-1.

trapolation must eventually decline, even if they rise or fluctuate at first. Thus, optimal forecasts of stationary time series are, at least eventually, convex. However, convex forecasts may be optimal for structure (3) even when the latter is not stationary. This happens when the w_i coefficients decline, but do not converge.

It can be shown that convex forecasts are optimal for certain stationary time series whose stochastic process is somewhat more complex than (3):

Consider the following time series:

$$Y_t = X_t + U_t, \tag{25}$$

$$X_t = \sum_{i=1}^{n} \alpha_i X_{t-i} + \epsilon_t,$$

$$U_t = \sum_{h=1}^{m} \delta_h U_{t-h} + \eta_t.$$

ϵ_t and η_t are neither autocorrelated nor intercorrelated; Y_t and X_t are stationary.

For this latent structure of time series, a class of forecasts which attains a minimal mean square error is a weighted sum of several exponentially weighted averages of past values of Y_t: [20]

$${}_{t+1}Y_t^* = \sum_{j=0}^{\infty} \sum_{i=1}^{n} C_i \lambda_i (1 - \lambda_i)^j Y_{t-j}. \tag{26}$$

In terms of the general linear autoregression (4), the coefficients attached to past values Y_{t-j} are:

$$\beta_j = \sum_{i=1}^{n} C_i \lambda_i (1 - \lambda_i)^j. \tag{27}$$

The interesting thing about function (26) is that, provided the λ_i are distinct, positive, and less than unity, *it is necessarily convex.*

Recall the condition of convexity, when $0 < \beta_j < 1$:

$$\beta_j^2 < \beta_{j-1} \cdot \beta_{j+1}. \tag{21}$$

Applied to (27), the following inequality must hold as a condition of convexity:

[20] See Bailey [10] for derivation.

$$(28)\quad \Big[\sum_i C_i\lambda_i(1-\lambda_i)^j\Big]^2 < \Big[\sum_i C_i\lambda_i(1-\lambda_i)^{j-1}\Big]\Big[\sum_i C_i\lambda_i(1-\lambda_i)^{j+1}\Big].$$

Define
$$a_i = [C_i\lambda_i(1-\lambda_i)^{j-1}]^{1/2},$$
$$b_i = [C_i\lambda_i(1-\lambda_i)^{j+1}]^{1/2.}$$

By the Schwartz inequality:

$$\Big(\sum_i a_ib_i\Big)^2 < \sum_i a_i^2 \sum_i b_i^2.$$

Since, in our case, $b_i = (1-\lambda_i)a_i$, it is easily seen that the equality sign holds only when $(1-\lambda_i)$ is the same for all i (or when $n = 1$). Otherwise the convexity of the forecasting function (26) must hold.[21]

Function (26) may arise as an aggregation phenomenon rather than as an optimal extrapolation for an assumed type of time series. If individuals ($i = 1, \ldots n$) forecast exponentially, the aggregated (market?) forecast will appear to be convex, in terms of the β_j coefficients, if not all λ_i are identical;[22] similarly if a forecast of an aggregate, say GNP, was obtained by aggregating sectoral forecasts, each of which was exponential with different parameter λ_j.

Note, however, that even if the aggregated function (26) appears to have convex coefficients β_j, it does not imply declining revision coefficients γ_i: If each individual, or sector forecast is exponential, individual multispan forecasts are identical for each span. Therefore, the aggregated (weighted averages) multispan forecasts also remain fixed regardless of span, and the (aggregated) revision coefficients remain fixed. Thus, when reported on aggregates, the revision equations (10) provide better insight into the true nature of forecasting behavior than the extrapolation function itself.[23]

[21] Intuitively, the conclusion that a linear combination of exponentials is necessarily convex is perhaps best visualized as follows: Since exponentials are linear in logs, only geometric averages of exponentials are exponential. Arithmetic averages exceed geometric averages, hence (21) is nonlinear in logs. It is convex, because the arithmetic average is biased toward the higher and more steeply declining log $\lambda_i(1-\lambda_i)^j$ for small values of j, and again toward the higher and flatter log $\lambda_i(l-\lambda_i)^j$ for larger values of j.

[22] See Bierwag and Grove [11].

[23] If individual forecasting is exponential, the revision coefficient is a weighted average of the λ_i. The degree of convexity in the aggregated extrapolation function clearly depends on the variance of λ_i across individuals. Taken together, the extrapolation and revision functions provide information on the distribution of λ_i among individuals.

VI. FORECASTING STOCK OR FLOW VARIABLES

The discussion in the preceding sections suggests that, despite the widespread use and asserted success of the exponential forecasting formula, convexity may often be a better description of extrapolative expectations.

Successful exponential forecasting need not be inconsistent with this conclusion. First, the degree of convexity may not be sufficiently strong to affect the forecasting errors very much. More important, the relevant variables which need to be forecast for purposes of management decisions are often discounted flows of predicted future values rather than single-span predictions. The redeeming feature of the exponential forecast is that even when it is unsatisfactory as a forecast of a single future value, its error as a forecast of a discounted multispan flow is likely to be much smaller:

Denote the single span prediction of the flow at i by ${}_{t+i}P_t^*$, and the prediction of the stock (converted into the same dimension as a flow in perpetuity) by P_t^*. Then, *exponential* forecasting of ${}_{t+i}P_t^*$ means that $P_t^* = {}_{t+1}P_t^*$. This is because ${}_{t+i}P_t^* = {}_{t+1}P_t^*$ for all i.

$$(29) \qquad P_t^* = r\left[\frac{1}{1+r}\,{}_{t+1}P_t^* + \frac{1}{(1+r)^2}\,{}_{t+2}P_t^* + \cdots\right]$$

$$= {}_{t+1}P_t^* r \sum_{i=1}^{\infty} \frac{1}{(1+r)^i} = {}_{t+1}P_t^*.$$

When nonexponential forecasting is appropriate but exponential is used, the exponential forecasts of flows ${}_{t+i}P_t^*$ will be too high for near spans and too low for higher spans, or conversely. Since P_t^* is a weighted average[24] of ${}_{t+i}P_t^*$, the error of using it as a forecast of P_t will be smaller than the average error of ${}_{t+i}P_t^*$ as a forecast of P_{t+i}.

In the case of the consumption function, for example, an incorrect exponential formulation of "permanent" income may yield errors in predicting consumption which are only slightly larger than the errors resulting from a correct convex formulation. At the same time, the

[24] Weighted by $\frac{r}{(1+r)^i}$. The attentuation of error will be greater the lower the discount rate r.

differences in the errors of forecasting the next period's income by the two forecasting functions could be sizeable. Similarly, even a relatively poor exponential forecast of the next period's sales may be a relatively good forecast of the discounted flow of future sales. If the latter decision variable is superior to the former, the incorrectly formulated forecast may be sufficiently useful.

Equation (29) suggests two important properties of exponential forecasts of perpetuities (discounted stocks): First, they are the same as forecasts of single-period flows, and second, the forecast value does not depend on the discount rate. For example, if permanent income (Y_P) is estimated by exponential extrapolation of past incomes, its forecast is the same as the forecast of the next period's income, $Y_P = {}_{t+1}Y_t^*$. More important, the exponential forecast does not depend on the discount rate, and cannot, therefore, be used to estimate the discount rate: The exponentially declining weights with parameter β do not provide any information on the size of the discount rate r.[25]

When the relevant variable to be forecast is a discounted future flow, it is natural to raise the question about possible relations between the discount rate r and the expectational coefficients β_j. A small r denotes "longsightedness" into the future. Similarly, a small β_1 means that a longer past was taken account of in forming expectations. It is tempting to postulate a positive correlation between the two parameters of behavior: Horizons are both longer or shorter symmetrically with respect to the future and to the past. And, as discount rates change, so do expectational coefficients.[26] It is clear, however, that no such connection needs to exist if the extrapolative weights β_j are dictated exclusively by the structure of the time series, while the size of the discount rate does not depend on the experienced variations in the time series.

[25] See Friedman's analysis of the consumption function [12] and [13]. In the latter article Friedman proposes a different expectational interpretation of his estimating procedure in [12], precisely for the reasons indicated above.

[26] A hypothesis which has some resemblance to this one was introduced by Allais [7].

VII. EXTRAPOLATION, AUTONOMOUS FORECASTING, AND EMPIRICAL INVESTIGATION OF FORECASTING BEHAVIOR

It is not reasonable to assume that forecasts F_t of a future value of Y_t are based exclusively on extrapolations of that series. We may represent the actual forecast as consisting of two parts:

$$ {}_{t+1}F_t = {}_{t+1}Y_t^* + {}_{t+k}u_t, \tag{30} $$

where ${}_{t+1}Y_t^*$ is the extrapolative component and ${}_{t+k}u_t$ an independent remainder, or autonomous component of the forecast. The autonomous component presumably utilizes information based on relations with other series and other objective or subjective data.[27]

The preceding discussion of forecasting functions refers to the extrapolative component in (30). Before empirical observations can be analyzed, we need to know in what way the presence of autonomous components affects our conclusions. We proceed to an analysis of revision equations in the presence of autonomous components.

Expression (30) refers to a one-span forecast. If forecasts for longer spans are obtained recursively, that is, by substitution of intervening-span forecasts for as yet unknown values of Y_{t+i}, so that:

$$ \begin{aligned} {}_{t+k}Y_t = \beta_1({}_{t+k-1}F_t) + \beta_2({}_{t+k-2}F_t) + \cdots + \beta_k Y_t \\ + \beta_{k+1}Y_{t-1} + \cdots + {}_{t+k}u_t, \end{aligned} \tag{31} $$

then the general term of (30) for k spans is:

(32)

$$ {}_{t+k}F_t = {}_{t+k}Y_t^* + \sum_{j=0}^{k-1} \gamma_j({}_{t+k-j}u_t) \quad (\gamma_0 = 1, \quad \text{and} \quad \gamma_k = \sum_{j=i}^{k} \beta_j\gamma_{k-j}). $$

Revision functions (10) now become:

$$ \begin{aligned} {}_{t+k}F_t - {}_{t+k}F_{t-1} &= \gamma_k(Y_t - {}_tF_{t-1}) \\ &\quad + (\Delta_k u + \gamma_1\Delta_{k-1}u + \cdots + \gamma_{k-1}\Delta_1 u), \end{aligned} \tag{33} $$

where $\Delta_k u = {}_{t+k}u_t - {}_{t+k}u_{t-1}$.

[27] See Chapter 1 of this volume, pp. 23 ff.

It is possible to conceive of expectational behavior in which multi-span forecasting is recursive only in the extrapolative component. The autonomous component is then superimposed. In such a case expression (30) generalizes directly for any span k:

$$ {}_{t+k}F_t = {}_{t+k}Y_t^* + {}_{t+k}v_t. \tag{30a}$$

And the revision functions (33) become:

$$ {}_{t+k}F_t - {}_{t+k}F_{t-1} = \gamma_k(Y_t - {}_tF_{t-1}) + \Delta_k v + \gamma_k\, {}_tv_{t-1}, \tag{33a}$$

since

$$ {}_tF_{t-1} = {}_tY_{t-1}^* + {}_tv_{t-1}.$$

Even if expectational behavior is not described by (30a), empirical data which we take as representing F may contain some systematic nonforecasting components[28] or, more commonly, measurement errors. Revision equation (33a) would then be interpreted as reflecting extrapolation observed with some error.

If data on forecasts are available for several spans, empirical estimates of revision equations can be used to ascertain important features of expectational behavior:

1. Estimated coefficients γ_k indicate whether expectational behavior is exponential, convex, or concave.

2. The extrapolation function can be reconstructed from the γ_k coefficients.

3. The coefficient of determination in revision equations (33) is less than unity, because of the presence of nonextrapolative components in forecasts. Its size reflects the importance of revisions of autonomous components or of change in nonforecasting components in the observed revisions of F.

It is of interest to note that, under model (31), the residual variance in empirical regressions of the revision functions increases with span, as the right-hand term in (33) cumulates. This is not true in (33a), where the residual variance grows or declines together with the coefficients γ_k, thus *decreasing* in convex forecasting.

Denoting the forecast error $(Y_t - {}_tF_{t-1}) = \epsilon$, and assuming $\Delta_k u$ are uncorrelated over spans and of equal variance, we can derive coefficients of determination for the various spans of the revision functions (33):

[28] Such as the liquidity premium in forward interest rates. See Kessel [14].

(34) $\sigma^2(\Delta_k F) = \gamma_k^2\sigma^2(\epsilon) + (1 + \gamma_1^2 + \gamma_2^2 + \cdots + \gamma_{k-1}^2)\sigma^2(\Delta u).$

Hence, coefficients of determination in empirical revision equations for span k are given by

(35) $$\frac{R_k^2}{1 - R_k^2} = \frac{\gamma_k^2}{1 + \gamma_1^2 + \gamma_2^2 + \cdots + \gamma_{k-1}^2} \cdot \frac{\sigma^2(\epsilon)}{\sigma^2(\Delta u)}.$$

Putting $k = 1$ into (34) makes it possible to replace the unobservable term $\dfrac{\sigma^2(\epsilon)}{\sigma^2(\Delta u)}$ in (35) by the observable $\dfrac{1}{\gamma_1^2 \dfrac{1 - R_1^2}{R_1^2}}$, since

$$\frac{R_1^2}{1 - R_1^2} = \frac{\sigma^2(\epsilon)}{\gamma_1^2\sigma^2(\Delta u)},$$

which yields:

(36) $$\frac{R_i^2}{1 - R_i^2} = \frac{\gamma_i^2}{1 + \gamma_1^2 + \gamma_2^2 + \cdots + \gamma_{i-1}^2} \cdot \frac{1}{\gamma_1^2} \cdot \frac{R_1^2}{1 - R_1^2}.$$

Clearly, R_k^2 declines as k increases in convex and in exponential forecasts. It may increase, though it need not, in concave forecasts. If the revision functions are interpreted as (33a) rather than (33), the coefficients of determination follow:

(35a) $$\frac{R_k^2}{1 - R_k^2} = \frac{\gamma_k^2\sigma^2(\epsilon)}{\gamma_k^2\sigma^2(v) + \sigma^2(\Delta_k v)} = \frac{\sigma^2(\epsilon)}{\sigma^2(v) + \dfrac{1}{\sigma_k^2} \cdot \sigma^2(\Delta v)}.$$

Assuming that $\sigma_k^2(v)$ and $\sigma^2(\Delta_k v)$ do not vary systematically with k, it follows from (35a) that R^2 declines with increasing span in convex forecasts (even though the residual variance declines), increases in concave forecasts, and remains unchanged in exponential forecasting.

Taking models (33) and (33a) together—and they are not mutually exclusive if a nonforecasting component is present in the data on F—it appears that the coefficient of determination in the revision equations is most likely to decline in convex forecasts, remain constant in exponential forecasts, and increase in concave forecasting.

If empirical data on forecasts are available for two spans only, discrimination between exponential and nonexponential forecasting can still be achieved by a comparison of estimated revision functions (8) with (2):

(8) $$_{t+1}F_t - {}_{t+1}F_{t-1} = \beta_1(Y_t - {}_tF_{t-1}),$$

(2) $$_{t+1}F_t - {}_tF_{t-1} = \beta(Y_t - {}_tF_{t-1}).$$

If forecasting is exponential, the two equations should yield the same results. However, estimated β_1 should be greater than β if forecasting is convex, and smaller than β if concave.

To see this, perform a Koyck-type reduction in the general case:

$$_{t+1}F_t = \beta_1 Y_t + \beta_2 Y_{t-1} + \cdots$$

$$(1 - \beta_1)_tF_{t-1} = (1 - \beta_1)(\beta_1 Y_{t-1} + \cdots).$$

Subtracting, we obtain:

(2′) $$_{t+1}F_t - {}_tF_{t-1} = \beta_1(Y_t - {}_tF_{t-1}) + [\beta_2 - (1 - \beta_1)\beta_1]Y_{t-1} + [\beta_3 - (1 - \beta_2) \cdot \beta_2]Y_{t-2}.$$

In the exponential case, all terms beyond the first on the right-hand side of (2′) vanish, yielding adaptive equation (2). However, in the convex case, lagged terms of Y_t enter with negative coefficients [since $\beta_2 < \beta_1(1 - \beta_1)$], and in the concave case, with positive coefficients.

With positive serial correlation in Y usually present, leaving out the lagged terms [that is, using (2) instead of (2′)] will make the estimated β smaller than β_1 in the convex case, larger in the concave case.

If data are available for one span only, equation (2′) can, in principle, still serve the purpose: Convexity is suggested by significant lagged terms with negative coefficients, concavity by the same terms with positive coefficients. No lagged terms appear in the exponential case.

THE TERM STRUCTURE OF INTEREST RATES

Revision equations (10) were first introduced into the empirical literature by Meiselman [15] in his study of the term structure of interest rates. While previous research based on adaptive hypothesis (2) was invoked to justify the use of (10), the distinction between the two formulations of adaptive behavior received no attention in that study.[29]

Meiselman tested the hypothesis that the "forward" rate $_{t+k}F_t$ is

[29] The expectational aspects of the term structure are intensively explored within the present framework by Stanley Diller in Chapter 4 of this volume. In this section we draw on some of his findings.

a forecast of the future (spot) rate A_{t+k}, by means of empirically fitted revision functions (10), for $k = 1, 2, \ldots$, eight spans. The fact that good fits were obtained is consistent with a hypothesis that forward rates embody linear autoregressive forecasts of future spot rates. They also embody autonomous forecasting components, as well as nonforecasting components such as liquidity premia. The existence of the nonextrapolative component in the forward rate creates correlations that are less than one in Meiselman's revision function. This component is responsible for the weakening of the Meiselman correlations as the span increases. As Diller shows (Table 4-1), the pattern of decline in R^2 is closely predictable on the assumption of recursive multispan forecasting, as in our equation (36).

Meiselman's revision equations show continuously declining estimates of revision coefficients γ_i from .703 in the first span to .208 in the eighth span. This pattern is clearly consistent with convex forecasting.[30]

While the pattern of eight revision coefficients constitutes more comprehensive evidence of convex forecasting, it might be of interest to illustrate the discrimination between hypotheses of exponential and nonexponential forecasting, using only the first revision equation in a

[30] It is also interesting to note that the revision coefficients (γ_i) in Meiselman's Table 1 seem to decline almost geometrically. If the coefficients for the more remote spans are disregarded, the pattern can be approximated by a straight line in logs (as noted by Meiselman, p. 21).

(24a) $$\lg \gamma_i = \alpha + i \log \gamma$$

with α close to unity, so that:

(24b) $$\gamma_i = \gamma^i.$$

If such an approximation is imposed, it turns out that the linear autoregressive extrapolation which would give rise to such revision coefficients is of a very simple form:

(24c) $${}_{t+1}A_t^* = \beta_1 \cdot A_t.$$

And, in terms of forward rates:

(24d) $${}_{t+1}F_t = \beta_1 \cdot A_t + u_t.$$

Proof: Recall (11)

$$\gamma_i = \sum_{j=1}^{i} \beta_j \gamma_{i-j} \quad (\gamma_0 = 1).$$

Substituting (24b) into (11) yields $\beta_1 = \gamma_1$, all other $\beta_j = 0$. In a recent article, Pye [16] shows that Meiselman's revision coefficients could have been produced by a particular first order Markov chain. This is equivalent to result (24c) which is a first order autoregression. See also L. Telser [17].

comparison with adaptive equation (2). The results were:

$$(8) \qquad {}_{t+1}F_t - {}_{t+1}F_{t-1} = .703\ (A_t - {}_tF_{t-1}), \quad R^2 = .906,$$

$$(2) \qquad {}_{t+1}F_t - {}_tF_{t-1} = .558\ (A_t - {}_tF_{t-1}), \quad R^2 = .774.$$

As expected in convex forecasting, the regression coefficient in (8) exceeds the regression coefficient in (2). Equation (2) also shows a weaker fit and a larger residual variance, while the variance of the dependent variable is smaller than in (8).

Diller finds suggestions of convex forecasting also in other bodies of interest rate data. However, Conard [18, Table 10] reports a study of government securities in which neither revision coefficients nor (the very high) coefficients of determination change with span. If these data reflect forecasting behavior, then according to (35a) the findings suggest exponential extrapolation, without autonomous components.

BUSINESS FORECASTS

In a recent NBER study of short-term economic forecasting, Victor Zarnowitz [19] compiled and analyzed a variety of recent forecasts of aggregate economic activity in the United States. The forecasts come from a variety of sources.[31] Most of them are predictions of the next year's business, but some include forecasts of several semiannual or quarterly spans.

One of the conclusions of Zarnowitz's study is that these forecasts in part represent extrapolations of the past. In order to ascertain whether business forecasts are better characterized as exponential, concave, or convex, regressions were fit to the two alternative forms of revision functions (2) and (8).

Columns 1 to 4 in Table 3-1 show results of fitting the shifting-target function (2); columns 5 to 8 are results of fitting the fixed-target function (8).

Clear patterns of convexity are visible in GNP forecast G, for which five spans are available. Otherwise, the evidence is unclear. Since the business forecasts contain apparently sizeable autonomous components, the correlations are not strong.[32]

[31] For a detailed description, see [19, Chapter 1].

[32] Another reason is that the forecast base values contain errors of measurement. Forecasters use preliminary available data which are subject to revision. Data revisions are, in effect, a part of the forecasting error. For a discussion of this issue, see Rosanne Cole, Chapter 2 of this volume.

TABLE 3-1. Relations Between Forecast Revisions and Forecast Errors, Selected Forecasts of GNP and Plant and Equipment Outlays for Spans Varying From Three to Eighteen Months [a]

	Revision Function (2) Regressions of Constant-Span Revisions				Revision Function (8) Regressions of Reduced-Span Revisions			
Line	Span of Forecast, in Months (before and after revision) (1)	Intercept a_1 (2)	Coefficient of Regression b_1 (3)	Correlation Coefficient r_1 (4)	Span of First and Span of Revised Forecast (in months) (5)	Intercept a_2 (6)	Coefficient of Regression b_2 (7)	Correlation Coefficient r_2 (8)
				GNP Forecasts: C				
1	3	5.07 (2.25)	2.034 (.684)	.668	3:0	−1.57	.892	.428
2	6	6.68 (2.65)	1.391 (.803)	.463	6:3	0.54 (2.07)	.814 (.627)	.365
3	9	8.44 (2.35)	.253 (.714)	.106	9:6	0.72 (1.94)	.657 (.588)	.319
				GNP Forecasts: D				
4	6	12.87 (4.54)	−.050 (.359)	−.037	12:6	9.07 (2.86)	−.112 (.227)	−.131
				GNP Forecasts: G				
5	6	12.23 (2.40)	.265 (.256)	.250	6:0	−3.77 (2.00)	.565 (.214)	.551
6	9	13.23 (2.33)	−.063 (.249)	−.063	9:3	−2.97 (1.59)	.422 (.170)	.528
7	12	14.15 (2.45)	−.226 (.262)	−.211	12:6	−1.07 (1.94)	.289 (.207)	.329
8	15	14.51 (3.04)	−.372 (.334)	−.306	15:9	−1.17 (2.49)	.070 (.274)	.074
9	18	14.81 (2.99)	−.452 (.330)	−.368	18:15	−.13 (2.56)	−.041 (.281)	−.042
				Anticipations of Plant and Equipment Outlays (OBE-SEC)				
10	3	.66 (.16)	1.033 (.228)	.548	6:3	.13 (.09)	.675 (.127)	.609

[a] Period covered: 1952-II through 1964-III. The figures in parentheses are standard errors.

VIII. OPTIMALITY, ONCE AGAIN

In the early sections of this paper the formal generation of optimal forecasts was exclusively determined by assumptions about the stochastic structure of time series. This was a matter of mathematical and expositional simplicity. In general, an optimal formulation of forecasts depends not only on the stochastic structure of time series but also on the criterion of optimization, that is, on the "loss function" of the forecast error. Minimization of an economically motivated loss function need not yield the same results as, for example, the mean square error criterion.

In particular, minimizing the cost of error may lead to convex forecasting even when the mean square error criterion implies concavity, or conversely.[33] We have seen that convex forecasting means larger revisions of short-term than of long-term expectations, in response to current (unexpected) developments. If short-term plans are based on short-term expectations, and if economic considerations lead to greater flexibility in the short run than in the long run, such considerations may lead to convex forecasting. For example, if revisions of (short-run) production schedules are less costly than those of (longer-run) capital investment plans, economic optimization would influence the formation of convex forecasts of future demand.[34]

We noted previously (p. 90) that, if multispan forecasts are available, the revision coefficients of equation (10) not only provide a means for testing the form and reconstructing the extrapolation function (4), but also for ascertaining whether the extrapolation function is optimal, provided the time series can be described by the stochastic process (3). When the forecasts consist of extrapolations only, the answer is obtained by testing the equality (14) $\gamma_i^2 = \frac{M_{i+1} - M_i}{M_1}$, where M_i is the mean square error of the ith span forecast. If the equality holds, then $\gamma_i = w_i$ in (3), and forecasting behavior is indeed optimal in the mean square error sense. If the equality does not hold, forecasting behavior may still be optimizing, but either (3) is false or forecasters follow a different optimization criterion.

The test becomes less meaningful in the presence of nonextrapolative components in forecasting. To the extent that these components are either nonforecasting (e.g., measurement errors) or ineffective as forecasting components, they enter the mean square errors M_i.

The contribution of autonomous components to the size of forecasting error was observed to increase relative to that of extrapolation with increasing span in the NBER collection of business forecasts. We

[33] In this case, the revision coefficients γ_k in (10) are no longer equal to the coefficients w_i in the latent structure (3), even if such a structure could be assumed.

[34] As another example, Harberger [9] advocates convex forecasting of future interest rates in optimal planning schemes.

may infer from this pattern that, even if equality (14) held for pure extrapolations, the addition of autonomous components augments the numerator $(M_{i+1} - M_i)$ more than the denominator (M_i). Hence, estimates of $\frac{M_{i+1} - M_i}{M_i}$ would exceed estimates of σ_i^2, and by an increasing proportion with increasing span. If the ratios $\frac{M_{i+1} - M_i}{M_1}$, while differing from γ_i^2, nevertheless vary in the same direction, this is consistent with a weak hypothesis of optimization, in the sense that convex (concave) forecasting is used because the true series itself is convex (concave).

Table 3-2 compares $\frac{M_{i+1} - M_i}{M_1}$ of forward interest rates with corresponding γ_i^2 in Meiselman's revision equations.

TABLE 3-2. Observed and Predicted Revision Coefficients in Forward Rates

Span	i	1	2	3	4	5	6	7	8
Observed [a]	γ_i^2	.50	.28	.16	.10	.08	.06	.06	.04
Predicted [b]	$\frac{M_{i+1} - M_i}{M_1}$	.99	.90	.87	.60	.52	.42	.24	.34

[a] γ_i are regression coefficients in Meiselman revision equations.
[b] Calculated using residual variances in the regressions of Y_{t+k} on ${}_{t+k}F_t$, from Diller's Table 4-21, this volume. The residual variances are, in effect, mean square errors adjusted for bias.

The observed $\frac{M_{i+1} - M_1}{M_1}$ do, indeed, decline as the γ_i^2, though they are larger and decline more slowly. Here the joint hypothesis of optimizing forecasting in a linear time series process (3) cannot be rejected.

No comparable statements can be made about the business forecasts analyzed by Zarnowitz. The forecasts are prima facie not optimal, since they vary by forecaster for the same time series.

IX. CONCLUSION

Can we learn from available forecast data how these forecasts were generated? The analysis presented above arises from an attempt to answer this question. The answer is positive, to a degree believed to be useful, provided forecasts are available for several successive future periods at a given time.

When direct forecast data are not available, empirical insights on how expectations are formed should provide some guidance for specification of expectational forms in econometric models. The usual procedure in this usual case has been to assume a simple extrapolation value for the expectational magnitude on which the observed behavior is based. This extrapolation is often a naive projection of past values or past changes in them, or a more sophisticated geometrically weighted or exponential extrapolation. In addition to relative simplicity, the following claims have been put forth on behalf of the exponential extrapolation: That it represents a type of error-learning forecasting behavior, and that it is an optimal predictor, in the mean square error sense, in certain nonstationary time series.

In this paper we have shown that wide classes of nonexponential extrapolations can also be interpreted as error-learning behavior, and that they can be optimal in types of time series for which the exponential is not optimal. For example, the extrapolation which is optimal for certain stationary linear processes is not exponential but convex, at least in some range. Convex forecasts have properties of regressivity, a behavioral characteristic often desired in the specification of the model.

We also recognize that forecasts or anticipations do not consist exclusively, or even mainly, of extrapolations. We have shown that revision functions (10), which relate revisions of forecasts to the last observed error of forecast, permit not only an analysis of the type of extrapolation embedded in the forecast but also an analysis of the nature and importance of nonextrapolative components in the observed forecasts.

The diagnostic usefulness of the analysis developed in this paper is illustrated more concretely in Diller's investigation of the term structure of interest rates in the following chapter.

REFERENCES

[1] Mills, Edwin, *Price, Output, and Inventory Policy*, New York, 1962.
[2] Griliches, Zvi, "Distributed Lags: A Survey," *Econometrica*, January 1967.

[3] Koyck, L. M., *Distributed Lags and Investment Analysis*, Amsterdam, 1954.
[4] Muth, J. F., "Optimal Properties of Exponentially Weighted Forecasts," *Journal of the American Statistical Association*, June 1960.
[5] Nerlove, M. and Wage, S., "On the Optimality of Adaptive Forecasting," *Management Science*, 1964.
[6] Whittle, P., *Prediction and Regulation by Least Squares*, London, 1963.
[7] Allais, M., "A Restatement of the Quantity Theory of Money," *American Economic Review*, December 1966.
[8] Bossons, J. and Modigliani, F., "Statistical vs. Structural Explanations of Understatement and Regressivity in Rational Expectations," *Econometrica*, April 1966.
[9] Harberger, A., "Techniques of Project Appraisal," in *National Economic Planning*, Universities-National Bureau Conference 19, New York, 1967.
[10] Bailey, M. J., "Prediction of an Autoregressive Variable . . . ," *Journal of the American Statistical Association*, March 1965.
[11] Bierwag, G. O. and Grove, M. A., "Aggregate Koyck Functions," *Econometrica*, October 1966.
[12] Friedman, Milton, *A Theory of the Consumption Function*, Princeton University Press for NBER, 1957.
[13] ———, "Windfalls, the Horizon, and Related Concepts . . . ," in *Measurement in Economics: Studies in Memory of Y. Grunfeld*, C. Christ *et al.*, Stanford, Calif., 1963.
[14] Kessel, R., *The Cyclical Behavior of the Term Structure*, NBER Occasional Paper 91, New York, 1965.
[15] Meiselman, D., *The Term Structure of Interest Rates*, Englewood Cliffs, N.J., 1962.
[16] Pye, G., "A Markov Model of the Term Structure," *Quarterly Journal of Economics*, February 1966.
[17] Telser, L. G., "A Critique of Some Recent Empirical Research . . ." *The Journal of Political Economy*, Part II, August 1967.
[18] Conard, J., *The Behavior of Interest Rates*, NBER, New York, 1966.
[19] Zarnowitz, V., *An Appraisal of Short-Term Economic Forecasts*, NBER Occasional Paper 104, New York, 1966.