

**RATIONAL EXPECTATIONS UNDER CONDITIONS
OF COSTLY INFORMATION**

By

Michael R. Darby*

Discussion Paper Number 45
February 1974

Department of Economics
University of California, Los Angeles

Preliminary Report on Research in Progress
Not to be quoted without permission of the author.

Not to be quoted without the author's
permission:
February 8, 1974

Rational Expectations Under Conditions
of Costly Information

By

Michael R. Darby*

I. Introduction

This note analyzes optimal formation of expectations in the case of costly information. It is seen that under these conditions expectations will normally be formed by an estimator intermediate between Muth's "rational expectations" based on all available information and optimal extrapolative predictors. These two classes are the corner solutions for costless information and prohibitively expensive information, respectively. Since the costs of acquiring information are likely to vary less than the value of information, this model can be applied to explain systematic differences in expectations in markets such as labor for which the value of accuracy of expectations differs between or among buyers and sellers. Similarly expectations about the same variable may differ between markets where arbitrage is also costly.

Section II contains a brief review of the expectations literature. In Section III, I extend a model due to Theil to analyze the effects of costly information. The results are summarized in Section IV.

II. Previous Literature

Expectations of future values are inherent determinants of rational behavior in decisions involving future outcomes. Economists have long been faced with the task of estimating these expectations in order to explain such variables as savings, investment, and market interest rates.

There have been three main approaches to estimation of expectations: (1) extrapolative expectations, (2) rational expectations, and (3) surveys of expectations. The latter approach is not useful for considerations of policy or for predictions of the future. For these purposes, a model of the formation of expectations is necessary.^{1/}

Extrapolative expectations refer to any model based solely upon the past series of values of the variable for which expectations are to be derived, and (perhaps) a time index. Much of this work has been strictly in the framework of distributed lag models. Distributed lags were introduced by Irving Fisher (1923, 1925, 1930) and Tinbergen (1933). Cagan's classic "The Monetary Dynamics of Hyperinflation" (1956) introduced exponentially declining weights into expectations models on the basis of adaptive expectations or error-learning. Muth (1960) provided a statistical justification for the procedure in terms of its optimality as a predictor for a certain class of stochastic models. Error-learning has been generalized in such work as Meiselman (1962), Allais (1966), and Darby (1970). The concept of optimal extrapolative predictor has been applied to broader classes of stochastic models by Mincer (1969) and in ultimate form by Nelson (1970)

who allows the data to select the form of the predictor by utilization of the methods developed by Box and Jenkins (1970).

All of these extrapolative estimators of expectations have met with success when applied to reasonably consistent data periods, but all implicitly ignore the existence of economic theory or deny its usefulness in predicting the future. Muth (1961) has observed that "rational expectations" in a market imply that information is not wasted, and specifically that the market will utilize the information contained in economic theory. This insight has become the basis of many interesting theoretical models in such papers as Gordon and Hynes (1970), Lucas and Prescott (1971) and especially Nelson (1972).

Despite the many attractive properties of these rational expectations models, they possess two interrelated flaws which appear to me to be at times critical: (1) each individual will have identical expectations;^{2/} and (2) they are generally inconsistent with maximizing behavior if it is costly to acquire information. The first difficulty can be surmounted somewhat artificially by making different information available to different groups, but this may break down if "outsiders" can rationally infer "insider's" information from the latter group's behavior.^{3/} The second difficulty is more refractory. Even the "free" publications of the Federal Reserve Bank of St. Louis on money supply growth are not costless to acquire, read, and digest. It is reasonable to assume that this information cost is trivial for bond speculators forming predictions of future rates of inflation, but that seems much less reasonable for millions of individual consumers who have very little payoff from any marginal improvement in accuracy of expectations.

III. Formation of Rational Expectations Under Conditions of Costly Information

The effects of differential costs of information and values of accuracy can be analyzed by use of a simple model of single period decision making. The certainty equivalence approach is adopted for expositional simplicity.

The decision maker in this problem is primarily concerned with the values of two vectors.^{4/}

x is the vector of m instruments controlled by the decision maker.

y is the vector of n market variables unaffected by the individual decision maker.

The criterion function is quadratic in the vectors x and y :^{5/}

$$(1) \quad w(x,y) = a'x + b'y + 1/2(x'Ax + y'By + x'Cy + y'C'x) ,$$

where A and B are symmetric matrices. This criterion function can be viewed as either exact or as a Taylor expansion of the exact criterion function to the second order terms. Any remainder term will not be discussed in the current presentation. It is assumed that the decision maker must choose the value of x for the period at the beginning of the period before it is possible to observe the actual value of y, \hat{y} . This introduces uncertainty, and it is assumed that the decision maker wishes to maximize the expected value^{6/} of $w(x,y)$, $Ew(x,y)$.

Let \bar{y} be the predicted value of the realized value \hat{y} , such that

$$(2) \quad \hat{y} = \bar{y} + u \text{ where } Eu = 0, E(uu') = \Sigma \equiv [\sigma_{ij}] .$$

Theil has shown that under these conditions the value \bar{x} of x which maximizes $Ew(x,y)$ is the same as the value of x which maximizes $w(x,\bar{y})$. The conclusion is that the decision maker operates as if he were maximizing $w(x,\bar{y})$ with \bar{y} known with certainty. Such a \bar{y} is what is referred to in economic discussions of "expectations."

It is further possible to show, following Theil, that the expected loss due to uncertainty, EL , is given by

$$(3) \quad EL = E[-1/2u'Fu]$$

where $F = C'A^{-1}C$, and the assumption of a unique optimal value of x implies A^{-1} exists. This F is either negative definite or semidefinite. It follows that

$$(4) \quad EL = -1/2 \sum_i \sum_j f_{ij} \cdot E(u_i u_j)$$

$$EL = -1/2 \sum_i \sum_j f_{ij} \sigma_{ij} \quad .$$

The expected loss due to uncertainty is therefore a double sum of products reflecting the interaction of x and y and the quadratic term in x in the preference function and the variances and covariances of the errors u in the prediction of y .

In general the prediction \bar{y} will be formed on the basis of information contained in certain exogenous variables (which may include lagged observations) and on past observed values of y . If only past observed own-values

of each y_i are used in the formation of the prediction \bar{y}_i , then the decision-maker is using pure extrapolation of a more or less sophisticated type. If the decision-maker uses some other information than past values of the own-series, then a mixed extrapolation-exogenous information technique is begin used.

Consider the reduced form equations for y :

$$(5) \hat{y} = z' G + \epsilon'$$

where

$$(6) E\epsilon = 0 ; E(\epsilon\epsilon') = \Lambda = [\lambda_{ij}] ;$$

and where z is of order $s \times 1$ and G of order $s \times n$. If all elements of z were perfectly known, and \bar{y} predicted as $z'G$, then the prediction errors u would be identical to the disturbance vector ϵ .^{7/} If, however, some of the elements were impossible or costly to know before \hat{y} , then the decision-maker will be faced with a problem of economizing in his use of information. Partition z into two subvectors z_1 and z_2 , where z_1 consists of the r elements of z which the decision maker knows, and z_2 is the $s-r$ elements which he doesn't know. A similar partition of G into an $r \times n$ submatrix G_1 and an $(s-r) \times n$ submatrix G_2 leads to

$$(7) \bar{y}' = z_1' G_1 + z_2' G_2 + \epsilon'$$

Now consider the prediction of z_2 on the basis of known information.^{8/} Refer to known data other than z_1 as h . Then the actual value of z_2 is

related to z_1 and h according to

$$(8) \quad z_2' = [z_1' \quad h'] \Gamma + v' = z_1' \Gamma_1 + h' \Gamma_2 + v'$$

where the disturbances of the relationship are given in v . Substituting (8) into (7),

$$(9) \quad \hat{y}' = z_1' [G_1 + \Gamma_1 G_2] + h' \Gamma_2 G_2 + v' G_2 + \epsilon' .$$

If the decision maker predicts by setting the disturbances v and ϵ equal to their expectation of zero, then the prediction error $u = \hat{y} - \bar{y}$ is

$$(10) \quad u = G_2' v + \epsilon .$$

Now $E u = 0$, and

$$(11) \quad \begin{aligned} \Sigma &= E (u u') = E [(G_2' v + \epsilon) (G_2' v + \epsilon)'] \\ &= E (G_2' v v' G_2 + \epsilon \epsilon' + G_2' v \epsilon' + \epsilon v' G_2) \\ &= G_2' P G_2 + \Lambda \end{aligned}$$

where $P \equiv [\rho_{ij}] = E(vv')$; $E(v\epsilon') = 0$.^{2/} Since P is positive semidefinite, so is $G_2' P G_2$, and the variance-covariance matrix of the prediction errors thus exceeds the variance covariance matrix of reduced-form disturbances by a positive semidefinite matrix. Note that the coefficients of z_1 are altered in equation (9) in a manner analogous to standard specification analysis.

Thus far the analysis has proceeded as if the data series included in z_1 and h were given, but surely this is not the case. The addition of a data series to h -- and this may be formally extended to include an element

of z_2 -- will in general involve some cost of information gathering and some benefit in the form of reduction in the expected loss given in expression (4). The decision-maker's problem can be expressed most generally as a (rather difficult) nonlinear programming problem: Let $p' = (p_1 \dots p_T)$ be the vector of information costs for the s elements of z and all potential candidates for h (some p_i may be infinite). Let $\delta' = (\delta_1 \dots \delta_N)$ be a vector which represents whether or not each potential data series is used. That is $\delta_i = 0$ if the series is not used. The total information cost is $IC = p' \delta$. Now, note that Σ was shown to be a function of the data series used in the formation of the prediction \bar{y} , so that the σ_{ij} can be taken as a function of δ , $\sigma_{ij}(\delta)$. The total cost of uncertainty is

$$(12) \quad UC = EL + IC$$

$$UC = -\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n f_{ij} \sigma_{ij}(\delta) + p' \delta .$$

Therefore, the nonlinear programming problem is:

$$(13) \quad \min_{\delta} \left\{ -\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n f_{ij} \sigma_{ij}(\delta) + p' \delta \right\}$$

subject to $\delta_i = 0$ or $1, i = 1, \dots, T$.

This form of the problem (and its solution) is too general to be of much interest other than as an organizing framework which emphasizes the functional relationship of the σ_{ij} 's to δ . Let $R \equiv [r_{ij}] = G_2' P G_2$. Then,

$$(14) \quad UC = -\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n f_{ij} \lambda_{ij} - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n f_{ij} r_{ij}(\delta) + p' \delta .$$

The first right hand side term reflects the irreducible loss due to the stochastic nature of the world. The second and third terms however vary with δ and reflect the marginal cost of any particular prediction scheme. A local maximum will exist at δ^0 if adding any variable (changing a zero δ_i to 1) or subtracting any variable (changing a unit δ_i to 0) would increase UC. That is those variables with the largest (weighted) effect of decreasing prediction error variance relative to their cost of acquisition will be included. These are variables which are highly correlated with the residuals of the higher order predictions (v),^{10/} particularly with residuals corresponding to variables having large coefficients (g_{ij} 's) in the reduced form equations for important (in terms of the f_{ij} 's) components of y .^{11/}

For the important special case of the prediction of a single market series, the situation is more tractable. Here y is scalar ($n=1$), so

$$(15) \quad UC = -\frac{1}{2} f_{11} \lambda_{11} - \frac{1}{2} f_{11} r_{11}(\delta) + \sum_{i=1}^T \delta_i P_i .$$

The incremental change in the cost of uncertainty in moving from δ^0 to δ^1 which is identical to δ^0 except for one δ_k ($k \geq r+1$) which is 1 instead of 0, is:

$$(16) \quad UC(\delta^1) - UC(\delta^0) = -\frac{1}{2} f_{11} \sum_{i=1}^{s-r} \sum_{j=1}^{s-r} g_{1,i+r} g_{1,j+r} [\rho_{ij}(\delta^1) - \rho_{ij}(\delta^0)] \\ + P_k .$$

Recalling that the value of general variance reduction ($-1/2 f_{11}$) is positive, it is seen that the addition of the k th variable adds its cost P_k but reduces

loss due to prediction error by $-1/2 f_{11}$ times the reduction in the variance or covariance of the predictions the unknown determinants of y weighted by the associated coefficients in the determination of y . If $r+1 \leq k \leq s$, so that an element of z was utilized then the prediction variance and covariance for that element would go to zero. Other ρ_{ij} ($i \neq k-r, j \neq k-r$) will also be affected to the extent that the development of the added variable adds information about the remaining unknown values of z .

This model suggests that the actual method of prediction (expectation formation) used will vary (1) with differences in the value of variance reduction ($-\frac{1}{2} f_{11}$), (2) with differences in the predictability of exogenous variables ($\rho_{ij}(\delta)$), and (3) with differences in the cost of data acquisition (p).

The value of variance reduction can normally be derived directly from the objective functions posited for derivations of "rational expectations." Different participants in a market will generally have different objective functions because they are on different sides of the market or because they face different constraints or because they are not identical. The first two differences may be observable and so be the basis for a theory of differential surprise in the face of a policy shock. A simple example of different values of information is given below.

Differences in predictability of exogenous variables provide a theoretical basis for changes in methods of expectations formation between a period of stable monetary growth and periods of large variance in monetary growth such as the 1950's against the 1960's in the U.S. or stable economies against hyperinflations.

Costs of data acquisition are difficult to quantify and use. Clearly, if they are infinite no exogenous information will be used and purely extrapolative models of expectations would be appropriate. If information were really free, everyone would have identical "rational expectations" based on the best reduced form predictive equation for all information. Considering the intermediate case where information costs are neither trivial or infinite, some structure must be posited to relate costs to the variables which determine the value of information. The simplest assumption is that p is a constant to all participants. It may be that there are market structures which share the information costs among users to whom accuracy has relatively little value, such as mutual funds for individuals facing relatively small wealth constraints. Put another way, constant informational costs are a basis of scale economies leading to elimination of small participants from the market (and their replacement by mutual funds in the example).

One simple example will illustrate the operation of the model.^{12/} Farmers deciding on production of hogs can be characterized by objective functions of the type in equation (1). The instrument is the quantity of hogs to produce q and the market variable to be predicted is the price π when production is completed. Assume that total cost TC depends on the quantity of pigs produced and the fixed capital of the firm measured in pig pens n :

$$(17) \quad TC(q,n) = \alpha_0 n + \alpha_1 q + \alpha_2 \left(\frac{q}{n}\right)^2 .$$

That is--at least approximately in the relevant range--marginal cost is $\alpha_1 + \frac{\alpha_2}{n^2}$. The objective function of a hog farm--revenue less variable costs--is

$$(18) \quad w(q, \pi; n) = q\pi - \alpha_1 q - \left(\frac{\alpha_2}{n}\right)q^2.$$

In terms of the general form of equation (1), $a = -\alpha_1$, $b = 0$, $A = -2\alpha_2/n^2$, $B = 0$ and $C = 1$, all scalars. So,

$$f_{11} = F = C' A^{-1} C ;$$

$$(19) \quad f_{11} = \frac{-n^2}{2\alpha_2}.$$

So the value of variance reduction is $\frac{1}{4\alpha_2} n^2$ which rises in a very predictable way with farm size as measured by n . If we assume π fixed and α_2 distributed independently of farm size, the probability that a farmer will use a particular costly information series^{13/} to predict the future price π rises with the square of farm size. This has a common sense interpretation: The larger is n , the flatter is the marginal cost curve and hence the further will be q from the optimal q for any given difference between actual and expected prices. The loss in profits is (at least approximately) equal to $\frac{1}{2}$ this distance times the price difference.

The most interesting uses of the model will arise, I believe, in cases where specific information structures can be posited and their implications for reactions to specific shocks analyzed, as perhaps in search models of unemployment. The desirable implication of different expectations is inconsistent not only with optimal extrapolative and "rational" models of

expectations formation, but also with "efficient markets." Differences of opinion imply unexploited opportunities for profit if there are no transactions costs, no wealth constraints, and so forth. This does not seem to characterize the sort of markets for which acquisition of information involves nontrivial costs. Thus bond speculators will not be able to arbitrage on the basis of more accurate expectations of inflation than are used by consumers in determining their purchase of a washing machine.

I have neglected in the formal model the effects of uncertainty about the parameters of the true model and of more complicated structures of autocorrelation in disturbances. While both are appropriate areas for future research, it is to be hoped that the results derived on the basis of these simplifying assumptions will hold approximately.

IV. Summary

Optimal extrapolative and "rational" models of expectations formation are based on polar assumptions of the cost of acquisition and use of information other than the series for which expectations are being formed. These assumptions imply that identical methods of expectations formation and hence identical expectations will be used by all market participants. There are interesting cases for which information is neither infinitely costly nor free and for which systematically different methods of expectations formation will be used by different market participants. A general model is presented for analysis of such problems. Factors causing differences in expectations models are (1) differences in the value of accuracy, (2) differences in the predictability of exogeneous variables, and (3) differences in the cost of data acquisition.

FOOTNOTES

* Acting Associate Professor of Economics, University of California, Los Angeles. The author would like to acknowledge helpful comments by Jack Hirshleifer, Henri Theil, and Arnold Zellner on earlier versions of this paper. Any remaining errors should be attributed to the author.

1/ Survey data may permit direct tests of these models.

2/ This is also true for optimal extrapolative predictors.

3/ See Jaffe (1973).

4/ It is possible that the values of x may be subject to some linear constraints, but this merely effects the particular values of the f_{ij} in equation (4) below, and these values are not of primary interest; so this complication is suppressed here in the interest of simplicity. The following presentation through the derivation of the expected loss, EL , is based on the work of Theil (1958, 1964), although the interpretation of some of the concepts is of course significantly different in philosophy and detail. Other important references in the certainty equivalence framework are to Simon (1956), Theil (1957), and Malinvaud (1969). An alternative to this approach is found in decision theory, which however has yet to be successfully applied to macro-economic analysis. Nevertheless it seems a fruitful avenue for exploration as to whether a simplified decision theoretic approach will yield interestingly different results. References are to Raiffa and Schlaifer (1968), Pratt, Raiffa, and Schlaifer (1965), and Aoki (1967).

5/ The criterion function will be assumed to be expressed in terms of

dollars, which corresponds to our ignoring any linear constraints on x . It follows from Theil's work that for, say, utility maximization under constraints, the definition of F in equation (3) is merely complicated, and we must use the marginal utility of income to convert the expected loss from units of utility to dollars.

6/ To avoid possible confusion with the statistical concept of expected value, the term prediction will be used for the decision maker's "expectations" in this section.

7/ It is necessary to add the restriction $\bar{y} = z'G$, since the current ϵ vector might be correlated with past ϵ 's, so that an improved prediction would be possible if the noncontemporaneous correlation of disturbances was utilized in the prediction procedure. This possibility is usually assumed away in econometrics, and can probably be justified in terms of a fully specified reduced form.

8/ This is what Theil (1958, p. 17) refers to as higher order prediction.

9/ Depending particularly on the nature of h , this assumption might not be completely fulfilled, but its relaxation would not seem to alter the substance of the argument below.

10/ And therefore fairly orthogonal to the other included variables.

11/ See Theil's (1971, pp. 167-171) discussion of the incremental contribution of explanatory variables in a multiple correlation framework for a thorough discussion of a similar problem.

12/ I am indebted to Kenneth Koford for suggesting this example.

13/ Or alternatively, the probability that farmers will pay the cost of insuring against price fluctuations through use of the futures markets.

REFERENCES

- Allais, M., "A Restatement of the Quantity Theory of Money," Am. Econ. Rev., December 1966, 56: 1123-57.
- Aoki, M., Optimization of Stochastic Systems: Topics in Discrete-Time Systems, New York: Academic Press, 1967.
- Box, G. E. P. and Jenkins, G. M., Time Series Analysis: Forecasting and Control, San Francisco: Holden-Day, Inc., 1970.
- Cagan, P., "The Monetary Dynamics of Hyperinflation," in M. Friedman, ed., Studies in the Quantity Theory of Money, Chicago: Univ. of Chicago Press, 1956.
- Darby, M. R., "Allais' Restatement of the Quantity Theory: Comment," Am. Econ. Rev., June 1970, 60: 444-46.
- Fisher, I., "The Business Cycle Largely a 'Dance of the Dollar'," J. Am. Stat. Ass., December 1923, 18: 1024-28.
- _____, "Our Unstable Dollar and the So-Called Business Cycle," J. Am. Stat. Ass., June 1925, 20: 179-202.
- _____, The Theory of Interest, [1930], New York: Augustus M. Kelley, 1965.
- Gordon, D. F. and Hynes, A., "On the Theory of Price Dynamics," in E. Phelps, et. al., Microeconomic Foundations of Employment and Inflation Theory, New York: W. W. Norton & Co., Inc., 1970.
- Jaffe, J. F., "The Value of Information for Trading Purposes," unpublished manuscript, 1973.
- Lucas, R. E., Jr. and Prescott, E. C., "Investment Under Uncertainty," Econometrica, September 1971, 39: 659-81.
- Malinvaud, E., "First Order Certainty Equivalence," Econometrica, October 1969, 37: 706-718.
- Meiselman, D., The Term Structure of Interest Rates, Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1962.
- Mincer, J., "Models of Adaptive Forecasting," in J. Mincer, ed., Economic Forecasts and Expectations: Analyses of Forecasting Behavior and Performance, New York: National Bureau of Economic Research, 1969.

Muth, J. F., "Optimal Properties of Exponentially Weighted Forecasts," J. Am. Stat. Ass., June 1960, 55: 299-306.

_____, "Rational Expectations and the Theory of Price Movements," Econometrica, July 1961, 29: 315-35.

Nelson, C. R., "Rational Expectations and the Predictive Efficiency of Economic Models," Report 7209, Center for Mathematical Studies in Business and Economics, Univ. of Chicago, March 1972.

_____, "Testing a Model of the Term Structure of Interest Rates by Simulation of Market Forecasts," J. Am. Stat. Ass., September 1970, 65: 1163-79.

Nerlove, M. and Wage, S., "On the Optimality of Adaptive Forecasting," Management Science, January 1964, 10: 207-224.

Pratt, J. W., Raiffa, H., and Schlaifer, R., Introduction to Statistical Decision Theory, New York: McGraw-Hill Book Company, 1965.

Raiffa, H. and Schlaifer, R., Applied Statistical Decision Theory, Cambridge, Massachusetts: The M.I.T. Press, 1968.

Simon, H. A., "Dynamic Programming Under Uncertainty with a Quadratic Criterion Function," Econometrica, January 1956, 24: 74-81.

Theil, H., Economic Forecasts and Policy, Amsterdam: North-Holland Publishing Company, 1958.

_____, "A Note on Certainty Equivalence in Dynamic Planning," Econometrica, April 1957, 25: 346-49.

_____, Optimal Decision Rules for Government and Industry, Amsterdam: North-Holland Publishing Company, 1964.

_____, Principles of Econometrics, New York: John Wiley & Sons, Inc., 1971.