ECONOMIC RESEARCH REPORTS

UNFALSIFIED EXPECTATIONS: AN ALTERNATIVE PERSPECTIVE ON MODELLING EXPECTATIONS IN MACROECONOMICS

by

Peter Rappoport

R.R. #85-16

May 1985

C. V. STARR CENTER
FOR APPLIED ECONOMICS

NEW YORK UNIVERSITY
FACULTY OF ARTS AND SCIENCE
DEPARTMENT OF ECONOMICS
WASHINGTON SQUARE
NEW YORK, N.Y. 10003

First draft December 1983
Second draft December 1984
Third draft May 1985

UNFALSIFIED EXPECTATIONS: AN ALTERNATIVE PERSPECTIVE ON MODELLING EXPECTATIONS IN MACROECONOMICS

Peter Rappoport

Department of Economics

New York University

I am grateful for helpful comments from Jim Albrecht, Marcia Marley, Mark Schankerman, and especially Roman Frydman, without whose encouragement this paper would not have been written. Mary McCarthy provided invaluable research assistance. I am also grateful for financial support provided by the C. V. Starr Center for Applied Economics at New York University.

1. Introduction

The rational expectations approach to modelling expectations seeks to impose on it the rigors of rational behavior. Muth (1961, p. 317) stated that this would require that economic agents' "expectations . . . are essentially the same as the predictions of the relevant economic theory" and "tend to be distributed . . . about the predictions of the theory." In macroeconomic applications, this is followed by modelling the expectations of a rational representative economic agent as the predictions of the analyst's model. Under the assumption that this is the true model of the system, the agent's prediction errors will be free of systematic components. I will refer to expectations that satisfy this condition as "RE." In the course of the articulation of this approach, it has come to be regarded as irrational to forecast using any prediction method other than that mandated by the underlying model. The explanation for this is that the errors so generated will contain systematic components, the exploitation of which will increase profits (cf. Barro and Fischer (1976, p. 162)).

Rationality is more usually characterized by the optimal use of given resources. Many authors have pointed out that RE can be reconciled with this conventional definition only at the cost of assuming individuals to be endowed with accurate and exhaustive information on the economy.[1] In the absence of an account of how people acquire this information, RE comes preciously close to perfect foresight. On the other hand, there are problems with using the conventional definition unmodified in any way. The drawback of pursuing this line too far is that it can be employed to justify the rationality of any expectations scheme, on the grounds that the individual steadfastly believes it to predict the best, irrespective of its performance. Between these two extremes, there lies the informal idea that rational people will attempt to learn how to predict as well as possible. They may be regarded as experimenting with a particular model until it proves to be inadequate, at which time they modify it or reject it in favor of another.[2]

This paper seeks to formalize these loose ideas, and to examine their implications for empirical modelling of expectations. It is concerned with the content of rational behavior for an individual who attempts to learn the true model. The representative individual is assumed to adopt a particular model for forecasting at a given date. Ignorant of whether it is the true model or not, at each subsequent date, he or she subjects the experience of using it to statistical tests that examine whether the model commits systematic mistakes. For as long as the model does not reveal the presence of systematic errors, it is said to be <u>unfalsified</u> and to generate unfalsified expectations. Once systematic errors are detected, the model is said to be <u>falsified</u>. It bears stressing that it is quite possible that the forecast errors of an unfalsified model contain systematic mistakes. However, the crucial point is not whether systematic portions of errors <u>exist</u>, but whether they can be <u>detected</u> on the basis of available information.

Whether it is rational to use an unfalsified model depends again on the alternatives available to the individual. Many models may be unfalsified at a point in time. The problem arises of distinguishing among them using some other criterion, for example, one that chooses the model with minimum past or expected squared forecast errors.[3] Such a criterion may also rank falsified models above unfalsified ones. Consequently, caution has to be exercised before it is concluded that an agent will use a particular unfalsified model. However, there is an unambiguous relationship between falsified models and both RE and conventional rationality. Subject to the fallibility of the testing procedures employed (i.e., the possibility of committing a Type I error) a falsified model cannot qualify as RE, since the model exhibits systematic errors. Similarly, as the agent has perceived the presence of these errors, they constitute a part of available information, and suggest to the agent that the model can be bettered. Thus, a falsified model also runs afoul of the conditions required for conventional rationality. In short, unfalsifiedness is a necessary but insufficient condition for RE and conventional rationality.

The relationship between unfalsifiedness and the two concepts of rationality is exploited in this paper to examine how various empirical expectations proxies would have fared in the hands of rational economic agents. Since unfalsifiedness is a necessary condition for RE and conventional rationality, any legitimate proxy for rational expectations (in either sense) must remain unfalsified for the duration of any interval in which it is to be employed.

In order to impart substance to the concept of unfalsified expectations, it is necessary to have a theory of how agents come to perceive the systematic mistakes of their forecasting models and of what they do with this information. This is the subject of the next section, where the concept of the model used by the representative agent and the assumptions surrounding it are discussed. It is assumed that, at a particular historical date, the agent adopts an economic specification that is used to forecast. As experience of using the forecasting method accumulates, the agent is assumed to subject these data to statistical tests. The null hypotheses of the tests describe certain properties of non-systematic errors. Thus, the agent tests each period to see if the accumulated errors are serially correlated, orthogonal to the predictions, and have a zero mean. For as long as these hypotheses remain unrejected, the model is unfalsified, and no evidence has been assembled that it is not RE or rational in the conventional sense.

The following two sections apply this framework in two types of model of expectation formation that have been prominent in the empirical literature: adaptive expectations and RE models. Over the last decade, the accepted practice for constructing empirical proxies for RE has been to take the fitted portion of a least squares projection of the variable to be forecast on the predetermined variables in the system. I shall refer to data so constructed as "empirical RE." Under the maintained hypothesis that the model (i.e., the list of predetermined variables) is correctly specified, this provides an efficient estimate of RE, given the available data. However, there is no guarantee that the maintained hypothesis is correct. This hypothesis is the null hypothesis of the falsification tests.[4]

Adaptive expectations models were early casualties of the RE approach. They were ruled out because their use amounted "to supposing that the public's method of forming expectations of inflation was very irrational in the sense of being widely inconsistent with the actual inflation process" (Sargent, 1971, p, 724). Sargent's criticism does not necessarily imply that the public had, at the time, sufficient information to substantiate the supposed inadequacy of adaptive expectations in tracking the inflation process. Thus, while an arbitrarily chosen adaptive expectations mechanism may, in general, be irrational in Muth's sense, because it implies that the public believes the inflation process to differ from that which actually generates it, it is not automatically falsified at a given point in time by data generated by that process.[5]

These two types of models are described in section 3. Certain econometric problems which have to be surmounted before the falsification tests can be applied to empirical RE data are also dealt with in this section. The following section examines the performance in falsification tests of examples of these models in forecasting the quarterly rates of change of the GNP deflator and the money supply (M1). In general, the empirical RE models fare badly, often experiencing early, persistent and summary rejections of at least one of the falsification tests. The evidence of systematic errors thus assembled implies that they cannot be considered legitimate measures of RE, and, subsequent to the date of their falsification, their use for forecasting would not be rational in the conventional sense. In comparison, more "naive" models fare better, some of them passing all falsification tests in each period in the sample.

Two principal implications of the paper are discussed in the final section. First, the results of falsification tests entail that RE has been mismeasured in empirical tests of the major propositions of the new classical economics, and the deleterious effects of this mismeasurement on inference are briefly mentioned. Second, the failure of economists' models to survive the falsification tests suggests that learning needs to be introduced explicitly into accounts of rational expectations. The falsification framework suggests that learning involves the forecaster in experimentation with a sequence of models, each of which is used to forecast until it is falsified. However, it does not seem that the "choice" or "discovery" of new models is amenable to an analysis based on rational behavior. Hence, the principle of rationality does not provide a complete guide to modelling expectations.

2. Falsification Tests

The forecasting behavior of economic agents is typically characterized by attributing to them a model of the variable they are supposed to predict. The perspective of this paper augments this description by assuming that economic agents also carry out falsification tests, to ascertain whether the model with which they have been endowed commits systematic mistakes. This section makes explicit the nature of the forecasting equations attributed to agents, and of the falsification tests they are assumed to carry out on them. In addition to offering a positive description of learning behavior, this framework supplies a means by which a researcher can check whether a model he or she wishes to attribute to agents satisfies the conditions for RE or conventional rationality. To accomplish this, the investigator must replicate the falsification tests carried out by agents. The next section deals with this topic.

In pursuit of concreteness, I consider the activities of a stylized economic agent, of female gender, referred to as a "forecaster," who is concerned with making "efficient" use of available information in predicting the variable y with a one-period forecast horizon.[6] The forecaster is assumed to be endowed with a model of y. Her information set for predicting y_t one period in advance comprises available data on the variables in the model, denoted "X_t." (This does not imply that X_t necessarily contains any data realized at t.) It also contains numerical values of the parameters of the model, denoted by "γ". All models considered in the sequel are linear in the variables, and the coefficients used by the forecaster do not vary over time. The typical model, m, may therefore be written:

$$(1) \qquad y_t = X_t^m \gamma^m + \varepsilon_t^m$$

Thus, the forecast from model m, denoted y_t^m, will be $X_t^m \gamma^m$ (X^m is the list of variables in model m), and the associated forecast error will be ε_t^m.

The forecaster's behavior in theoretical RE models falls under the rubric of this description. Here, γ^m is the coefficient of the "population projection"[7] of y on X^m, which includes all predetermined variables in the semi-reduced form equation for y. The error, ε^m, is accordingly guaranteed to be white noise.[8] Empirical RE models can also be characterized in this framework. The investigator again specifies a list of variable X^m that are supposed to account for the observed historical variation in y. However, in contrast to the theoretical RE case, the errors are not guaranteed to be white noise[9], since the investigator cannot be sure that X^m includes all variables in the semi-reduced form for y. The investigator, while ignorant of "the" value of γ^m, assumes that it is known to the forecaster. This is an unambiguous assumption in the case where X^m is exhaustive, and ε^m is indeed white noise. However, when the investigator omits variables capturing parameter shifts or structural changes, there is no unique value of the population projection coefficient. Its value depends on the way in which limits are taken. In order to illustrate this problem, it will be useful to introduce a simple example to which I will return subsequently. Suppose y follows the process:

(2) $$y_t = 3 + u_t \quad t < t_o$$
$$y_t = 5 + u_t \quad t \geq t_o$$

where u_t is a white noise series. The investigator, however, attributes to the forecaster a model in which y is projected only on a constant, so $X^m_t = 1$ for all t. Thus, the finite sample projection coefficient of y on X^m is simply the sample mean

$$\hat{\gamma}^m = \sum_{t=t_\ell}^{t_u} y_t, \quad t_\ell \leq t_u$$

and it is simple to see that, for example,

$$(3)\qquad \gamma^m = \text{plim}\ \hat{\gamma}^m = \begin{cases} 5 \text{ if } t_\ell \text{ fixed, } t_u \to \infty \\ 4 \text{ if } t_\ell,\ t_u \to \infty \text{ at the \underline{same} rate} \\ 3 \text{ if } t_u \text{ fixed, } t_\ell \to \infty \end{cases}$$

Which of these values of γ^m is used is somewhat arbitrary. One may make an argument for $\gamma^m = 3$ as follows. In the theoretical RE case, one justification for endowing the forecaster with the population projection coefficient is that, if the model has always been in a rational expectations equilibrium, the value of the population coefficeint can be inferred by observing correlations on infinite samples of past data. Note that, if this argument is extended to the case of an appropriately misspecified empirical RE model, the doubly infinite vectors of forecasts and errors will not be orthogonal. These two vectors will be orthogonal if $\gamma^m = 4$, but this is not the value that the forecaster could have inferred from regressions on the infinite past data set. The case where $\gamma^m = 5$ is unsatisfactory for similar reasons. In the sequel, γ^m is understood as the population coefficient computed on an infinite sample of past data, which results in the value 3 in the example.

The example also illustrates that (whichever interpretation of γ^m is chosen) the errors will not be white noise, since, in this case, the mean of the error changes over time. Thus, if X^m is misspecified, it will not be rational in the RE sense for the forecaster to use that model, nor will it be rational in the conventional sense for her to use it once the misspecification has been discovered. The falsification tests carried out by the forecaster enable her to check whether (continued) use of the model is rational.

If the forecaster starts to use the model at date 0, at the end of the T-th period (T > 0) she has available a history of the forecast errors of m, summarized by the vector[10]

$$\underline{\varepsilon}^m_{1-T} = (\varepsilon^m_1, \ldots, \varepsilon^m_T)$$

The falsification tests examine whether the accumulated data on ε^m evidences the presence of systematic components.

Systematic errors will occur when (1) is misspecified, either because variables are omitted or coefficients change over time. (By a suitable redefinition of terms, both of these cases can be subsumed under the case of omitted variables, but it will be useful to keep them separate.) In general, the forecaster may consider the correlation of ${}_1\underline{\varepsilon}_T^m$ with indicators of structural changes or with available data on variables suspected to be relevant. Each model or historical period may suggest a different set of such variables. For practical purposes, it would be useful to develop a single battery of tests applicable to all models. Those described below use only ${}_1\underline{\varepsilon}_T^m$ and the corresponding history of forecasts,

$${}_1\underline{y}_T^m = (y_1^m, \ldots, y_T^m)$$

However, they are likely to detect in the errors the presence of the systematic components that are symptomatic of the misspecifications described above.

As macroeconomic time series are typically highly autocorrelated, the presence of omitted variables in the forecast errors should be revealed by an examination of their serial dependence. The most comprehensive (and mechanical) method of accomplishing this is by the use of "Q-statistic" (Box and Pierce (1970), Ljung and Box (1978)):

$$Q_T = T\cdot(T + 2) \sum_{k=1}^{M} \hat{\rho}_k^2({}_1\underline{\varepsilon}_T^m)\cdot(T - k)^{-1} \tag{4}$$

where $\hat{\rho}_k^2({}_1\underline{\varepsilon}_T^m)$ is the estimate of ρ_k, the k-th autocorrelation of the deviation of ${}_1\underline{\varepsilon}_T^m$ from its mean vector. Box and Pierce show that, under the null hypothesis

(5) $R_0 : \rho_k = 0, \quad k = 1,\ldots,M$

the distribution of Q_T may be approximated by a chi-square distribution.[11,12] Since the computation of the Q-statistic examines the serial dependence of the deviation of ${}_1\varepsilon_T^m$ from its mean vector, it will not detect deviations of the mean error from zero. The forecaster is assumed to examine for presence of this systematic component by testing

(6) $M_0 : E\varepsilon^m = 0$

against

$M_1 : E\varepsilon^m \neq 0$

by use of a simple t test, which has T - 1 degrees of freedom when the test is run at time T.

A final test examines the correlation between errors and predictions. At the end of each period, T, the forecaster runs the OLS regression:

(7) $\varepsilon_t^m = \alpha + \beta y_t^m + \eta_t, \; t = 1,\ldots,T$

at the end of period T, which generates the estimates $\hat{\alpha}_T$, $\hat{\beta}_T$. She then tests the hypothesis

(8) $0_0 : [\alpha,\beta] = [0,0]$

against the alternative

$0_1 : [\alpha,\beta] \neq [0,0]$

This is accomplished by computing F_T, the F-statistic under 0_0, and referring it to an F(2,T - 2) distribution. The test thus examines, each period, whether m satisfies the conditions that its forecast errors have zero mean and are orthogonal to the forecasts. This condition is imposed on empirical RE proxies by the least squares regression procedure used to construct them. While testing this orthogonality condition may thus be regarded as a test of the empirical RE restriction, the motivation for the orthogonality test in fact lies in its potential for detecting parameter shifts in the RE context.

The operation of the orthogonality test in the case of empirical RE models can be seen by considering the example (2), in which γ^m is taken to be 3, hence $y^m_t = 3$, for all t. Then the forecast error is

$$\text{(9)} \qquad \varepsilon^m_t = \begin{cases} u_t & \text{for } t \leq t_o \\ u_t + 2 & \text{for } t > t_o \end{cases}$$

Thus, as t increases past t_o, the correlation of the forecasts and errors becomes more marked, and would eventually be detected by a forecaster who ran orthogonality tests on the successively larger samples, ${}_1\underline{\varepsilon}^m_T$, ${}_1\underline{y}^m_T$, T = 1, 2, ... Subsequent to the date which the forecaster could falisfy the model in this way, it could not be maintained that use of the model to forecast would be consistent with RE or conventional rationality[13].

To recapitulate, the forecaster is assumed to be endowed at T = 0 with a forecasting equation. To forecast y_t she combines the vector of observations on X^m_t with γ^m, which is the population regression coefficient of y_t on X^m_t. At each date T, she also carries out tests of the hypotheses O_0, R_0 and M_0. These are necessary conditions for ε^m to be white noise. While these tests were discussed in the context of (empirical) RE models, they can also be applied to "ad hoc" models attributed to the forecaster by an investigator.

It is thus proposed that the forecaster's behavior be modelled by the outcomes of the tests of R_0, M_0 and 0_0 at each date T. Given that the forecaster employs these classical statistical tests, her decision variable is the probability of a Type I error, α. Ideally, the critical values for each test are to be selected so that the probability of rejecting the true model in the set of all tests at all dates (suitably discounted) taken as a whole is α.[14] These critical values would then take into account the sequential nature of the testing procedure. Unfortunately, the difficulties in designing sequential tests for all but the simplest of cases render this problem intractable for the statistics examined in this paper.[15] Therefore, it is well to state the procedure used here. The three tests are treated each period as if they were independent, and each test is repeated each period at the same significance level. This amounts to ignoring the fact that the survival of the model up to any date is conditional on its having passed the tests to which it was subjected prior to that date. The results reported below examine tests at the 5% and 1% levels of significance. While probabilities of Type I error are usually employed in the econometrics literature, their use in repeated tests lacks a rigorous justification.

The consequences of using these significance levels in the orthogonality test may be analyzed from a Bayesian standpoint, following the approach of Leamer (1978, Ch. 4). The classical decision rule instructs the forecaster to reject the null hypothesis when F_T exceeds the 1% or 5% critical value for a sample of size T. An alternative (strict) rule is one that instructs the forecaster to discard the null hypothesis if the posterior odds are less favorable to it than the prior odds. Since the odds ratio is a function of the sums of squared residuals from the restricted and unrestricted regressions, this rule can be translated into a critical value for F_T. The decisions it entails can then be compared with those of the classical rule. For samples smaller than 130 observations, the 1% classical strategy is more favorable and the 5% classical strategy less favorable to the null hypothesis than this "Bayesian" rule. The 1% critical value will tolerate small, steady deteriorations in the odds in favor of the null hypothesis. Conversely, a 5% critical value requires the odds to improve each period. This suggests it is most advisable to focus attention on tests at the 1% level, at least as far as the orthogonality test is concerned.[16]

3. Econometric Considerations

The previous section outlined the activities of a rational forecaster that determine when she will not use a particular model for forecasting. The purpose of this section is to demonstrate how the investigator may replicate the forecaster's behavior and so ascertain whether the model is admissible as a proxy for the expectations of rational agents. The analysis is limited to two types of expectations models, adaptive expectations and RE models.

(i) Empirical RE proxies

Over the last decade, the practice for specifying expectations variables has followed the approach used in Sargent (1973). The investigator specifies a list of variables, X, and postulates that the rational forecast of y is given by the expectation of y_T conditional on X_T. The expectation attributed to the forecaster is thus the population projection of y on X:

(10) $$y_T^r = X_T\gamma,$$

where γ is the appropriate population projection coefficient. To be consistent with the notation of the last section, X and γ should have the superscript "r," to indicate that they belong to the model r. This is dropped in this section for simplicity. Although the forecaster is assumed to know γ, the investigator does not know it, and has to estimate it, in order to create an empirical proxy for y^r. This is usually accomplished by regressing y on X over the sample available to the investigator, say $t = 1,\ldots,S$, and yields the OLS estimate g_S. The investigator's proxy for y^r is then given by [17]

$$(11) \qquad y_T^r(S) = y_T - y_T^r(S).$$

The corresponding forecast error proxy is denoted:

$$(12) \qquad e_T(S) = y_T - y_T^r(S).$$

Given the information available to the investigator, and the belief that the model is correctly specified, g_S is regarded as an efficient estimate of γ, and so $y_T^r(S)$ and $e_T(S)$ are the best approximations the investigator can make to y_T^r and ε_T respectively. However they are not useful for reconstructing the forecaster's falsification tests, which will contribute towards determining whether the forecaster would actually use the investigator's model. To see this, consider first the investigator's attempt to replicate the orthogonality test carried out at T = S. Equation (7) shows that this involves projecting ${}_1\underline{e}_S(S)$ on a constant vector and ${}_1\underline{y}_S^r(S)$. These are precisely the residual and fitted vectors from the investigator's regression of y on X (using the sample $1,\ldots,S$). Hence, they are orthogonal by construction, and the residual vector will be orthogonal to a constant vector, if X includes one. Hence, $\hat{\alpha}_S$ and $\hat{\beta}_S$ will be identically zero, as will the test statistic, F_S. Thus, using this method, the investigator will never infer that at S the forecaster would reject the orthogonality hypothesis, irrespective of the actual performance of the forecasts (10). This result is purely an artifact of the way the investigator's proxies, ${}_1\underline{y}^r_S(S)$ and ${}_1\underline{e}_S(S)$, are constructed. The same problems occur, <u>mutatis mutandis</u>, with the mean test. The serial correlation test is not affected, as the OLS procedure does not constrain the autocorrelation of ${}_1\underline{e}_S(S)$.

Other difficulties arise if it is attempted to replicate orthogonality tests occurring at T > S using expectations proxies constructed on the basis of g_S:

$$(13)\qquad \begin{aligned} {}_1\underline{y}^r_T(S) &= (X_1 g_S, \ldots, X_S g_S, \ldots, X_T g_S) \\ {}_1\underline{e}_T(S) &= (y_1 - X_S g_S, \ldots, y_s - X_S g_S, \ldots, y_T - X_T g_S) \end{aligned}$$

The forecast error is given by:

$$(14)\qquad e^r_T(S) = X_T(\gamma - g_S) + \varepsilon_T,$$

where ε_T is white noise under the assumption that (1) is the true model. The correlation of $X_T g_S$ and (14), which is examined by the orthogonality test, depends on the variance of g_S, which is fixed as $T \to \infty$. Moreover, as long as the X's have a non-zero mean, that of $e_T(S)$ will be non-zero, and as long as the X's are serially correlated, $e_T(S)$ will be serially correlated. Hence, one would expect all three tests to yield rejections, even when the forecaster's model is correct. The source of these rejections is the error in the investigator's estimate of the parameter values used by the forecaster, which is measured by $g_S - \gamma$, a number that is in general non-zero, and fixed as $T \to \infty$. Thus, the problem with falsification tests based on the empirical RE proxies e(S) and $y^r(S)$ is that their results are sensitive to the sample used by the investigator. From the point of view of modelling the forecaster's falsification tests, the terminal date of this sample is arbitrary.

An alternative approach, which is pursued in this paper, is to use the following proxies for y^r_{T+1} and ε_{T+1}:

(15) $$\tilde{y}_{T+1} = X_{T+1}g_T$$

(16) $$\tilde{e}_{T+1} = y_{T+1} - \tilde{y}_{T+1} = X_{T+1}(\gamma - g_T) + \varepsilon_{T+1} \qquad T = 0,1,\ldots$$

Following Brown, Durbin and Evans (1975), I shall call the $\tilde{e}$ variables "recursive residuals" and shall adopt the term "recursive forecasts" for $\tilde{y}$. The variables defined in (15) and (16) result from running the regression of y on X using data from periods 1 through T, and forecasting one period ahead. This procedure is to be carried out for each period T. The recursive residuals and forecasts differ from the corresponding variables defined in (13), because the estimator of γ is recomputed each period, rather than just at S. The recursive forecast for T + 1 is constructed using g_T, the estimate of γ based on the sample 1,...,T.

The replications of the forecaster's falsifications tests undertaken in the next section will be based on the recursive forecasts and residuals. The latter must be transformed if tests designed for i.i.d. random variables are to be applied to them. In the case where the X_T's are fixed in repeated samples, it may be shown that $\tilde{e}^r_{T+1}$ has a zero mean and is not serially correlated at any lag. The variance is given by

$$\mathrm{Var}(\tilde{e}^r_{T+1}) = \sigma^2_\varepsilon A^2_T, \quad A^2_T = (1 + x_{T+1}(X'_T X_T)^{-1}x'_{T+1})$$

where x_{T+1} is the last row of X_{T+1} (cf. for example Brown, Durbin and Evans (1975)). The variance of $\tilde{e}_{T+1}$ thus differs from that of e_{T+1} by a correction that takes into account the fact that γ is estimated each period. The quantities

$$w_{T+1} = A_T^{-1}\tilde{e}_{T+1}^r$$

will be independently and identically distributed under the forecaster's null hypothesis, that (1) is the true model.[18]

The variables $\tilde{y}_t$ and $\tilde{e}_t$ which are used by the investigator, differ from the corresponding y_t^r and ε_t used by the forecaster because the former has to estimate the parameter of the model. The corresponding sequences of tests based on ${}_1\tilde{y}_T, {}_1\tilde{e}_T$ and ${}_1y_T^r$, ${}_1\varepsilon_T$ will also differ as T, the end date of the investigator's sample, increases. To see this, consider again the example (2), (3) where $\gamma^m = 3$. Thus, the empirical RE model which the investigator hypothesises to be used by the forecaster implies that $y_t^r = 3$ and ε_t is given by equation (9). As $T \to \infty$, the sample covariance between the elements of ${}_1y_T^r$ and ${}_1\varepsilon_T$ tends to 6, and so the forecaster would <u>continue</u> to reject the orthogonality hypothesis. However, as $T \to \infty$, the investigator's estimator $\hat{g}_T$ tends to 5, since it results from samples in which observations of the form $y_t = 5 + u_t$ dominate those of the form $y_t = 3 + u_t$, which are present only in the initial finite segment of the data. Hence, from (16), $\tilde{e}_t$ tends to $-2 + \varepsilon_T$, which is u_t from (9). This is orthogonal to the forecast (15), which tends to 5. Consequently, while the investigator's orthogonality and mean tests may evidence rejections early on, as $T \to \infty$ the test statistics will tend back to insignificance.

This example suggests that structural changes not accounted for by the investigator's model may be present when the orthogonality test statistics first rise and then fall as the sample size grows. It can also be shown that this test statistic will tend to zero if the model omits only non-stationary variables. Thus, if the test statistic evidences persistent rejections as the sample size becomes large, one must suspect that it may be the result of repeated structural changes or some other form of pertinent non-stationarity.

So far, the analysis has proceeded under the assumption that the forecaster knows the population projection coefficient, γ. This assumption is used widely in the RE literature, but there is no economic reason why the forecaster's information should include this knowledge. Some authors, for example Sheffrin (1979), have proposed that the forecaster estimates γ each period. Under this assumption, the recursive forecasts and residuals describe the behavior of the forecaster (assuming she adopted the model at $T = 0$). Consequently, the investigator's falsification tests are exactly those the forecaster would carry out.[19]

(ii) Adaptive expectations

When expectations are assumed to be adaptive, the investigator attributes to the forecaster expectations governed by the following equation, in which "y^a" represents the adaptive forecast:

$$y^a_{T+1} = \lambda y_T + (1 - \lambda) y^a_T \,, \quad 0 < \lambda < 1 \tag{17}$$

While the parameter λ may, in principle, vary over time, I consider only the case where the value of λ is fixed for the duration of the model's use.[20] If the fixed value of λ is also attributed to the forecaster, then the investigator can reconstruct exactly the time series of expectations and replicate exactly the falsification tests carried out at each date.[21] This is done in the next section for different values of γ and different starting dates.

4. Empirical Results

In this section, the falsification tests described above are applied to various proxies for expectations of the quarterly growth rates of the GNP deflator and the money supply (M1). Expectations of these variables have appeared in a large number of articles on inflation and on tests of the policy neutrality hypothesis.

In order to simulate the forecaster's falsification tests of a particular expectations mechanism, one first specifies its date of initial adoption and generates the time series of forecasts and errors to be used in the tests. For adaptive expectations, the parameter λ is fixed for the duration of the simulation run; for empirical RE models, the parameters are reestimated each period. For each period in the sample, the accumulated history of forecasting is then subjected to the falsification tests.

The output of a simulation is a quarterly time series of test statistics for each of the three tests. The patterns of these time series are of interest in their own right but are too voluminous to present in full. Instead, the charts that follow exhibit certain qualitative features of the behavior of each model. For each period, they illustrate whether the model passed a given test at the 5% significance level, failed at that level but passed at the 1% level, or failed at the 1% level (and hence at the 5% level as well). This permits one to address the question of how long a model would have survived as unfalsified in the hands of a rational person. Since the significance levels have not been derived from the optimization problem of the forecaster, one must be cautious about concluding that failure of a test will result in immediate rejection of the model. Rather, one would expect that the model would not be maintained in the face of persistent rejections. The charts also exhibit the length of the periods for which each model fails falsification tests.

(i) GNP deflator

Three different methods of forecasting the rate of change of the GNP deflator were examined. Four simulations were run on the adaptive expectations model, in which λ was set for the duration at 0.2, 0.4, 0.6, and 0.8 respectively. Two empirical RE specifications that have been used in tests of the neutrality hypothesis were also subjected to the testing procedure.[22] One is that used by Mishkin (1983), who forms his RE proxy by regressing the quarterly inflation rate of the GNP deflator on a constant and four lagged values each of the dependent variable, the Treasury bill rate, and the quarterly growth rate of M2. Mishkin chose these variables from a set of eleven macroeconomic series by an informal stepwise search procedure that seeks to include only variables whose coefficients are statistically significantly different from zero.[23]

The other RE method for forecasting inflation is derived from an equation used by Sargent (1976). He regresses the logarithm of the level of the GNP deflator on a constant, a time trend, three seasonal dummies and four lagged values each of the dependent variables, straight-time manufacturing wages, the labor force participation rate and the unemployment rate.[24] This delivers a forecast of the log of the price level, which was converted into an inflation forecast by subtracting the lagged value of the log of the deflator.

The summary results of the simulations, where 1951.I is the date for which the first forecast is made, are given in Chart 1.[25] The table shows

KEY TO CHARTS

: indicates that the test is passed at the 5% level

: indicates that the test is passed at the 1% level, but not at the 5% level

: (no line) indicates that the test is failed at the 1% level

A vertical bar indicates the date of the first test.

F : Orthogonality Test

Q : Serial Correlation Test

T : Mean Test

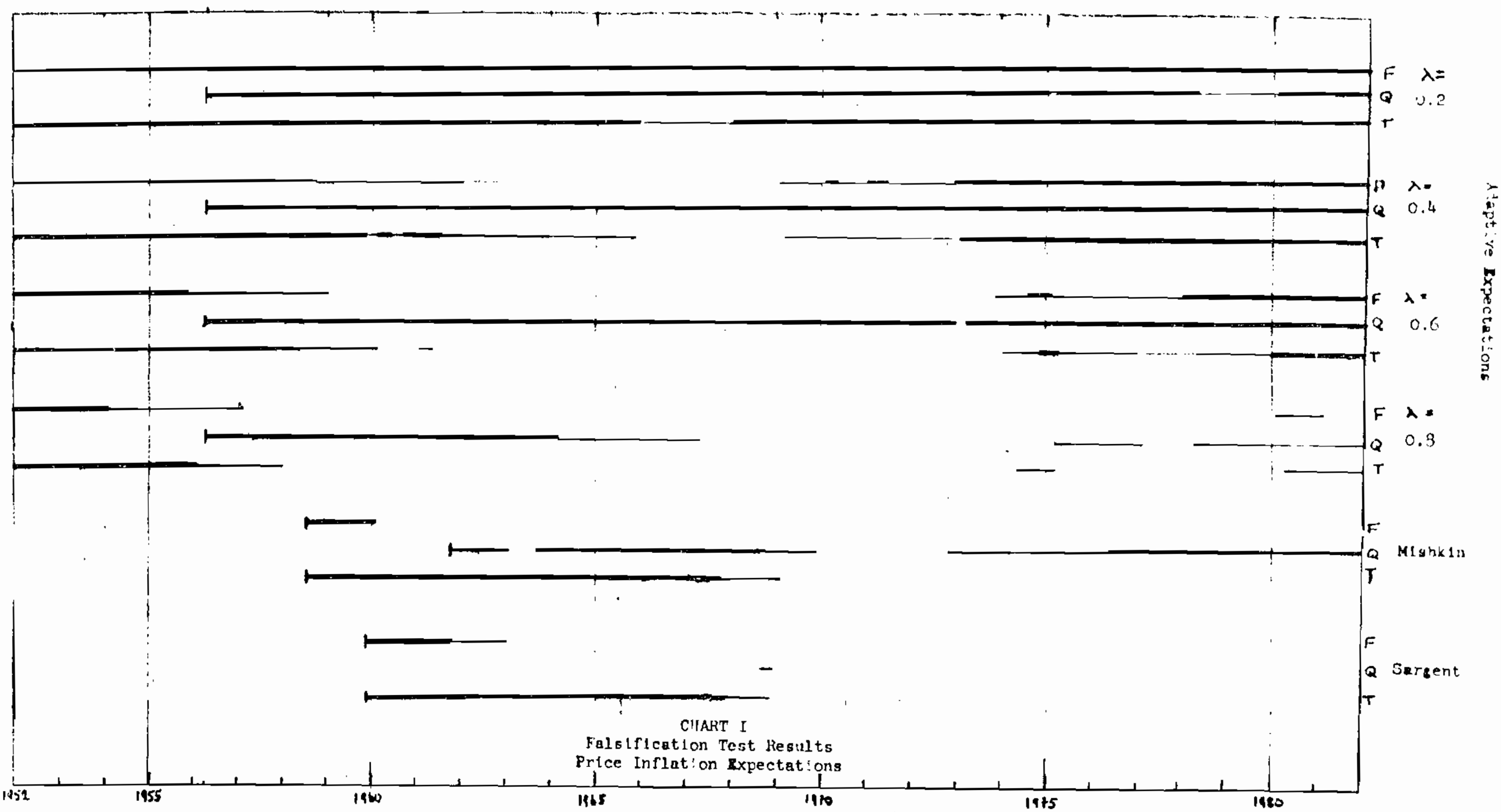

CHART I
Falsification Test Results
Price Inflation Expectations

that only one model passes every test at the 1% significance level, that is, adaptive expectations when $\lambda = 0.2$. At the 5% level, this model fails the serial correlation and mean tests for two brief periods: 1978.IV-1980.I in the first case, and 1966-67 in the second.

Generally, the performance of the adaptive models worsens as λ increases. The time paths of the test statistics are similar in shape. Their values are generally higher and their movements more pronounced for larger values of λ, although this may not be evident from the chart.[26] In most tests, the problem period is centered at about 1966, although its range expands dramatically as λ increases. For example, when $\lambda = 0.4$, rejections at the 1% level are first encountered in 1963.IV and last until 1969.I. The corresponding range is 1959.I-1974.I for $\lambda = 0.6$ and 1957.I through the late 1970's for $\lambda = 0.8$. From this, it may be concluded that if one were to consider modelling expectations from 1951 adaptively with, say, $\lambda = 0.4$, it would have to be assumed that the forecaster is prepared to ignore evidence of systematic errors for six years. Alternatively, one could model expectations as starting with this model in the 1950's and switching to some other model at or after the beginning of 1963. These remarks apply, mutatis mutandis, to the results for $\lambda = 0.6$ and $\lambda = 0.8$. On the other hand, the results support the proposition that, were the forecaster to commence forecasting using an adaptive model with the parameter λ set at 0.2, by the 1980's she would have accumulated scant evidence of the model's inadequacy. Again, other criteria may indicate that another model should be used in preference to this one. Furthermore, no reasons have been advanced explaining why the forecaster would adopt such a model in 1951.

The performance of Mishkin's and Sargent's proxies is considerably worse than that of the adaptive expectations models for which $\lambda \leq 0.6$. While the mean and orthogonality tests cross their critical values in the early 1960's, as they do for adaptive expectations, subsequently they rise gently on average. In contrast, the corresponding statistics for the adaptive models return to insignificance by the early 1970's. In view of the discussion of the last section, these results for the regression based models suggest that they fail to capture the non-stationarity that is apparently present in the series. Had the process been stationary, or experienced only one structural change, the continual updating of estimates should have caused the test statistics to tend to insignificance in large samples. The serial correlation statistic for Mishkin's model has a profile similar to those for adaptive expectations. However, that computed for Sargent's model is markedly different, largely remaining between 50 and 60, moving below the 1% critical value of 32.0 only in one quarter of 1968. In the face of these results, it is difficult to maintain the view that a rational forecaster would persist in using either of these models beyond the mid 1960's.[27]

The poor performance of the sophisticated RE models, and the contrasting results for adaptive models, are surprising in view of the criticism that proponents of the RE approach have levelled at ad hoc forecasting methods such as adaptive expectations (cf., for example, Sargent (1971)). One possible explanation of the results is that adaptive expectation in fact constitutes a "correct" model, in the sense that its errors are white noise. Eliminating the forecasts from (17) yields the following relationship between y_t and the forecast errors:

$$(18) \qquad y_t = y_{t-1} + \varepsilon_t - (1 - \lambda)\varepsilon_{t-1} .$$

If the errors are white noise, equation (17) says that y_t admits of an ARIMA(0,1,1) representation with parameter $(1 - \lambda)$.[28] In this case, adaptive expectations with the correct value of λ should pass the falsification tests. Two points are of interest here. First, examination of the autocorrelation and partial autocorrelation functions of the rate of change of the GNP deflator indeed suggests an ARIMA(0,1,1) model for the period. However, when this model is fitted to the data, the resulting estimate of $1 - \lambda$ is 0.41, with a standard error of .08. This suggests a 95% confidence interval for λ of $.59 \pm .16$. Values of λ in this range turned in mediocre or bad performances in the falsification tests. Second, the fact that the price variable admits of this representation does not explain why the empirical RE proxies fared so dismally. Both of these models include distributed lags of the dependent variable, and so may be regarded as low-order autoregressive approximations to (18).[29]

A second possibility is that the empirical RE proxies are "better" models, in the sense that they have a lower standard error than the adaptive models. However, because there is less noise in the empirical RE forecast errors, the systematic components are more easily detectable.[30] To examine this possibility, the squared one-step ahead forecast errors were cumulated each year for each model, starting in 1958.IV, the date of the first test of Sargent's model. The cumulated sums were 1.5 to 2 times larger at all dates for the empirical RE models than for the adaptive models. This indicates that "noise" in the adaptive expectations errors does not account for their performance.

A further anomalous finding is the date at which the first rejection of a model occurs. For the adaptive model (with the exception of $\lambda = 0.2$) rejections at the 5% level first appear in the mid 1950's, with those at the 1% level coming in the early 1960's. For the empirical RE proxies, the corresponding dates are 1960-1962.[31] It is not customary to associate difficulties in forecasting inflation with the pre-1968 era, when the inflation rate was in the range 0.9% - 3.4%. Rather, one would expect problems to occur during the 1970's, when the inflation rate increased steadily each year. Under these circumstances, adaptive expectations of inflation consistently underpredict. Hence, one would expect to see the adaptive models fail the orthogonality and mean tests. However, for $\lambda \leq 0.6$, these tests are passed at the 1% level from 1974 onwards. These results may occur because the 1970's are outweighed by the experience of the 1950's and 1960's. During those decades, the inflation rate appeared roughly to fluctuate around a constant mean. Such a process is adequately modelled (in terms of the mean forecast error) by adaptive expectations. While plausible, this attempt at explanation only compounds the problems of accounting for the behavior of the empirical RE proxies in the face of the changing nature of inflation. Again, one would expect them to perform no worse than adaptive expectations, to which they approximate. Furthermore, the regression coefficients are reestimated each period, which permits some accommodation of changes in the underlying parameters of the price process.

(ii) Money supply (M1)

In addition to adaptive expectations proxies, three other models of the growth of the money supply were considered. The results most favorable to the policy neutrality hypothesis came from the approach of Barro, whose examination of quarterly data is given in Barro and Rush (1980). Their forecasting equation regresses the quarterly growth rate of M1 on a constant, six lagged values of the dependent variable, three lagged values of log $(U/(1 - U))$, where U is the unemployment rate, and a variable that measures the contemporaneous ratio of real federal government expenditure to its "normal value."[32] Additional lags of these variables proved insignificant (Barro and Rush (1980, p. 33)).

A subset of their regressors is employed by Sheffrin (1979). On the basis of identification and estimation of a univariate ARIMA model over the period 1952.IV-1975.III, Sheffrin found the money growth process to be AR(2) with a non-zero mean. Consequently, his model contains a constant term and two lagged values of the money growth rate.

The third model of this type that is examined is used by Mishkin (1983). On the basis of the search method used for the GNP deflator equation, Mishkin regresses the money growth rate on a constant and four lagged values each of the dependent variable, the average Treasury bill rate, and the high employment federal budget surplus.

Because of the wide variation in the starting dates of the empirical studies from which the models are drawn, simulations were run starting with Barro's and Mishkin's initial dates (1941.I and 1954.I respectively) as well as 1948.III. The latter is the earliest date that permits the use of data series each constructed from only one source, rather than spliced together from a variety of sources. As the data required for Mishkin's model are not available prior to 1947.I, no simulation of this model could be run for the starting date of 1941.I.

The results for the earliest starting date are not illustrated, but their story is simple enough to recount. In general, all the models fail the orthogonality and serial correlation tests by the end of the 1960's. Having crossed the critical values, they rise monotonically throughout the rest of the period. Performance in the mean test is markedly different. Barro's model and adaptive expectations for $0.4 \leq \lambda \leq 0.8$ pass the test at the 1% level for the entire period. Sheffrin's and the remaining adaptive expectations model fall marginally during the 1950's and 1960's, but subsequently return to insignificance. Notwithstanding this result, it is apparent that none of the models examined would have survived in the hands of a rational forecaster, were it adopted in 1941.

The performance of the estimated models is considerably improved when the starting date is set at 1948.III (see Chart II). In particular, Sheffrin's model passes all tests at the 1% level, and problems develop only at the 5% level in the late 1970's. The test statistics for this model are everywhere markedly lower than for Barro's and Mishkin's models. Of these two, Barro's appears to fare the best. Even at the 1% level, however, this model fails at least one test during the periods 1960-65 and 1968-74, leading one to doubt that it would have been retained for forecasting for the entire period. On the other hand, it returns to insignificance at the end of the period. The major problem with Mishkin's model is that it fails the orthogonality test for the entire period.

The results for the adaptive models with $\lambda = 0.4$ and $\lambda = 0.8$, which are representative, are also shown in Chart II. As λ increases, performance in the orthogonality test generally deteriorates, while the results for the serial correlation test improve. By 1967 all models have experienced a rejection in at least one test at the 1% level. Beyond this date, therefore, they cannot be considered legitimate proxies for rational expectations of the money growth rate.

The above pattern of test statistics for adaptive expectations becomes more pronounced when the starting date is moved to 1954.I: the F-statistics for the orthogonality test generally fall, while the Q-statistics are generally higher than for the earlier date. Since, for $\lambda = 0.6$, the serial correlation test is soundly rejected even at the 1% level subsequent to 1960,

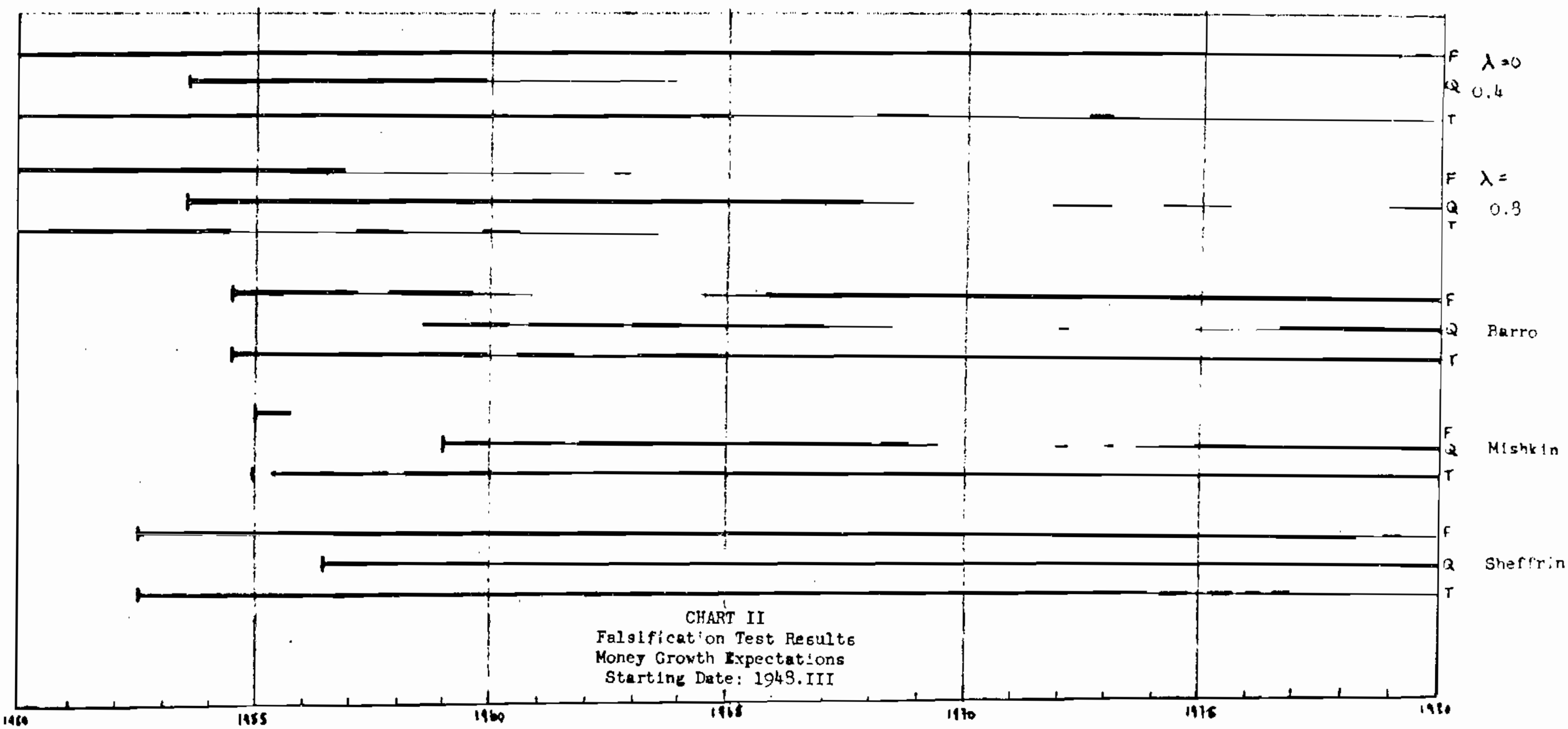

CHART II
Falsification Test Results
Money Growth Expectations
Starting Date: 1948.III

these models must again be dismissed as legitimate measures of rational expectations. The model in which $\lambda = 0.8$ survives the serial correlation test until 1968.

The results for the empirical RE models for the starting date of 1954.I are given in Chart III. In comparison with Chart II, the performance of Mishkin's model improves, while that of Sheffrin's deteriorates. Barro's model performs better in the orthogonality tests, but considerably worse in the serial correlation tests. Their results are similar to those for the $\lambda = 0.8$ adaptive model. It is difficult to judge which model is the best, but this is not really at issue. Each model experiences prolonged problems at least one test at the 1% level. Hence, none of them can be regarded as an adequate proxy for rational expectations of the money growth rate.

Again, it is difficult to reconcile these results with the "stylized facts" of the money growth data over the postwar period. For the period 1948.I-1979.IV, the series appeared to be characterized adequately by an AR(1) process with coefficient of 0.68 and a mean of 0.41. The adaptive expectations forecast and error are then both stationary ARMA(2,1) processes, whose white noise error is the one from the univariate representation of the money growth process.[33] In this representation, the forecast error has a zero mean, which may explain performance in the mean test for $\lambda \leq 0.4$, but not for larger values of λ. When $0 < \lambda < 1$ and the autoregressive coefficient is 0.68, the forecast error cannot be made to approximate white noise. Thus, one would expect the adaptive expectations models to fail the serial correlation test. This is so, except for the case $\lambda = 0.8$, whose performance in this test is similar to that of the empirical RE models. The correlation between the contemporaneous forecast and forecast error should be large if the money growth process is AR(1). Consequently, the orthogonality test should exhibit

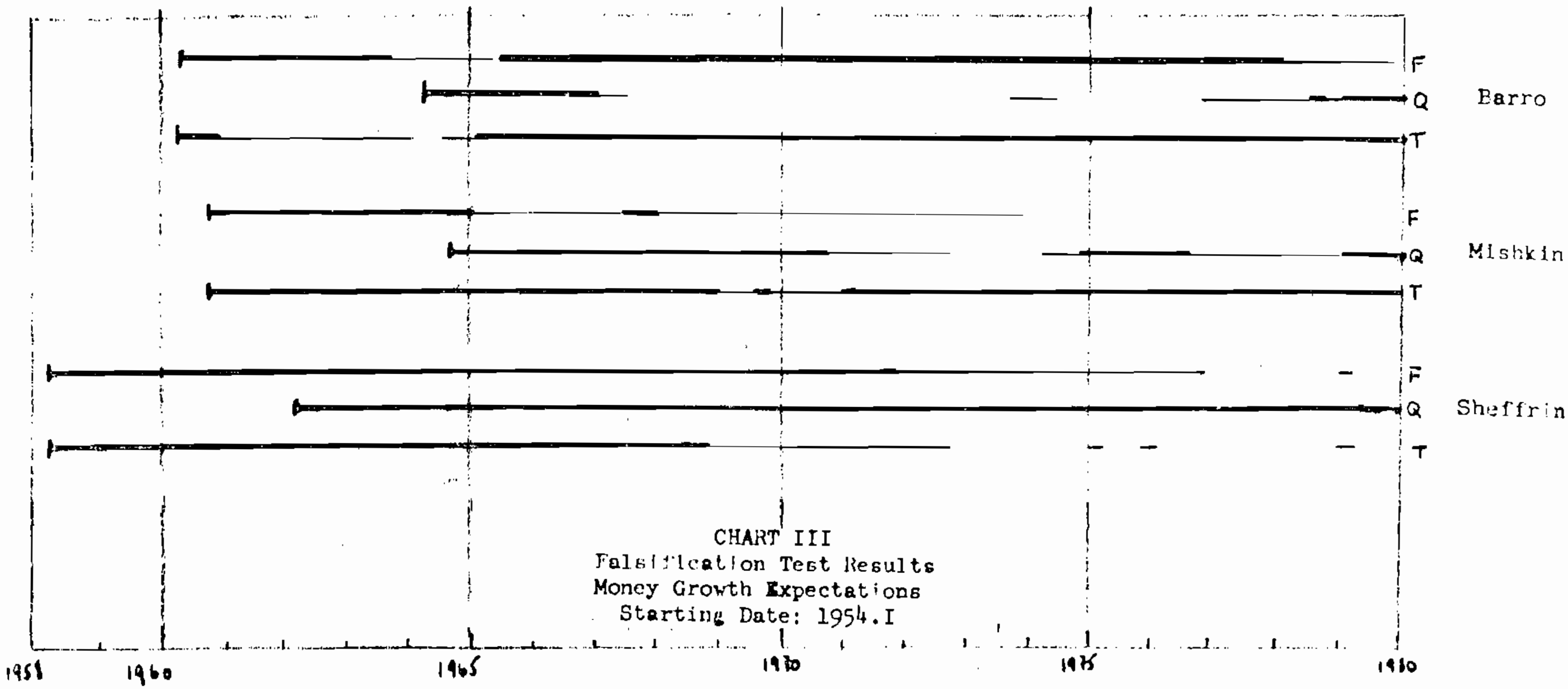

CHART III
Falsification Test Results
Money Growth Expectations
Starting Date: 1954.I

rejections. However, these occur only for larger values of λ.

The empirical RE models exhibit a puzzle similar to those experienced in the case of the GNP deflator. Barro's and Mishkin's models include the univariate model as a special case, and yet they fail the falsification tests summarily.[34] For the two later starting dates, each of the empirical RE models experiences problems with a particular test. For Barro's model it is the serial correlation test. Orthogonality is a problem for Mishkin's model, while Sheffrin's fits the mean test for the sample commencing in 1954.I. Barro's model evidences rejections of these tests in the early 1960's and subsequently returns to insignificance. Failure of the orthogonality and mean test is taken as evidence of non-stationarity in the process. The results suggest that there may have been a discrete shift in the money growth process in the early 1960's. However, the principal increase in the money growth rate took place in the 1970's. Rejections of the orthogonality and mean tests do occur in this decade for Mishkin's and Sheffrin's models. However, the results of the orthogonality test for Mishkin's model also suggest that the change in the process may have occurred earlier. Clearly, these results suggest that money growth is not adequately modelled by a process with unchanging coefficients.

5. Conclusion

The tests carried out above yield information on two closely related properties of proposed measures of expectations. First, if the expectations series fails the tests, there is prima facie evidence that it does not come from the true model. Second, since the test results that yield this information are computable on the basis of the historical data assumed available to economic agents, rationality of their expectations, and *a fortiori* RE, would demand that they not persist in using the model. The implications of the results of these tests for the rational expectations approach are the subject of this section.

The application of the testing framework to proposed empirical RE proxies indicated that they did not satisfy the necessary conditions, as the one-step ahead prediction errors did not exhibit the properties of white noise variables. None of these models passes the serial correlation tests for the duration of the periods examined and, in particular, they are frequently found not to be orthogonal to the predictions. Thus, the orthogonality of predictions and errors of empirical RE models over the sample period of their estimation must be regarded as an artifact of the ordinary least squares procedure used in the literature to construct them and can maintain no claim to reflect the validity of the assumed model. The advantage of using recursive residuals is that they permit one to examine the prediction performance of a model during the sample period over which it has been estimated in empirical RE studies. Thus unconstrained, the models reveal that they would have been rejected by rational agents *during* these sample periods. For example, most empirical RE models experience problems during the 1960's.

One interpretation of this result is that the empirical RE models measure rational expectations (in the sense of use of the true model) with an error. The falsification tests suggest that, in general, this error will be serially correlated. Frydman and Rappoport (1984) examine the effects of this mismeasurement on inferences concerning three central tenets of new classical macroeconomics. These are the hypotheses that anticipated monetary policy is neutral and that expectations are rational in the sense of RE, and Lucas' proposition relating the variance of nominal demand to the slope of the observed "output-inflation tradeoff." The particular type of measurement error found here in empirical RE models is shown to render crucial parameter estimators inconsistent. This, in turn, implies that the test statistics derived from them are uninterpretable.

In contrast to the results for the large empirical RE models, for each of the two series examined, there is a simple model that passes all tests.[35] In the case of the GNP deflator, this is the adaptive expectations model with $\lambda = 0.2$. Sheffrin's autoregressive proxy of the M1 growth rate passes all tests when the starting date is 1948.III, although it fails orthogonality and mean tests in the 1970's, when the starting date is 1954.I. These results may suggest that a low-order time series process correctly measures RE. In order to be tractable, such a model would involve parameters that are constant over time. However, if the true model has this property it becomes impossible to comprehend the large sample failures of the orthogonality and mean tests experienced by empirical RE models. The continual updating of parameter estimates involved in computing recursive residuals should cause the orthogonality and mean tests to be passed when the sample size is large. In addition, it is not possible to reconcile the performance of the empirical RE models with that of the simple ones examined above. If the parameters are constant, the models with more regressors should perform no worse, except possibly in small samples. Thus, the results of this paper should not be interpreted as advocating particular parsimonious models as proxies for rational expectations, even though these models are not diagnosed inadequate on the basis of the available information. Rather, the results suggest that the large models do not improve upon the simple ones.

It becomes feasible to understand these results only if non-stationarity of the processes in question is entertained. In particular, one would expect the money growth process to change over time as the Fed changes the direction and priorities of its monetary policy. The most obvious piece of informal evidence on this point is that the mean of the money growth rate appears to have shifted substantially over the postwar period. In the 1950's it was approximately 2.4% per annum, while in the following decades it averaged 3.87% and 6.6% respectively.

Thus far, the falsification framework has been employed as a tool for diagnosing the lifespan of a forecasting model in the hands of a rational forecaster. The falsification tests attributed to the forecaster can also be regarded as an aspect of her attempts to learn the true model. This learning process would involve the forecaster adopting a particular model at a point in time, and using it to forecast until it failed the falsification tests. Then these steps would be repeated with another model. The expectations of a rational agent over a period of historical time would thus be characterized by the predictions of a series of different models. At any point when a particular model was used, it would not have evidenced systematic mistakes. Transitions between models would reflect that the old model had experienced systematic errors.

Such a process is often discussed briefly and informally as the motivation for modelling expectations by RE. Thus, for example, Begg (1983) and Cyert and DeGroot (1974) both suggest that a rational agent will not persist in using a false model, due to the nature of its forecast errors. They conclude from this that the correct way to model rational expectations is by use of the true model. Implicitly, they invoke the assumption that agents will learn the specification of the true model. In order to support this conclusion, a formal model of learning is required.

The account of learning behavior normally employed in economics involves the specification of a model with parameters whose numerical values are unknown by the agent. With the accretion of information, the agent updates his or her prior distribution on the parameters. Under certain regularity conditions (DeGroot (1970, Ch. 10)), and if the true model is contained in the support of the prior, the posterior distribution collapses on the parameter values of the true model as the sample grows. Some models of this type suggest that agents will learn the true model over time.[36]

The crucial assumption in this account is that the true model is a special case of the set of possible models entertained at the outset by the agent. Under this circumstance, learning is, in effect, modelled as a process of elimination. The framework of this paper suggests another important aspect of the process of learning, which must be formalized. The forecaster must first discover the model with which she is endowed in the Bayesian learning studies mentioned above. In general, one must assume that the forecaster commences with a "false" model (in the sense that it does not contain the true one as a special case), whose inadequacy is eventually revealed by falsification tests. The rules which the rational forecaster would follow to select a new model must now be specified. The forecaster has information only on the inadequacy of the old model. This does not suggest which model it will be fruitful to use in the future. Once a new model is adopted, the forecaster may then act rationally with respect to that model.[37] But no theory of the <u>rational</u> way to discover a new model is apparent.

This hiatus of rational behavior in the transition between models has precedents both in the methodology of science and in economic theory. The economic theory of the firm shows how it can maximize profits subject to a given technology. It does not provide an account of how the rational firm may best discover a new technology. Popper's work on scientific methodology advocates that the current theory be subjected to attempts to falsify its predictions. It does not suggest how a new theory may be developed. Popper does not encourage one to believe that a theory of the rational discovery of models will be possible. He expresses the view that "there is no such thing as a logical method of having new ideas . . . every discovery contains 'an irrational element' or 'a creative intuition'" (1932, p. 32). A complete model of learning thus involves some behavior that cannot be readily explained by resorting to rationality. The forecaster's choice of a model is in part arbitrary.

It is apparent that one cannot conclude that rational people will learn the true model over time. One is at liberty to assume that people know the true model, but this assumption cannot be defended by appealing to rational behavior. As a consequence, rationality alone cannot account for the way people form their expectations. This is determined in part by their "choice" of a model (i.e., by the models they have discovered), which lies outside the purview of rational behavior. To define what constitutes rational expectations at any date is thus a tricky and difficult matter, because it involves an assumption about the models known to agents. Firm statements can be made only about those models that are ruled out because they have been falsified.

Bibliography

Amemiya, Takeshi [1981] "Selection of Regressors," International Economic Review, Vol. 21, no. 2, pp. 331-354.

Barro, Robert J. [1977] "Unanticipated Money Growth and Unemployment in the United States," American Economic Review, Vol. 67, no. 1, pp. 101-115.

Barro, Robert J. and Mark Rush [1980] "Unanticipated Money and Economic Activity," in Stanley Fischer (ed.), Rational Expectations and Economic Policy, Chicago: University of Chicago Press.

Box, George E. P. and D. A. Pierce [1970] "Distribution of Residual Autocorrelations in Autoregressive-Integrated-Moving Average Time Series Models," Journal of the American Statistical Association, Vol. 65, Dec., pp. 1509-1526.

Bray, Margaret, Lawrence Blume and David Easley [1982] "Introduction to the Stability of Rational Expectations Equilibrium," Journal of Economic Theory, Vol. 26, no. 2, April, pp. 313-317.

Brown, R. L., J. Durbin and J. M. Evans [1975] "Techniques for Testing the Constancy of Regression Relationships Over Time," Journal of the Royal Statistical Society, Series B, no. 2, pp. 149-192.

Cyert, Richard M. and Morris H. DeGroot [1974] "Rational Expectations and Bayesian Analysis," Journal of Political Economy, Vol. 82, no. 3, pp. 521-536.

Davies, Neville and Paul Newbold [1979] "Some Power Studies of a Portmanteau Test of Time Series Model Specifications," Biometrika, Vol. 66, no. 1, pp. 153-155.

Davies, Neville, C. M. Triggs and Paul Newbold [1977], "Significance Levels of the Box-Pierce Portmanteau Statistic in Finite Samples," Biometrika, Vol. 64, no. 3, pp. 517-522.

DeGroot, Morris H. [1970] Optimal Statistical Decisions, New York: McGraw-Hill.

Dhrymes, Phoebus J. et al. [1972] "Criteria for Evaluation of Econometric Models," Annals of Economic and Social Measurement, Vol. 1, no. 3, pp. 291-324.

Friedman, Benjamin M. [1979] "Optimal Expectations and the Extreme Information Assumptions of 'Rational Expectations' Macromodels," Journal of Monetary Economics, Vol. 5, no. 1, pp. 23-41.

Friedman, Milton and Anna J. Schwartz [1970] Monetary Statistics of the United States, Studies in Business Cycles, no. 20, New York: National Bureau of Economic Research.

Frydman, Roman and Edmund Phelps [1983] "Individual Expectations and Aggregate Outcomes: An Introduction to the Problem," in Frydman and Phelps (eds.), Individual Forecasting and Aggregate Outcomes: "Rational Expectations" Examined, New York: Cambridge University Press.

Frydman, Roman and Peter Rappoport [1984] "An Examination of Econometric Tests of the Central Propositions of the New Classical Macroeconomics," New York University, Department of Economics, mimeo.

Fuller, Wayne A. [1976] Introduction to Statistical Time Series, New York: Wiley.

Gordon, Robert J. [1982] "Price Intertia and Policy Ineffectiveness in the United States," Journal of Political Economy, Vol. 90, no. 1, pp. 22-51.

Leamer, Edward [1978] Specification Searches, New York: Wiley.

Leiderman, Leonardo [1980] "Macroeconometric Testing of the Rational Expectations and Structural Neutrality Hypotheses for the United States," Journal of Monetary Economics, Vol. 6, pp. 69-82.

Ljung, G. M. and G. E. P. Box [1978] "On a Measure of Lack of Fit in Time Series Models," Biometrika, Vol. 65, no. 2, pp. 297-303.

Meltzer, Allan [1981] "Rational Expectations, Risk, Uncertainty and Market Responses," paper prepared for the New York University Conference on Crises in the Economic and Financial Structure, mimeo.

Mishkin, Frederic H. [1983] A Rational Expectations Approach to Macroeconomics, Chicago: University of Chicago Press.

Mullineaux, Donald [1978] "On Testing Rationality: Another Look at the Livingston Price Expectations Data," Journal of Political Economy, Vol. 86, pp. 329-336.

Muth, John F. [1960] "Optimal Properties of Exponentially Weighted Forecasts," Journal of the American Statistical Association, June, pp. 299-306.

——— [1961] "Rational Expectations and the Theory of Price Movements," Econometrica, Vol. 39, no. 3, pp. 315-335.

Pesando, James E. [1975] "A Note on the Rationality of the Livingston Price Expectations," Journal of Political Economy, Vol. 83, no. 4, pp. 849-858.

Popper, Karl [1959] The Logic of Scientific Discovery, London: Hutchinson.

Rappoport, Peter [1983] "Unfalsified Expectations: A New Approach to Modelling Expectations in Macroeconomics," New York University, Department of Economics, mimeo.

Sargent, Thomas J. [1971] "A Note on the 'Accelerationist' Controversy," Journal of Money, Credit and Banking, pp. 721-725.

——— [1973] "Rational Expectations, the Real Rate of Interest, and the Natural Rate of Unemployment," Brookings Papers on Economic Activity, no. 2, pp. 429-472; errata, no. 3, pp. 799-800.

——— [1976] "A Classical Macroeconometric Model for the United States," Journal of Political Economy, Vol. 84, no. 2, pp. 207-237.

Sheffrin, Steven M. [1979] "Unanticipated Money Growth and Output Fluctuations," Economic Inquiry, Vol. 17, pp. 1-13.

Shiller, Robert J. [1978] "Rational Expectations and the Dynamic Structure of Macromodels: A Critical Review," Journal of Monetary Economics, Vol. 4, no. 1, pp. 1-44.

Simon, Herbert A. [1957] Models of Man, New York: Wiley.

Turnovsky, Stephen J. [1970] "Empirical Evidence on the Formation of Price Expectations," Journal of the American Statistical Association, Dec., pp. 1141-1456.

Wetherill, G. Barrie [1975] Sequential Methods in Statistics, London: Chapman and Hall (2nd ed.).

Data Sources

The data used to reconstruct the empirical RE proxies followed the instructions given in the relevant articles. Unless otherwise stated, all data were taken from the Citibase Economic Data Base. Sargent (1976) describes the data sources and construction of his equation in footnotes 15 and 22 of his article. Mishkin's data are described on p.115 of his article. The series employed for M1 for all except Barro's model are from the 1980 revision, as stored on Citibase. Barro and Rush (1980) provide the series that they used in Table 2.3 on pp. 40-46 of their article, and this series is used as the dependent variable in falsification tests of their model. (This series is based on an earlier revision.) The 1966.III and 1966.IV observations appeared to have been reversed in the table; these were corrected in the empirical work in this paper. Simulations of Barro's model on the data from the 1980 revision did not yield qualitatively different results. In order to construct the real federal government expenditure variable, it was necessary to find a measure for the quarterly (GNP) price deflator for the years 1941-1946 inclusive. This was taken from the data appendix to Gordon (1982), supplied by the author. The interpolation procedure used is described on p.1114 of Gordon's article.

Data for the adaptive expectations proxies after 1947.I was taken from Citibase. In order to construct the series for money growth expectations from 1941.I, forty initial values starting in 1931.I were used. This data was taken, following Barro and Rush, from Friedman and Schwartz (1970, Table 2).

Footnotes

1. An early criticism of RE was that it endowed agents with too much information. This is voiced, for example, by Shiller (1978), B. Friedman (1979, p. 26) and Meltzer (1981, p. 3). Frydman and Phelps (1983) provide a more detailed discussion of the relationship between the optimization problems dealt with in microeconomics and the concept of rationality implicit in RE.

2. The elements of such a procedure are often outlined in informal discussions that seek to justify modelling rational expectations as use of the true model. Cf., for example, Begg (1983, p. 61), and Cyert and DeGroot (1974, p. 524).

3. Model selection criteria based on the expected sum of squared forecast errors are discussed in Amemiya (1980).

4. Falsification tests are obviously not the only vehicle for testing the adequacy of a model. The investigator may also use the entire data sample available to carry out Chow tests, or to examine the serial correlation of the residuals from a regression using this sample. These approaches are complementary to the one used in this paper, which has the added interpretation of constituting a description of a part of the agent's learning process.

5. Another casualty has been the use of reported forecasts, such as the Livingston data. A number of authors have tested this series for rationality in Muth's sense, yielding mixed but largely negative results. Cf., for example, Turnovsky (1970), Pesando (1975), Mullineaux (1978). It does not follow from these results that Livingston's forecasters persisted in using a model that had been falsified. In order to demonstrate irrationality in this sense, it is necessary to show that, after a certain date, it was possible to discern that the forecast error had a particular systematic structure and

that, subsequently, this systematic component persisted in the errors. This would show that the systematic components were discoverable on the basis of available information, but not exploited.

6. The problem considered is thus that of the representative economic agent modelled in econometric investigations of RE.

7. The population projection coefficient is the probability limit of the coefficient in the least squares regression of y on X, and the population projection is the corresponding fitted value.

8. For example, Sargent and Wallace (1976, p. 177) model the RE of the log of the money supply (m_t) by assuming that the monetary authority uses the feedback rule $m_t = g\theta_{t-1} + \eta_t$, where η_t is serially uncorrelated and uncorrelated with θ_{t-1}, "a set of observations on variables dated t - 1 and earlier." θ_{t-1} is assumed to be the information set held by the representative agent, whose (rational) forecast is consequently modelled by the population projection of m_t on θ_{t-1}, that is, $g\theta_{t-1}$. Hence, the assumptions of the analysis guarantee that the forecast error will be white noise. If necessary, the white noise property of η_t can be assured by including its lagged values in θ_{t-1}.

9. It should be noted that the condition that the errors are white noise is necessary but not sufficient for them to be devoid of systematic components. For example, in the case of stationary stochastic processes, the univariate Wold representation has white noise errors, but the variance of these is larger than the variance of the white noise errors that result from a higher dimensional Wold representation. The reduction in variance comes from the extra explanatory power of the variables in the multivariate representation. Frydman and Rappoport (1984) provide an example.

10. In this paper, it will be convenient to adopt the following norational conventions: Z_j stands for the value of Z at date j, and ${}_{i}Z_{j}$ stands for the vector of observations on Z dated i through j inclusive. Thus, ${}_{i}Z_{j} = (Z_i, \ldots, Z_j)$.

11. The correct degrees of freedom for this distribution depend on the number of parameters used to estimate the ARMA model for ε_t^m (if any) contained in equation (1). Full details are provided by Box and Jenkins (1970, p. 394).

12. Ljung and Box (1978) demonstrate that, with a 5% significance level, the power of the Q-statistic can be well below 50% in samples of 100, unless the true and estimated models differ greatly. Davies and Newbold (1979) come to a similar conclusion after examining a wider class of underlying models to which an autoregressive specification is fit. They also find a marked improvement for samples of 200. Ljung and Box and Davies, Triggs and Newbold (1977) also indicate that the χ^2 approximation understates the variance of Q when the null hypothesis is true. When $M = 20$, the magnitude of the understatement is 35% for samples of 100, and 17.5% for samples of 200. Thus, the Q-statistic will reject the null hypothesis with greater frequency than the chosen level of significance.

13. It should also be noted that the orthogonality test described above will not be successful in detecting all kinds of parameter shifts. For instance, if in (2) the intercept were to <u>alternate</u> between 3 and 5 successive periods, the orthogonality test would have zero power in samples of even size and asymptotically zero power in samples with an odd number of observations.

14. For the cases where the test statistic is the mean forecast error or the mean squared forecast error, Brown, Durbin and Evans (1975) develop procedures for a finite number of observations that have this property.

15. A discussion of the problems of sequential testing of point against composite hypotheses is given in Wetherill (1975, Ch. 4). The problem is further compounded in the current case by the need for a statistic to test the three hypotheses jointly at each point in time.

16. For more complete details, the reader is referred to the original working paper associated with this study (Rappoport, 1983).

17. The S in parentheses in expressions such as $y_T^r(S)$ signifies that the variable is constructed with the estimate g_S, based on the sample running from 1,...,S.

18. Here, is it necessary to consider the fact that the X's will generally include lagged dependent variables. Exact results on the distributions of $\tilde{e}^r$ cannot be derived easily, and the correction to be administered to the forecast errors will, in general, depend on the true value of γ. Following Fuller (1976, pp. 382-84), an asymptotic approximation may be made. The degree of approximation is the same as that used by Box and Pierce (1970, pp. 1512-15) in deriving the distribution oi the Q-statistic. In the present case, it amounts, in finite samples, to ignoring the error that arises from estimating γ, and yields the result that, asymptotically, $\tilde{e}_{T+1}^r$ and e_{T+1}^r have the same distribution. This suggests that no variance correction need be applied to the recursive residuals, in order to replicate the information in the underlying e's. (The reader is referred to Rappoport (1983) for a more complete account.) There does not appear to be any clear choice between using w_{T+1} and using $\tilde{e}_{T+1}^r$ in the case where the X variables include lagged values of the dependent variable. The procedure followed here

is to use the former except where the model contains only lagged values of the dependent variable. The two distributions computed above differ only in their variance terms. Since A_T^2 is of the order of $(1 + 1/T)$, this difference is negligible in large samples. However, one cannot infer from this that the outcome of the tests would be the same were the true distribution of e_{T+1}^r under the null hypothesis used. The results would depend on the covariances of the recursive residuals, which will be non-zero in general.

19. Brown, Durbin and Evans (1975) and Dufour (1982) have advocated using the recursive residuals to test for structural change in the standard regression context. This context describes the problem facing the forecaster when she is ignorant of γ. (See below.)

20. The case in which the forecaster estimates λ each period is not pursued here because of the computational complexity involved in the estimation of moving average models.

21. It is also necessary that an initial value of expectations be given in order that (9) can be applied recursively. The choice of an arbitrary initial value contaminates the values used for subsequent forecasts to an extent that varies inversely with λ. In order to avoid this problem, the approach taken here was to construct the expectations series by initializing it at a date prior to that of the first observation that would be used in calculating the test statistics (i.e., the data at which $T = 0$). In the following analysis, expectations were initialized at the earliest date of reliable data. In the case where least data were available, y_{-16}^a was set equal to y_{-16}. As a result, the error in the value of y_0^a constructed recursively from this initial value using (9) with $\lambda = 0.2$ is 1.2% of the difference between the postulated and the unknown "actual" value of y_{-16}^a.

22. The sources and construction of the data used are described in the Appendix.

23. Cf. Mishkin (1983, p. 155). His sample period is 1954.I-1976.IV.

24. Sargent's sample period is 1952.II-1973.III. A similar model, with more regressors, is used in Sargent (1973).

25. As these forecast and errors are generated from an estimated equation, it is necessary to accumulate some degrees of freedom. The procedure followed is to let data accumulate from the data for which T = 0. When T is sufficiently large to allow 10 degrees of freedom for the estimation of the equation that generates the RE proxy, the equation is estimated and the first error generated. Thus, for example, Sargent's model contains 21 regressors, so the first forecast does not occur until T = 32, i.e., 1958.IV. This generates one observation for the orthogonality and mean tests. Three degrees of freedom are permitted to accumulate before the first orthogonality test is run. Since this test involves the estimation of two parameters, five observations are required, and the first orthogonality test is thus run in 1959.IV. Since Mishkin's model has only 13 regressors, the date of the first orthogonality test is 1957.IV. The serial correlation test is run on 20 sample autocorrelations. Thus, at least 20 are observations on forecast are errors required. Hence, for Sargent's model, the date of the first autocorrelation test is 1963.IV, and for Mishkin's 1961.IV. The adaptive proxies do not require any parameters to be estimated. Hence, the first expectation and error will be dated 1951.I, and the first orthogonality test will occur in 1952.I. Similarly, the first autocorrelation test will be run when 21 observations have accumulated, that is, in 1955.II. Thus, at any given date, the tests run on adaptive proxies examine longer series of data than those run on the empirical RE proxies.

26. For example, in the case of the orthogonality test, each model attains its maximum value in 1967. In ascending order of λ, these maxima are 3.1, 6.4, 10.5, and 15.7. At this date, the 5% critical value of the F distribution is 3.14, and the 1% value is 4.95.

27. It should be stressed that, whereas the first test of the adaptive models is carried out in 1952, the regression-based models are not subjected to falsification tests until 1958 (Mishkin) or 1959 (Sargent). This is because the data of the 1950's is used to estimate the parameters of the model. The forecasts from these years would thus be of the "within sample" variety discussed in the last section, which are not amenable to falsification tests.

28. This property of adaptive expectations was first noted by Muth (1960).

29. For example, Mishkin's model includes four lags of the inflation rate. The coefficient of the fifth lag (obtained by inverting the polynomial in (16)) is $-(1-\lambda)^4 + (1-\lambda)^5$. For the estimated ARIMA model ($\lambda = 0.412$), the implied value of this coefficient is $-.015$.

30. I am grateful to an anonymous referee for this suggestion.

31. Again, the fact that, for these models, tests do not begin until the late 1950's should be borne in mind.

32. The sample period used by Barro and Rush is 1941.I-1978.I.

33. Say $y_t = \delta + \beta y_{t-1} + u_t$, where y is the money growth rate and u_t is white noise. Then $y^a = (1-\beta L)^{-1}(1-(1-\lambda)L)^{-1}\lambda L(\delta + u_t)$ and $e_t^a = (1-\beta L)^{-1}(1-(1-\lambda)L)^{-1}(1-L)u_t$.

34. Again, this may occur because the complicated models may have lower error variances that make systematic components easier to detect. As with the GNP deflator, the cumulated sums of squared forecast errors were examined for each model, for the 1948.I starting date. Sheffrin's model always performed better than Mishkin's according to this criterion. (Mishkin's model also

always exhibited a higher cumulated sum than the $\lambda = 0.8$ adaptive model.) Barro's model bettered Sheffrin's only in the last three quarters of 1979. In general, its performance was very erratic. Its cumulated sum of squared errors was higher than that of the $\lambda = 0.8$ adaptive model for more than half of the dates in the sample. Thus, examination of a crude alternative statistic does not lead one to qualify the results of the falsification tests.

35. The dominance of these models over proxies that include more regressors is perhaps surprising. The choice of right-hand side variables for empirical RE proxies was often motivated by informal specification searches based on long samples. For example, Mishkin (1983, p. 21) rationalized his model specification procedure in terms of Granger causality, i.e., he used a vector autoregressive model. The list of regressors in empirical RE studies is often chosen by appeal to F-tests of the hypothesis that the coefficients of a set of variables are equal to zero (cf. Mishkin (1983), Barro (1977)). It may be thought that these tests may result in the inclusion of more regressors than would be warranted by a loss function based on prediction performance. Amemiya (1980) derives model selection criteria designed to minimize the expected squared prediction error, which penalize less parsimonious models. He shows that these rule in favor of the less parsimonious model when the F-statistic is larger than 2. For two-tailed t-tests of single coefficients, the 5% critical value is above 1.96 (irrespective of degrees of freedom) implying an F-value of 3.84, and, for an F-test of a zero restriction on a group of four coefficients, the corresponding critical value is no less than 2.37. Thus, if the procedures used by Mishkin and Barro favor the larger model, then so do the model selection criteria.

36. A survey of these models is given by Blume, Bray and Easley (1982).

37. This account is reminiscent of Simon's concept of "bounded rationality." The complexity of the problems facing the decision-maker is beyond his or her psychological capacity. The decision-maker is thus compelled "to construct a simplified model of the real situation in order to deal with it. He behaves rationally with respect to this model, and such behavior is not even approximately rational with respect to the real world" (1957, p. 199). Unlike Simon's treatment, the discussion of learning behavior here does not preclude the agent from ever learning the true model due to its complexity, although it obviously suggests that the barriers to successful completion of this endeavor may lead one to Simon's position.