Volume Title: Annals of Economic and Social Measurement, Volume 3, number 3

Volume Author/Editor: Sanford V. Berg, editor

Volume Publisher: NBER

Volume URL: http://www.nber.org/books/aesm74-3

Publication Date: July 1974

Chapter Title: Tests for Structural Change and Prediction Intervals for the Reduced Forms of Two Structural Models of the US: The FRB-MIT and Michigan Quarterly Models

Chapter Author: T. Muench, A. Rolnick, N. Wallace, W. Weiler

Chapter URL: http://www.nber.org/chapters/c10174 Chapter

pages in book: (p. 491 - 519)

Annals of Economic and Social Measurement, 3/3, 1974

TESTS FOR STRUCTURAL CHANGE AND PREDICTION INTERVALS FOR THE REDUCED FORMS OF TWO STRUCTURAL MODELS OF THE U.S.: THE FRB-MIT AND MICHIGAN QUARTERLY MODELS

BY T. MUENCH, A. ROLNICK, N. WALLACE AND W. WEILER*

Prediction interval tests are applied to the reduced forms of two quarterly models of the U.S. (the "old" FRB-MIT model and the Michigan model). The results illustrate the range of tests one can perform on an estimated simultaneous equation model. In particular, the tests determine whether ex post forecast errors can be attributed to structural deficiencies of the models. The paper examines confidence regions and other aspects of forecast distributions—comparisons between mean forecasts and nonstochastic forecasts, comparisons between forecast variances from multiperiod endogenous simulations and those from one-period simulations, and comparisons between forecast variances and residual variances.

I. INTRODUCTION

In this paper we report the results of statistical tests for a variety of structural change in the coefficients of two quarterly models of the U.S. economy: the "old" FRB-MIT model and the Michigan model.[1] We test for structural change between two periods, the period over which each model was originally estimated and a post-sample period. Because the latter is very short, our tests reduce to prediction interval tests, analogous to tests for structural change in the coefficients of a single equation model when one of the comparison periods is short.

As far as we know, prediction interval tests have not previously been applied to the reduced form of a simultaneous equations model, let alone to that of a large nonlinear model. There have been studies in which differences between actual outcomes and what we call ex post nonstochastic (reduced-form) forecasts (forecasts generated from the point estimates of all parameters) have been compared across models including a variety of "naive" models, but those comparisons cannot offer statistical grounds for acceptance or rejection of a model. In contrast, the tests we perform determine in a probabilistic sense whether the magnitudes of ex post forecast errors can be attributed entirely to randomness in the economy and to uncertainty stemming from the size of the data set, or, must in part be attributed to structural deficiencies of the model, where structure includes a stochastic specification consistent with the particular estimation procedure employed.

The paper is organized as follows. In section II, we give a brief description of the models we test and describe the class of test statistics we use. Our grounds for

* T. Muench and N. Wallace are staff members of the Economics Department, the University of Minnesota, and Consultants to the Federal Reserve Bank of Minneapolis. A. Rolnick is an Economist at the Federal Reserve Bank of Minneapolis; and W. Weiler is Assistant Director, Information Services Division, the University of Minnesota. All views expressed herein are the sole responsibility of the authors and should not be interpreted as representing those of the Federal Reserve Bank of Minneapolis which provided financial support.

[1] The Michigan model is described in [6]. The version of the FRB-MIT model we test has not been published. Versions much like it are described in [2] and [3].

employing these statistics and our associated distribution assumptions are presented in Appendix I. The subsequent three sections are devoted to a presentation of results: section III to basic test results; section IV to aspects of the confidence regions and to tests on linear functions of the variables; and section V to other aspects of the forecast distributions—comparisons between mean forecasts and nonstochastic forecasts, comparisons between forecast variances from multiperiod endogenous simulations and those from one-period simulations, and comparisons between forecast variances and residual variances.

II. Specification of the Models and Description of the Test Statistics

A. The Models

As noted in the introduction, we test two models in this paper. The first, the Michigan model, is a relatively small model with 24 estimated equations. It has almost no financial sector and operates with the interest rate on 4–6 month commercial paper as its exogenous monetary instrument. The second model, an old version of the FRB-MIT model, has 75 estimated equations and a fairly elaborate financial sector which gives us a choice among possible monetary instruments.[2] We chose the money stock, because the model has most often been used that way, and, because that is consistent with the estimation procedure; the demand for demand deposits in the FRB–MIT model was estimated with an interest rate as dependent variable and demand deposits as an independent variable.

Both models are estimated on quarterly data, the Michigan model on data for the period 1954(1) through 1967(4), the version of the FRB–MIT model we test on post-Korean War data up through 1968(3). The Michigan model was estimated by two-stage least squares with a special adjustment for serial correlation in two of the equations. Many of the equations are in first-difference form. The FRB–MIT model was estimated by ordinary least squares. In a majority of the estimated equations first-order serial correlation coefficients were estimated, and partial differences taken.

The models are noncomparable not only with regard to estimation period but also, and perhaps more importantly, with regard to what is taken as exogenous. In all cases we set the forecast-period values of the exogenous variables at their actual values. To do otherwise would mean specifying equations for those variables and, in so doing, venturing far from the reported base models. On balance, the FRB–MIT model takes fewer variables as given than does the Michigan model, which one might expect given their relative sizes. The differences are summarized in a rough way in Table 1. Note that the set of exogenous variables for FRB–MIT is not simply a subset of that for the Michigan model. In particular, we should emphasize that we shall be examining reduced forms as functions of two quite different monetary instruments; the money stock in FRB–MIT, the commercial paper rate in Michigan.

[2] We altered two equations in the FRB-MIT model, those for capacity utilization and the unemployment rate. In both cases it was an alteration of form only, one that constrained the variables to their economically meaningful ranges, roughly speaking (0, 1). In both cases, residual standard errors for the variables themselves were lower for our forms than for those originally in the model.

TABLE 1

PRINCIPAL EXOGENOUS VARIABLES BY MODEL

		Mich.	FRB-MIT
Monetary	Narrowly Defined Stock of Money	—[a]	X[b]
	Interest Rate on 4–6 Month Commercial Paper	X	Y[c]
Fiscal	Ratio of Personal Tax Payments to Personal Income	X	X
	Ratio of Corporate Tax Liability to Before-Tax Profits	X	Y
	Transfer Payments to Persons	X	Y
	Federal Government Expenditures	X	X
GNP and Income Components	Capital Consumption Allowances	X	X
	Exports	X	X
	Imports	X	Y
	State and Local Government Expenditures	X	Y
	Farm Investment	Y	X
Deflators	Gross Auto Product	X	Y
	Gross Farm Product	X	X
	Total Government Purchases	X	Y
	Exports	X	Y
	Imports	X	X
	Inventories	X	Y

[a] Not in the model.
[b] "X" stands for independent or exogenous.
[c] "Y" stands for dependent or endogenous.
[d] In the Michigan model, net exports and its deflator are exogenous variables.

In order to make a test for which statistical properties can (in principle) be determined, the models must be specified in stochastic terms. This means that for the types of tests we wish to make, more must be specified or assumed about the models than has been reported. It follows that the model tested is, in effect, a composite between a base model reported by its originators and our addendum, which will be described in detail below. One point, however, deserves mention here. We assume that the structural equation residuals are independent across equations. This is consistent with both the reported estimation procedures and the lack of reported covariances. We admit, though, that abandoning that assumption could have far-reaching effects on test results.

B. Estimation and Forecast Periods

Since, in general, the specification (functional forms, variables included, etc.) of each model was not determined before viewing the base-period data, it seemed imperative to use a comparison period outside that used to estimate the model initially. Therefore, we identify the base period for test purposes with the reported estimation period and use for the comparison period a subsequent period which we refer to as the "forecast" period. Given the data available when we performed the computation, the result is a twelve-quarter forecast period for the Michigan model, 1968(1) through 1970(4), and a nine-quarter period for the FRB–MIT model, 1968(4) through 1970(4).

As we shall see, a disadvantage of such a breakdown is that a wider class of tests could be performed if the "estimation" period was shortened and the "forecast" period lengthened enough to allow all parameters to be estimated from data for the "forecast" period alone. In particular, a test of the hypothesis that all parameters changed—as opposed to tests that certain functions of the parameters changed—might then be possible. However, even then, tests of hypotheses similar to ours would still be of interest and the calculation of the statistics for them not any simpler.

C. *The Test Statistics*

In Appendix I we argue (i) that with a post-sample comparison period as short as ours, the structural change hypotheses that are testable are those equivalent to hypotheses about whether the observed values of the endogenous variables for the "forecast" period come from the distribution predicted by the model estimated from the sample period data, and (ii) that an appropriate test is a prediction interval test, where the rejection region is of the form

$$D = [C(y - \hat{y})]'[C\hat{\Sigma}C']^{-1}[C(y - \hat{y})]/r \geq F_\alpha(r, s).$$

Here y and $\hat{y}$ are nM-element vectors of actual and predicted values of the endogenous variables in the forecast period, n being the number of endogenous variables and M the number of quarters in the forecast period. $\hat{\Sigma}$ is the $nM \times nM$ estimated covariance matrix of y. C is an $r \times nM$ matrix of constants of rank r. As described below, $\hat{y}$ and $\hat{\Sigma}$ are computed conditional on the values of the endogenous variables during the estimation period. $F_\alpha(r, s)$ is the $1 - \alpha$ percent point of an F distribution with r and s degrees of freedom, where s is a rough average of the degrees of freedom (in estimating the residual) for the structural equations of the model. For both models, we used $\alpha = .05$ and $s = 48$. Since D is a positive-definite quadratic form in $C(y - \hat{y})$, the acceptance region $D \leq F_\alpha(r, s)$ is an ellipsoid in Cy centered at $C\hat{y}$.

If C is taken as an $nM \times nM$ identity matrix, we are asking for rejection if any detectable structural change took place. A detectable structural change is a change in an estimable function, estimable from the post-sample period data alone. The fact that our comparison period is "short" implies that there is, in fact, a set of undetectable changes. These are parameter changes constrained so that they do not affect the predicted distribution of the endogenous variables in our comparison (forecast) period.

By using different C matrices, we can attempt to delineate what type of change has taken place. The effect of C is to filter out certain subsets of detectable changes. By varying C, we can also make use of the fact that we can test for some types of change with greater power than others. This is because (with a fixed "normalization" for C) we can predict (if no change has taken place) some linear combinations of y with greater accuracy and, therefore, can detect smaller changes.

D. *Computation of the Statistic D*

Because the models consist of nonlinear structural equations, we compute $\hat{y} = (\hat{y}_{11}, \ldots, \hat{y}_{1M}, \hat{y}_{21}, \ldots, \hat{y}_{2M}, \ldots, \hat{y}_{nM})$ and $\hat{\Sigma}$ by way of Monte Carlo experiments. That is done by repeatedly drawing values of the structural parameters consistent with the estimation period mean and covariance estimates, and values for the forecast period residuals consistent with the estimation period residual variance estimates, and for each drawing generating an Mn element "observation" on y, with the estimation period values of the endogenous variables held fixed at the actual values. For each model we take 300 random drawings and take as $\hat{y}$ the (Mn-element) vector of averages of those observations and as $\hat{\Sigma}$ the sample ($Mn \times Mn$) covariance matrix.

The random parameters are generated one structural equation at a time.[3] Letting $\hat{\hat{\alpha}}_i$ stand for the column vector of random parameters of the i-th estimated equation, a priori sample values of $\hat{\hat{\alpha}}_i$ are generated by the matrix equation,

$$\hat{\hat{\alpha}}_i = \hat{\alpha}_i + R_i'v \tag{1}$$

where $\hat{\alpha}_i$ is the estimation period vector of point estimates, v is a column vector of independent, mean zero, variance one, random variables generated by a random number generator[4] (drawn independently for different equations), and R_i is a matrix such that $R_i'R_i$ equals the estimation period estimated covariance matrix of the point estimator. It follows, then, that $\hat{\hat{\alpha}}_i$ generated by equation (1) has mean $\hat{\alpha}_i$ and covariance matrix $R_i'R_i$, the estimated covariance matrix of the point estimator.

The additive disturbance for each estimated equation is random both among runs and among periods in each run. It is chosen independently across time and equations according to

$$w_i(j) = \sigma_i v \tag{2}$$

where $w_i(j)$ is the residual for the i-th equation at time j, σ_i is the estimation period residual standard error of the i-th estimated equation, and v is a random variable with the same properties as the v in (1). (Note that the v's referred to in (1) and (2) are drawn independently.)

Given (1) and (2), a single M-period simulation run may be thought of as generated as follows. First a random set of parameters is drawn for each estimated equation. Those drawings constitute the parameter values for the run. Then, residuals are drawn, one for each estimated equation. These are embedded in the equations, and a solution, $y^{(1)} = y_{11}, y_{21}, \ldots, y_{n1}$, obtained via the Gauss–Seidel iterative procedure. That solution is dependent on actual estimation-period values of all variables and on actual forecast-period values of exogenous variables. Then a new set of residuals is drawn, again according to (2), and a solution, an observation on $y^{(2)}$, obtained. That observation is again dependent on actual estimation-period values of all variables and on actual forecast-period values of exogenous variables, and, in addition, is dependent on the previously solved for value of $y^{(1)}$. Proceeding in this way, observations on $y^{(3)}, y^{(4)}, \ldots, y^{(M)}$ are obtained. As noted above, for the

[3] This follows from the assumed independence of disturbances across structural equations.

[4] The elements of v are drawn from a truncated normal distribution. Let x be a zero-one normal random variable. We draw values of x and accept only those for which $|x| < 2$. The accepted x's have mean zero and variance $(0.88)^2$, so that $v = (1.137)x$ has mean zero and variance one, the desired distribution. We choose v's from a truncated distribution, because most parameters and disturbances do not a priori have infinite range.

The above description applies to all parameters except first-order serial correlation coefficients in the FRB-MIT model. For their distribution, see Appendix II.

principal tests, we performed 300 such M-period endogenous simulation runs for each model.[5]

III. Basic Test Results

Before turning to test results, it may be helpful to focus on some of the raw data. Figure 1 shows a number of single-quarter forecast distributions for real GNP from the Michigan model; while Figure 2 shows such distributions for the GNP deflator. Figures 3 and 4 show corresponding distributions from the FRB–MIT model. There is a clear-cut relationship between the forecast span and the variances of those distributions: the greater the forecast span, the greater the variance. We shall argue below that this arises mainly from the presence in the models of lagged endogenous variables and the fact that the greater the forecast span, the greater the number of those variables generated randomly within the simulations. Notice that in Figure 4, at each date the actual value of the deflator lies outside the estimated distribution of possible outcomes forecast by the FRB–MIT model.

We limit all our testing to a subset of the endogenous variables of the models: for Michigan, the 12 variables listed in Table 2, for FRB–MIT, the 16 variables listed in Table 3. This means that the columns of C corresponding to all other variables have all zero elements. For Michigan, the list includes an exhaustive breakdown of the endogenous components of nominal GNP—variables 3, 5, 9, and 12—while for FRB–MIT it includes a similar breakdown except that imports, which are endogenous, are excluded. Tables 2 and 3 contain a variable-by-variable view of the output; for each variable and each date, we list the actual value, the actual minus the mean value (the means of distributions like those in Figures 1–4), and the standard error of forecast (standard deviations of distributions like those in Figures 1–4).

To the extent that the structure embodied in each estimated model applies over the forecast period, the standard errors of forecast in Tables 2 and 3 measure the precision of single-date, single-variable forecasts made conditional on values of the variables assumed to be exogenous. For some variables, those standard errors seem quite large. For real GNP for the Michigan model, they range from almost 1 percent of the level for the first quarter of the forecast period to about 5 percent for the twelfth quarter; for the FRB–MIT model they range from about three-fourths of 1 percent in the first quarter to almost 4 percent in the ninth quarter.

For any variable at any date, the ratio of the forecast error (the second entry) to the standard error of forecast (the third entry) is a single-variable version of the D of section II and can be treated as a t statistic with 48 degrees of freedom, $t_{0.05}(48) = 2.01$. The F statistics in the last column are for each variable over all quarters of the forecast period.[6] The relevant 5 percent critical values are

[5] We performed checks on both the input and the output; the output was checked for oscillatory within-run behavior, while the input was checked for coding errors (see Appendix II).

[6] In terms of the statistic D, the t^2 statistic for variable i in quarter j is found by using for C the relevant row of an identity matrix of order Mn: namely, the row with unity in the $[(i\text{-}1)M + j]$th column. The F statistic for the i-th variable is found by using for C the rows obtained by letting $j = 1, 2, \ldots, M$.

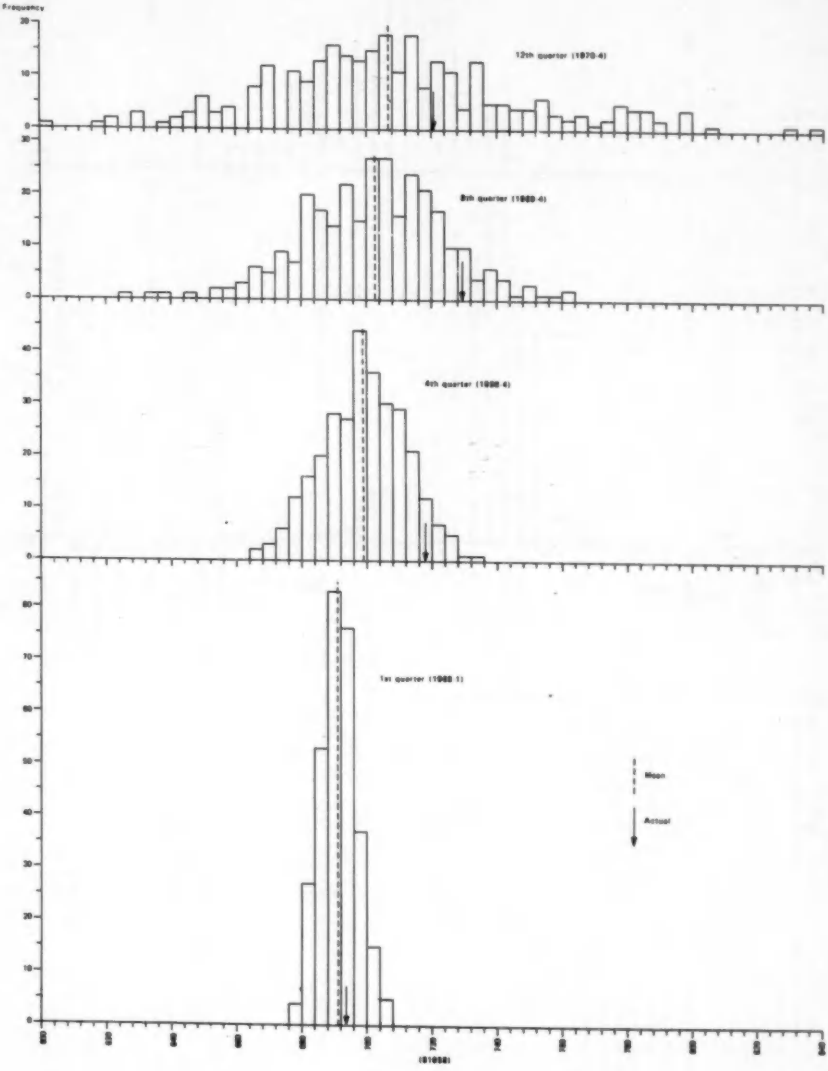

Figure 1 Michigan: forecast distributions of real GNP

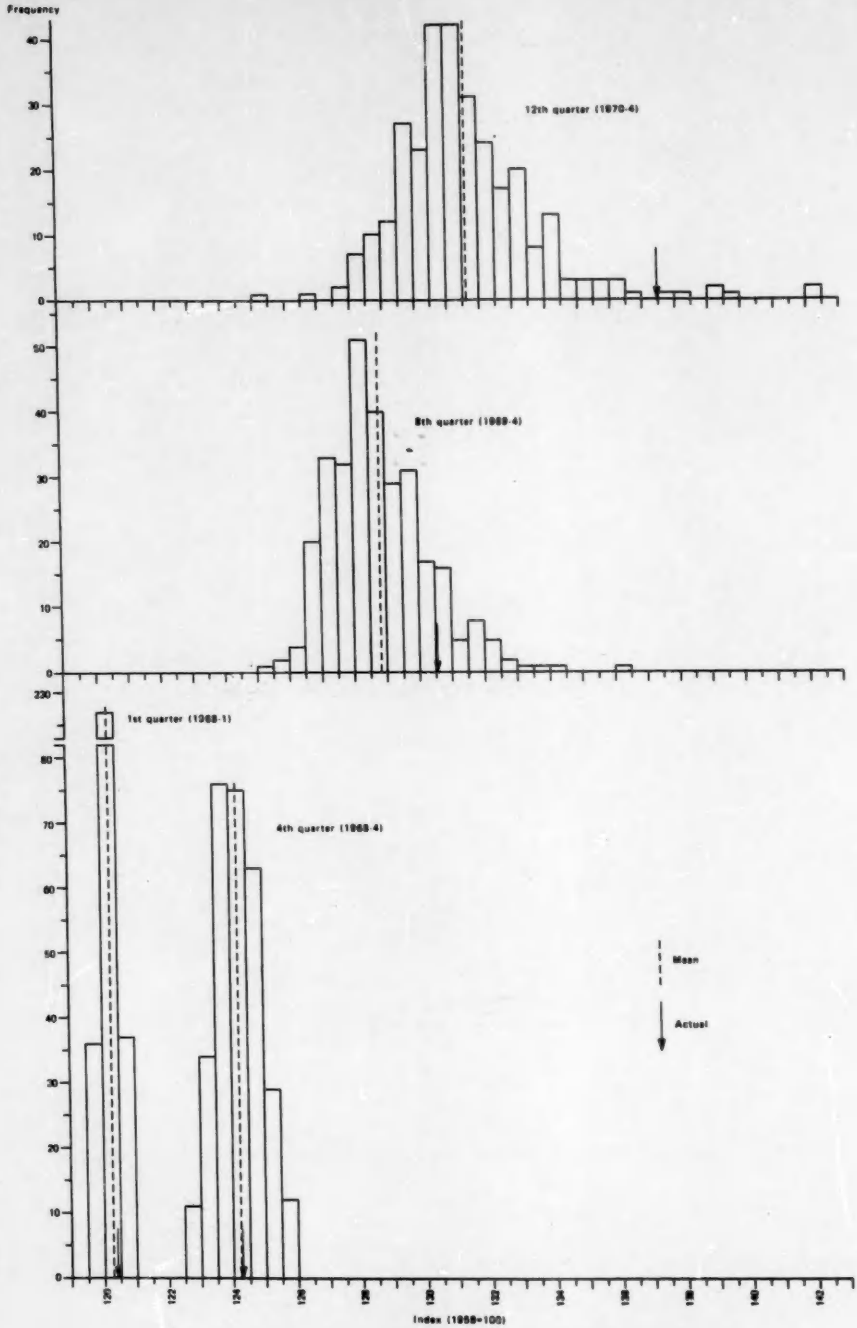

Figure 2 Michigan: forecast distributions of the GNP deflator

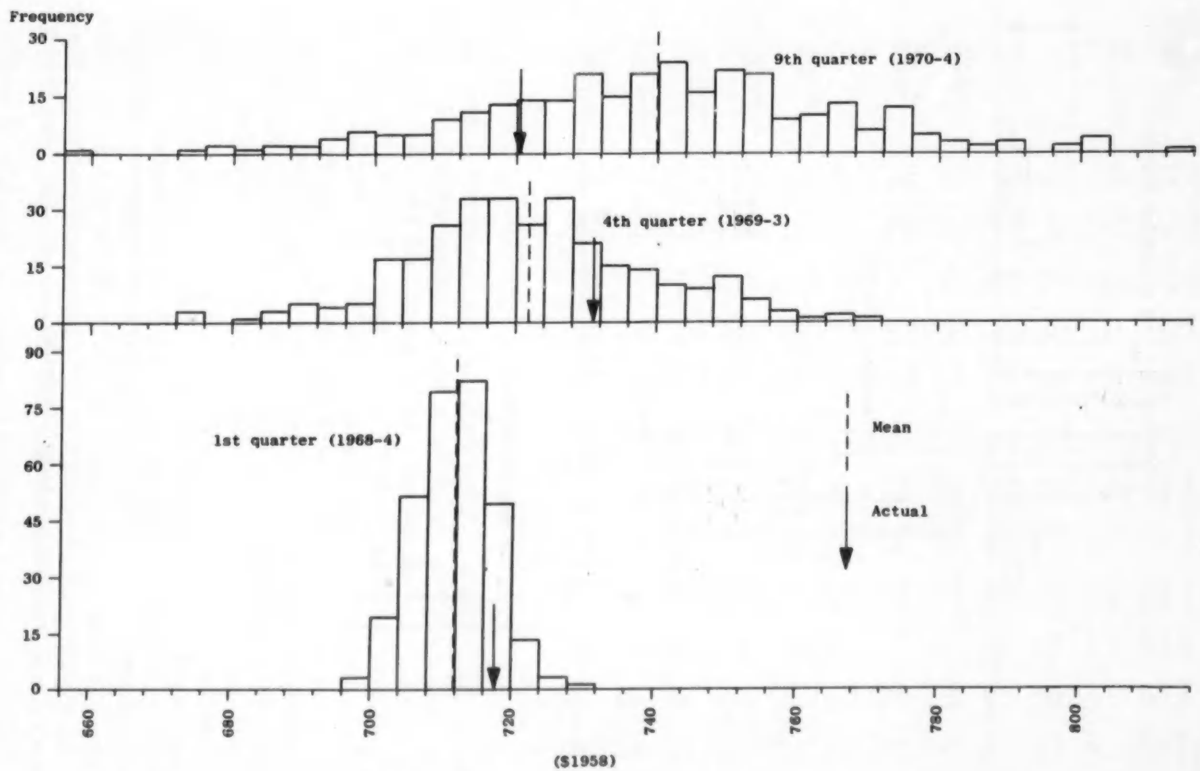

Figure 3 FRB-MIT: forecast distributions of real GNP

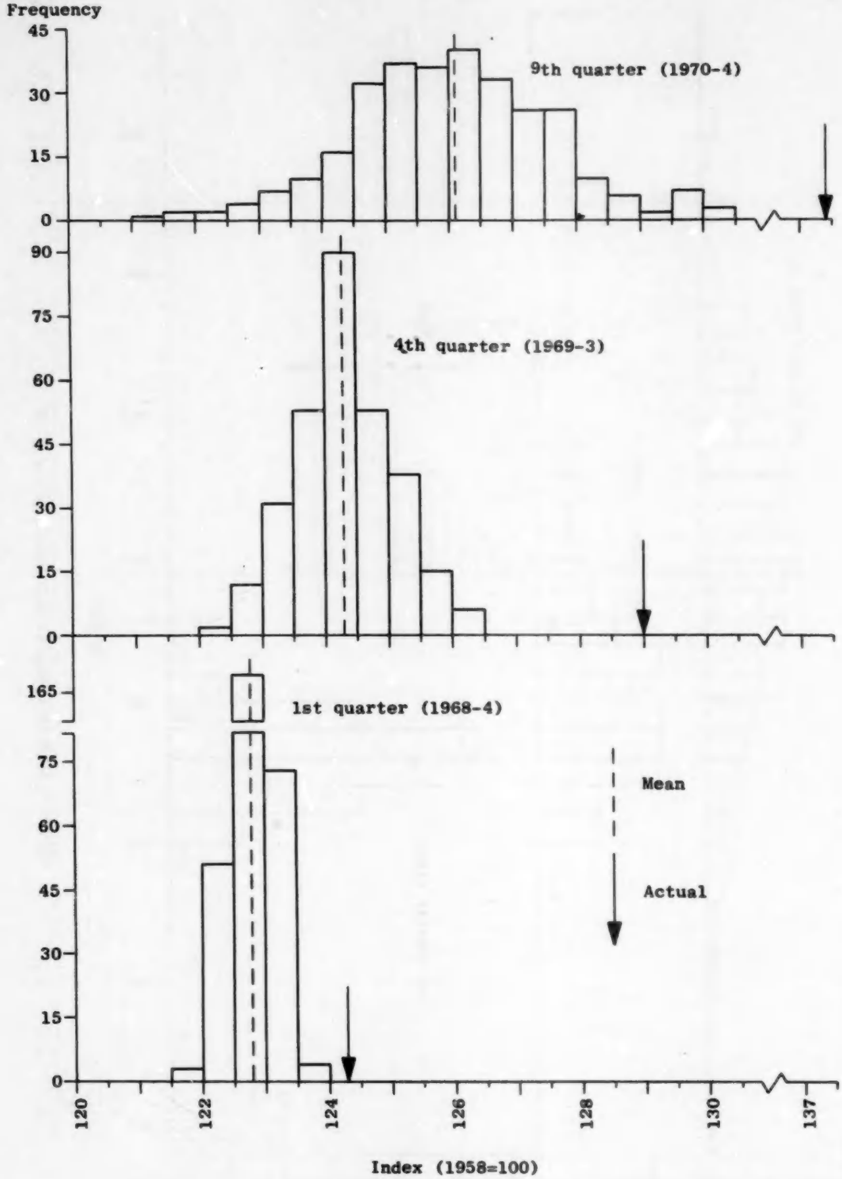

Figure 4 FRB-MIT: forecast distributions of the GNP deflator

TABLE 2

MICHIGAN: ACTUALS, FORECAST ERRORS, AND STANDARD ERRORS OF FORECAST

	1968-1	1968-2	1968-3	1968-4	1969-1	1969-2	1969-3	1969-4	1970-1	1970-2	1970-3	1970-4	F(12,48)
1. Gross National Product ($1958)	693.5 2.4 (5.4)	705.4 7.0 (8.2)	712.6 13.7 (9.9)	717.5 18.5 (12.7)	722.1 23.0 (14.4)	726.1 26.2 (16.1)	730.9 29.6 (18.9)	729.2 26.6 (21.9)	723.8 22.5 (24.7)	724.9 20.5 (28.2)	727.4 18.1 (34.4)	720.3 12.4 (38.3)	.45
2. Implicit Deflator for GNP (1958=100)	120.4 .2 (.2)	121.7 .1 (.4)	122.9 -.1 (.5)	124.3 .0 (.7)	125.7 .3 (.9)	127.2 .5 (1.1)	129.0 1.1 (1.4)	130.5 1.7 (1.6)	132.6 2.5 (1.7)	134.0 3.3 (1.9)	135.5 4.2 (2.1)	137.4 5.8 (2.2)	2.25*
3. Consumption ($)	519.6 7.2 (4.1)	529.0 6.1 (6.3)	543.8 13.3 (7.8)	550.8 14.1 (9.4)	562.0 19.3 (11.0)	573.6 22.6 (13.0)	582.1 23.0 (15.6)	592.5 25.8 (19.0)	603.1 27.7 (22.5)	614.4 28.2 (25.8)	622.0 28.0 (31.8)	627.2 34.7 (35.5)	1.04
4. Corporate Before-Tax Profits ($)	84.7 -2.1 (3.0)	88.6 .1 (4.0)	88.4 2.5 (4.8)	91.3 3.4 (6.0)	93.0 5.0 (6.7)	93.4 5.9 (7.5)	89.9 5.9 (8.5)	88.5 2.3 (9.5)	82.6 -1.5 (10.3)	82.0 -.5 (11.5)	84.4 -.2 (13.8)	79.2 -.2 (15.0)	.50
5. Business Fixed Investment ($)	88.4 2.6 (1.3)	86.4 -.3 (2.6)	88.3 1.7 (3.8)	91.6 3.6 (5.3)	95.7 6.4 (6.9)	97.5 7.8 (8.5)	101.5 11.6 (10.1)	102.7 13.4 (11.9)	102.6 15.0 (13.7)	102.8 17.4 (15.5)	103.6 19.9 (17.2)	101.3 19.0 (18.8)	2.02*
6. Private Nonfarm Housing Starts (0,000's)	146.9 -6.3 (8.0)	141.8 -14.3 (10.5)	132.4 -8.7 (13.4)	157.9 -6.6 (15.0)	169.2 6.7 (17.6)	149.6 -.1 (21.2)	142.9 4.2 (21.8)	135.7 3.9 (22.2)	125.2 -.1 (23.6)	128.6 5.1 (24.7)	151.2 18.7 (27.2)	175.3 26.4 (27.3)	.68
7. Corporate AAA Interest Rate (%)	6.1 .0 (.1)	6.3 .0 (.1)	6.1 -.1 (.2)	6.2 -.1 (.2)	6.7 .1 (.2)	6.9 .0 (.3)	7.1 -.1 (.3)	7.5 .2 (.3)	7.9 .5 (.3)	8.1 .7 (.4)	8.2 .6 (.4)	7.9 .7 (.4)	2.27*
8. Unemployment Rate (%)	3.7 -.4 (.3)	3.6 -.5 (.5)	3.6 -.8 (.7)	3.4 -1.4 (.8)	3.4 -1.9 (1.0)	3.5 -2.2 (1.2)	3.6 -2.6 (1.3)	3.6 -3.1 (1.5)	4.2 -3.0 (1.7)	4.8 -2.9 (1.9)	5.2 -2.8 (2.2)	5.8 -2.6 (2.5)	.68
9. Change in Business Inventories ($)	2.6 -6.2 (3.1)	10.4 2.5 (3.5)	8.2 2.0 (3.7)	9.3 5.4 (4.0)	7.4 4.7 (4.5)	7.9 4.1 (4.4)	11.3 9.5 (4.4)	7.2 5.2 (4.8)	1.6 2.3 (5.0)	3.1 2.7 (4.8)	5.5 4.2 (6.1)	3.6 1.4 (6.3)	1.21
10. Output Per Manhour Nonfarm Index 1957-1959=100	132.4 -.5 (1.1)	133.7 -.8 (1.5)	134.2 -1.6 (1.9)	134.6 -2.7 (2.4)	134.1 -4.7 (2.9)	134.0 -6.5 (3.4)	134.2 -8.5 (4.1)	134.3 -10.6 (4.8)	133.4 -13.8 (5.7)	134.7 -15.5 (6.6)	136.1 -17.2 (7.5)	137.2 -17.9 (8.6)	.67
11. Employment Rate of Males (20 Years and Over (%)	97.7 .2 (.3)	97.8 .3 (.5)	97.8 .6 (.7)	98.0 1.2 (.9)	98.1 1.8 (1.1)	98.0 2.1 (1.2)	97.8 2.4 (1.4)	97.8 2.9 (1.6)	97.3 3.0 (1.8)	96.6 2.8 (2.0)	96.2 2.7 (2.4)	95.8 2.7 (2.7)	.98
12. Residential Construction ($)	28.8 .1 (.7)	30.6 .7 (1.2)	29.9 -.9 (1.7)	31.7 .0 (2.2)	33.0 .9 (2.6)	33.9 2.6 (3.0)	31.0 1.5 (3.5)	30.4 2.2 (3.7)	29.1 2.0 (4.0)	28.4 2.0 (4.2)	29.2 2.3 (4.6)	32.2 3.1 (4.9)	1.10

Notes: (a) Forecast Error = Actual - Mean Forecast.
(b) The F statistics are by variable over 12 quarters. Here and in subsequent tables, F values in excess of the relevant .05 critical values are starred.

$F_{0.05}(12, 48) = 1.96$ for the Michigan model, and $F_{0.05}(9, 48) = 2.08$ for the FRB–MIT model. For the Michigan model, F statistics for the GNP deflator, business fixed investment and the corporate AAA bond interest rate exceed the critical value; for the FRB–MIT model, F's for the GNP deflator, the two interest rates, nonresidential structures, and state and local purchases exceed the critical value. It is interesting that despite differences between forecast periods and exogenous variable sets, the models fail on roughly similar sets of variables: sets which include the GNP deflator, business fixed investment, and the long-term interest rate.[7]

In interpreting the F statistics in Tables 2 and 3, it should be noted that if the model predicted zero correlations among outcomes for the same variable in different quarters, the F statistic for each variable would equal the average of the squared t statistics for the variable. Some examples of the correlations among variables are in Table 4 which contains a submatrix from the matrix of simple correlations between all pairs of the nM variables for the FRB–MIT model. The simple correlations between real GNP at different dates are given in the upper left-hand block; those between the GNP deflator at different dates in the lower

[7] It may also be of interest to note that the FRB-MIT model does poorly predicting the corporate AAA interest rate, but does well predicting the dividend-price ratio, variable 4, even though the former is an important determinant of the latter.

TABLE 3

FRB-MIT: Actuals, Forecast Errors, and Standard Errors of Forecast

	1968-4	1969-1	1969-2	1969-3	1969-4	1970-1	1970-2	1970-3	1970-4	F(9,48)
1. Gross National Product ($1958)	721.8 9.8 (5.3)	712.0 11.8 (10.1)	726.2 8.9 (14.1)	730.7 9.0 (16.9)	729.3 8.6 (19.9)	723.6 2.2 (21.8)	724.7 -3.5 (22.7)	727.3 -5.5 (24.0)	720.5 -19.5 (25.3)	.98
2. Implicit Deflator for GNP (1958=100)	123.5 .7 (.3)	125.7 2.3 (.5)	127.2 3.4 (.6)	129.0 4.7 (.8)	130.5 5.8 (1.0)	132.6 7.4 (1.1)	134.0 8.5 (1.3)	135.5 9.7 (1.4)	137.4 11.4 (1.6)	6.75*
3. Consumption ($)	550.8 3.1 (4.6)	561.8 6.7 (7.1)	573.3 10.4 (8.7)	582.1 11.5 (10.5)	592.6 16.6 (12.6)	603.1 19.8 (13.9)	614.4 21.8 (14.8)	622.1 20.8 (16.5)	627.0 17.5 (18.0)	.37
4. Dividend Price Ratio (%)	2.9 -.3 (.2)	3.1 -.2 (.3)	3.1 -.2 (.3)	3.3 -.1 (.4)	3.4 -.1 (.4)	3.6 .1 (.4)	4.0 .5 (.4)	4.0 .5 (.5)	3.6 .1 (.5)	1.22
5. Commercial Paper Interest Rate (%)	6.0 .8 (.6)	6.7 1.6 (.8)	7.5 2.3 (.9)	8.5 2.6 (1.2)	8.6 3.2 (1.2)	8.6 3.9 (1.1)	8.2 3.6 (1.1)	7.8 3.4 (1.0)	6.3 1.8 (1.1)	2.45*
6. Corporate AAA Interest Rate (%)	6.2 .4 (.2)	6.7 .8 (.3)	6.9 .9 (.3)	7.1 .8 (.4)	7.5 1.4 (.4)	8.0 2.0 (.4)	8.1 2.2 (.4)	8.2 2.4 (.5)	7.9 2.1 (.5)	5.18*
7. Deposits at S&Ls ($)	132.1 .0 (1.0)	134.1 .8 (2.1)	135.3 .8 (3.1)	136.0 .4 (4.0)	136.2 .2 (4.8)	136.8 -.9 (5.5)	139.2 -1.2 (6.1)	143.1 .5 (6.7)	147.4 3.0 (7.3)	.63
8. Corporate Before-Tax Profits ($)	95.7 8.7 (4.8)	93.0 10.9 (7.3)	93.4 8.6 (9.2)	89.9 6.1 (9.8)	88.5 7.2 (10.7)	82.6 2.2 (10.9)	82.3 3.1 (11.0)	84.3 2.4 (11.6)	76.3 -8.8 (12.7)	.82
9. Residential Construction ($)	31.7 1.6 (1.0)	33.0 3.0 (2.4)	33.9 4.6 (3.9)	31.0 2.5 (4.9)	30.4 3.1 (5.3)	29.1 2.5 (5.6)	28.4 1.0 (5.8)	29.2 -.2 (6.0)	32.2 .2 (6.7)	2.05
10. Producer Durables ($)	61.3 -.6 (1.1)	63.1 1.5 (1.8)	65.2 4.5 (2.8)	66.3 6.8 (4.0)	67.5 9.6 (5.1)	66.9 11.3 (6.2)	67.5 14.0 (7.1)	68.6 16.7 (7.7)	66.6 15.2 (8.1)	1.23
11. Nonresidential Structures ($)	30.3 1.2 (.7)	32.6 3.4 (.8)	32.3 3.1 (1.0)	35.2 6.1 (1.3)	35.1 6.2 (1.8)	35.7 7.2 (2.2)	35.3 7.2 (2.6)	35.0 7.5 (3.0)	34.7 7.5 (3.3)	3.40*
12. Change in Business Inventories ($)	9.7 .4 (2.7)	7.3 .6 (4.0)	7.6 3.1 (4.9)	10.8 7.1 (4.9)	6.5 4.6 (5.4)	.9 1.2 (5.6)	2.6 4.1 (5.5)	5.0 5.3 (5.9)	3.0 2.2 (5.9)	.44
13. State & Local Purchases ($)	104.6 7.5 (1.6)	107.6 10.4 (2.1)	110.0 11.3 (2.5)	111.7 11.4 (2.8)	114.3 11.4 (3.3)	117.4 11.6 (3.5)	118.6 7.7 (3.8)	122.4 11.0 (4.1)	125.0 11.1 (4.5)	4.01*
14. Employed Civilian Labor Force (mil.)	76.4 .3 (.4)	77.4 1.4 (.8)	77.6 1.5 (1.2)	78.1 2.0 (1.6)	78.6 2.5 (2.0)	79.0 3.1 (2.3)	78.5 2.7 (2.6)	78.5 2.7 (2.8)	78.6 2.7 (3.0)	1.72
15. Unemployment Rate (%)	3.4 -.4 (.2)	3.4 -.7 (.4)	3.5 -.9 (.6)	3.6 -1.0 (.8)	3.6 -1.3 (1.0)	4.1 -1.2 (1.2)	4.8 -.7 (1.3)	5.2 -.5 (1.4)	5.8 .0 (1.6)	1.07
16. Federal Taxes ($)	187.0 5.6 (3.4)	197.2 17.7 (5.0)	202.5 19.4 (6.5)	200.8 17.2 (7.4)	202.0 18.2 (8.2)	195.9 17.2 (9.0)	196.6 15.2 (8.9)	194.9 16.7 (9.2)	191.7 10.8 (9.7)	1.63

right-hand block; and those between the two variables in the upper right-hand block. The corresponding submatrix for the Michigan model (available upon request) is remarkably similar.

In each case, the correlations between forecasts of a variable at one date and at another date are positive. Moreover, the correlations decline as the time span between the dates increases: namely, looking from the diagonal either across a row or up a column. More interestingly, holding the span between dates fixed, the correlations tend to increase with time: namely, looking down from upper left to lower right on other than the main diagonals. This occurs despite the fact that the variances in Tables 2 and 3 increase with time and implies that the within-path covariance increases even faster. In a sense, it suggests that individual forecast paths become increasingly smooth as the fixed initial set of lagged endogenous variables gets less and less important.

TABLE 4

FRB-MIT: A Submatrix from the Correlation Matrix—Real GNP and the Deflator

REAL GNP									DEFLATOR										
68-4	69-1	69-2	69-3	69-4	70-1	70-2	70-3	70-4	68-4	69-1	69-2	69-3	69-4	70-1	70-2	70-3	70-4		
1.00	.77	.60	.50	.43	.36	.29	.25	.23	-.33	-.13	.02	.15	.18	.21	.22	.21	.18	68-4	REAL GNP
	1.00	.88	.73	.64	.55	.44	.36	.30	-.23	-.16	.03	.21	.26	.33	.35	.35	.33	69-1	
		1.00	.90	.80	.72	.60	.49	.41	-.15	-.12	-.01	.20	.28	.37	.41	.41	.40	69-2	
			1.00	.94	.86	.75	.64	.54	-.12	-.10	-.05	.09	.19	.31	.37	.39	.39	69-3	
				1.00	.95	.85	.75	.65	-.13	-.10	-.08	.04	.11	.25	.32	.35	.37	69-4	
					1.00	.94	.85	.74	-.11	-.10	-.11	-.01	.05	.16	.24	.29	.33	70-1	
						1.00	.95	.85	-.10	-.10	-.13	-.06	-.01	.08	.13	.20	.27	70-2	
							1.00	.95	-.11	-.11	-.16	-.11	-.07	.02	.05	.10	.18	70-3	
								1.00	-.11	-.12	-.17	-.14	-.12	-.05	-.02	.03	.10	70-4	
									1.00	.73	.58	.48	.44	.36	.33	.30	.31	68-4	DEFLATOR
										1.00	.85	.74	.67	.57	.51	.47	.44	69-1	
											1.00	.89	.83	.74	.67	.62	.58	69-2	
												1.00	.93	.87	.79	.74	.69	69-3	
													1.00	.94	.89	.83	.79	69-4	
														1.00	.95	.91	.87	70-1	
															1.00	.97	.93	70-2	
																1.00	.98	70-3	
																	1.00	70-4	

The similarity between correlation matrices for the two models extends to the off-diagonal block. The pattern of asymmetry is common to both models. Real GNP is negatively correlated with past prices and positively correlated with future prices, although the former gets weaker and the latter stronger the further one gets from the beginning of the forecast period.

The positive correlations between real GNP at t and at $t + j$ help explain, for example, why the F statistic for the Michigan model for the vector of GNP outcomes is lower than the average of the squared t's, which is 1.32. The actual forecast errors for real GNP for that model are all of the same sign; the model underpredicts real GNP in every quarter. But because of these positive correlations, those errors cast less doubt on the model than would a sequence of errors of similar absolute magnitude but with randomly varying signs. An average of the squared t's takes account only of the absolute magnitudes. In contrast, the F statistic credits the model for predicting correctly that forecast errors for different dates will be positively correlated.

Table 5 contains joint test results across variables and time. For the Michigan model, tests are performed for variables 2–12 in Table 2 for the first quarter ($m = 1$), the first four quarters, the first eight quarters, and all 12 quarters. Real GNP is omitted, because an identity connects it to the deflator and the endogenous components of GNP. (The test statistics are virtually unaffected by including real GNP and omitting one of the other variables entering the identity. They would be completely unaffected if the identity were linear.) For the FRB–MIT model, tests are performed on all 16 variables in Table 3 for the first quarter, the first four, the first eight, and all nine. Given the variable-by-variable tests in Tables 2 and 3 and the seemingly large standard errors of forecast exhibited there, these results are somewhat surprising. They suggest that neither model's structure is adequate during the forecast period, although that result comes through less strongly for Michigan than for FRB–MIT. Loosely speaking, if these results are put along side Table 2 and 3 results, they suggest that although the models predict fairly well the correlations over time between forecast errors for single variables, they do not correctly predict the correlations among forecast errors for different variables.

IV. Aspects of the Confidence Ellipsoids and Tests on Linear Functions of the Variables

As indicated above, the tests which we perform correspond to examining ellipsoids. In the last section, we to some extent examined $\hat{\Sigma}$ and performed tests

TABLE 5

Joint Test Results

Michigan		FRB-MIT	
m	$F(11\,m, 48)$	m	$F(16\,m, 48)$
1	2.23*	1	3.89*
4	2.62*	4	4.63*
8	2.26*	8	5.81*
12	3.90*	9	5.79*

TABLE 6

CHARACTERISTIC ROOTS AND VECTORS OF THE COVARIANCE MATRIX OF CORPORATE BEFORE-TAX PROFITS FROM THE MICHIGAN MODEL

Root as a Fraction of the Sum	Vector (element *i* multiplies quarter *i* value)											
	1	2	3	4	5	6	7	8	9	10	11	12
0.845	0.04	0.07	0.11	0.15	0.18	0.22	0.26	0.31	0.34	0.38	0.46	0.49
0.089	0.14	0.21	0.30	0.35	0.35	0.34	0.31	0.24	0.05	−0.11	−0.34	−0.45
0.023	0.21	0.34	0.32	0.34	0.21	−0.01	−0.26	−0.40	−0.37	−0.22	0.11	0.40
0.011	−0.43	−0.51	−0.23	0.09	0.39	0.34	0.19	−0.14	−0.29	−0.20	−0.02	0.20
0.007	−0.17	−0.11	−0.01	0.26	0.26	−0.02	−0.35	−0.42	0.16	0.59	0.11	−0.37
0.006	0.42	0.23	−0.34	−0.45	0.09	0.46	0.15	−0.25	−0.26	0.19	0.16	−0.14
0.004	−0.25	0.00	0.26	0.21	−0.51	0.10	0.28	−0.01	−0.44	0.09	0.44	−0.27
0.004	−0.06	−0.09	0.31	−0.09	−0.38	0.41	0.12	−0.51	0.47	−0.09	−0.23	0.14
0.003	−0.51	0.04	0.09	−0.33	0.16	0.24	−0.35	0.13	0.17	−0.32	0.31	−0.16
0.002	−0.38	0.44	−0.02	−0.18	0.17	−0.40	0.54	−0.28	−0.02	0.17	−0.19	0.09
0.002	0.23	−0.13	−0.15	0.12	0.15	−0.28	0.28	−0.27	0.35	−0.46	0.49	−0.26
0.002	0.13	−0.36	0.66	−0.51	0.29	−0.17	0.05	0.01	−0.13	0.08	0.09	−0.07

which involved choosing for C those matrices consisting of different sets of rows of the identity matrix of order nM. In this section we shall examine the shapes of the ellipsoids for certain subvectors of y and shall perform tests on linear functions of them: first, tests suggested by the shapes of the ellipsoids; and then a test of interest, a priori.

We are interested in the shape of the ellipsoid as a means of summarizing the forecast distributions. Thus, if y is a j vector and Σ is its $j \times j$ covariance matrix with characteristic roots $\lambda_1 \geq \lambda_2 \geq \ldots, \geq \lambda_j$ and corresponding unit vectors $v_1, v_2, \ldots, v_j$, then v_1 is to the length-one vector such that the variance of $v_1'y$ is a maximum equal to λ_1. In a sense, then $v_1'y$ is the linear combination about which the model has least to say. Similar interpretations can be given to $v_2'y, v_3'y, \ldots, v_j'y$, where $v_j'y$ is the linear combination with minimum variance. We are also interested in how well the model actually predicts these linear combinations.

We begin with results for the M vector of deviations of each variable for the different dates of the forecast period. It turned out that the shape of the M-dimensional ellipsoid is almost the same for every variable in both models. That allows us to illustrate the results by presenting the M roots and the corresponding vectors for any one of the variables.

As illustrated by the vectors in Table 6—those for a randomly chosen variable—the general pattern of characteristic vectors is that those associated with lower variance exhibit higher frequency oscillations. In each case v_1, the vector associated with the highest variance component, exhibits cycles with a period much greater than the forecast period (i.e., frequency near 0), while v_2 and v_3 exhibit periods with frequency close to the length of the forecast period. The vector associated with the lowest variance typically has a period of two quarters. A second feature of the canonical form is that the first one or two components account for a very large percentage of the variance.

We have also computed for each root the test statistic for the corresponding linear combination. In Tables 7 and 8 we give for each variable the M roots (ranked from largest to smallest and expressed as a fraction of the sum) and above it the corresponding test statistc, $[v_i'(y - \hat{y})]^2/\lambda_i$, which can be evaluated using an $F(1, s)$ distribution.[8] Note that the F statistics in Tables 2 and 3 are simply averages of these. Although we do not discern any clear pattern from these tables directly, by splitting the characteristic vectors into high and low variance groups, certain features can be noticed.

For each variable, we have divided the M-dimensional space into a space of high variance linear combinations (in a sense, those about which the model has little to say) and a space of low variance linear combinations (those about which the model has a lot to say). The test results for each subspace are given in Tables 9 and 10. The parameter k, which is the dimension of the high variance space was determined as follows. Given that the roots are ranked from largest (λ_1) to smallest—$k = 4$ if $\lambda_4/\lambda_1 > 0.05$, $k = 3$ if $\lambda_4/\lambda_1 < 0.05$ and $\lambda_3/\lambda_1 > 0.05$, $k = 2$ if $\lambda_3/\lambda_1 < 0.05$ and $\lambda_2/\lambda_1 > 0.05$, while $k = 1$ if $\lambda_2/\lambda_1 < 0.05$. Given the value of k for each variable, the high variance test statistic for that variable is the average of the corresponding

[8] This statistic is a special case of D, since if C is chosen to be a subset of the characteristic vectors of Σ, then $C\Sigma C'$ is a diagonal matrix with the corresponding roots as diagonal entries. ($F_{0.05}(1, 48) = 4.04$.)

TABLE 7

Michigan: Test Statistic and Root as a Fraction of the Sum

1. Gross National Product ($1958)	.68	3.03	.06	.38	.04	.25	.34	.00	.01	.00	.51	.12
	.87	.08	.02	.01	.01	.00	.00	.00	.00	.00	.00	.00
2. Implicit Deflator for GNP (1958=100)*	3.00	5.37*	6.05*	2.15	2.47	1.55	3.46	.99	1.23	.11	.13	.25
	.90	.07	.02	.01	.00	.00	.00	.00	.00	.00	.00	.00
3. Consumption ($)	1.37	1.43	.34	.29	.00	.04	1.64	.52	.09	.93	.52	5.33*
	.91	.06	.01	.01	.00	.00	.00	.00	.00	.00	.00	.00
4. Corporate Before-Tax Profits ($)	.02	.69	.02	3.08	.18	.04	1.06	.00	.07	.59	.14	.09
	.84	.09	.02	.01	.01	.01	.00	.00	.00	.00	.00	.00
5. Business Fixed Investment ($)*	1.21	.01	.45	.03	.01	1.06	.48	5.95*	.40	2.48	.00	12.11*
	.95	.04	.01	.00	.00	.00	.00	.00	.00	.00	.00	.00
6. Private Nonfarm Housing Starts (000's)	.11	.92	.01	4.16*	.49	.20	1.07	.78	.33	.02	.01	.02
	.73	.14	.06	.02	.01	.01	.01	.01	.01	.01	.00	.00
7. Corporate AAA Interest Rate (%)*	1.18	7.50*	.18	.18	6.50*	3.74	5.02*	.34	1.85	.11	.57	.00
	.86	.07	.03	.01	.01	.01	.00	.00	.00	.00	.00	.00
8. Unemployment Rate (%)	2.39	1.98	1.47	.03	.03	.01	.59	.97	.01	.00	.38	.33
	.90	.07	.02	.01	.00	.00	.00	.00	.00	.00	.00	.00
9. Change in Business Inventories ($)	1.08	1.58	.03	.31	.04	.64	2.29	.54	1.15	3.65	.09	3.09
	.41	.19	.11	.08	.04	.03	.03	.02	.02	.02	.02	.01
10. Output Per Manhour Nonfarm Index 1957-1959=100	5.07*	.13	1.19	.03	.50	.22	.45	.08	.23	.10	.02	.01
	.92	.04	.01	.01	.01	.00	.00	.00	.00	.00	.00	.00
11. Employment Rate of Males (20 Years and Over (%))	2.00	1.24	1.78	.10	.00	.06	1.01	.16	.01	.01	.09	.11
	.90	.06	.02	.01	.00	.00	.00	.00	.00	.00	.00	.00
12. Residential Construction ($)	.32	.01	.02	.88	.43	.36	4.12*	.40	.10	.18	.88	5.55*
	.79	.12	.05	.01	.01	.01	.00	.00	.00	.00	.00	.00

Note: Here and in Tables 8, 9, and 10, a "*" is affixed to any variable for which the *M*-period *F* statistic (see Tables 2 and 3) exceeds the .05 critical value.

TABLE 8

FRB-MIT: Test Statistic and Root as a Fraction of the Sum

1. Gross National	.00	1.84	.02	.87	.60	3.81	.31	1.35	.01
Product ($1958)	.81	.13	.03	.01	.01	.00	.00	.00	.00
2. Implicit Deflator	49.41*	.02	2.13	1.88	.59	2.83	.29	.06	3.49
for GNP (1958=100)*	.88	.08	.02	.01	.01	.00	.00	.00	.00
3. Consumption ($)	1.75	.12	.28	.68	.23	.00	.01	.12	.17
	.86	.08	.03	.01	.01	.01	.00	.00	.00
4. Dividend Price	.13	1.76	.47	1.82	4.59	.01	1.03	.84	.54
Ratio (%)	.72	.15	.16	.02	.01	.01	.01	.01	.01
5. Commercial Paper	11.37*	.10	.14	4.02	1.53	.23	1.59	.62	2.44
Interest Rate (%)*	.63	.18	.08	.03	.02	.02	.02	.01	.01
6. Corporate AAA	22.60*	3.60	.90	11.52	.07	5.09*	2.22	.05	.46
Interest Rate (%)*	.77	.08	.05	.03	.02	.02	.02	.01	.01
7. Deposits at	.01	.06	1.98	2.17	.44	.02	.46	.33	.22
S&Ls ($)	.88	.09	.02	.00	.00	.00	.00	.00	.00
8. Corporate Before-	.13	1.27	.18	2.62	.22	2.15	.39	.02	.42
Tax Profits ($)	.62	.22	.07	.03	.02	.02	.01	.01	.01
9. Residential	.14	.42	.50	.68	2.34	1.62	.23	8.75*	3.73
Construction ($)	.67	.25	.06	.01	.00	.00	.00	.00	.00
10. Producer	4.14*	.01	.00	.11	1.52	.88	1.09	3.01	.27
Durables ($)	.90	.07	.01	.00	.00	.00	.00	.00	.00
11. Nonresidential	8.53*	9.32*	2.47	.04	.04	.66	1.14	8.30*	.07
Structures ($)*	.85	.07	.03	.01	.01	.01	.01	.01	.00
12. Change in Business	1.02	.08	.17	.00	.08	2.45	.04	.08	.03
Inventories ($)	.40	.23	.09	.07	.06	.05	.04	.04	.03
13. State & Local	11.28*	13.58*	4.31*	.01	.94	.21	5.47*	.24	.04
Purchases ($)*	.82	.08	.03	.02	.01	.01	.01	.01	.01
14. Employed Civilian	1.28	.66	.02	.26	1.54	.81	2.74	1.83	6.09*
Labor Force (mil.)	.91	.06	.01	.00	.00	.00	.00	.00	.00
15. Unemployment	.37	2.64	.32	.18	2.23	3.11	.62	.02	.03
Rate (%)	.85	.11	.02	.01	.00	.00	.00	.00	.00
16. Federal Taxes ($)	5.11*	2.78	2.72	3.01	.24	.04	.59	.17	.04
	.70	.14	.05	.02	.02	.02	.02	.01	.01

first k test statistics in Tables 7 and 8, while the low variance test statistic is the average of the remaining $M - k$. The former can be treated as $F(k, s)$ and the latter as $F(M - k, s)$. Since the results for the FRB–MIT model (Table 10) are clearer than those for Michigan, we discuss them first.

For the FRB–MIT model, variables 2, 5, 6, 11, and 13 did not pass the nine-period test. None of these variables pass the joint test of the high variance linear combinations, but all of them except variable 6 pass the joint test of the low variance linear combination. Thus the actual data seem to exhibit a low frequency component with higher variance than the model itself. This can be interpreted to mean that the real world differs from the model in the direction of a naive model. Another way of stating this result is that the model tends to compensate sufficiently for high frequency autocorrelation but not for low frequency autocorrelation.

For the Michigan model where variables 2, 5, and 7 did not pass the twelve-period test, variables 2 and 7 fail the joint test of the high variance (low frequency) combination and pass the joint test of the low variance (high frequency) linear combination.

We also examined the ellipsoid generated by several variables jointly. In particular, we examined the characteristic vectors and values for the covariance matrix for real GNP, the GNP deflator, and the unemployment rate.[9] It would have conveniently fit with our interpretation of the eigen vectors of single variables as frequency components if the joint eigen vectors could have been described as the components of the (3×3) correlation matrix for each frequency, with (approximately) distinct frequencies uncorrelated. This, alas, was not the case. The components of the single variables are obviously correlated across components. For example, the highest variance (joint) component had (roughly) the same form as in the single-variable analysis for the GNP and unemployment partitions, but the price partition behaved in a manner similar to the second and third single-variable components. Indeed, we were not able to find any useful general interpretation of these joint components.

This completes our examination of linear combinations suggested by the forecast distributions themselves. We now examine annual averages, a set of linear combinations which might be considered of interest, a priori.

We present joint test results for all the variables for which quarterly forecasts were tested in Tables 2 and 3. For the Michigan model, we test annual forecasts for the first year, the first two years jointly, and all three years jointly. For the FRB–MIT model we omit the first quarter of the forecast period and test annual averages for 1969, and for 1969 and 1970 jointly. In each case, the test statistic is computed using the relevant matrix C. The results are given in Table 11.

As a forecaster of annual averages, the Michigan model fails the test for the whole forecast period, but passes it for one- and two-year horizons. While the relative standing of the model for different horizons is the same as in Table 5, the model is more consistent as a forecaster of annual averages. The same kind of comparison cannot be made for the FRB–MIT model, because all joint tests on quarterly forecasts were inclusive of 1968(4). Nevertheless, the poor showing of FRB–MIT as an annual forecaster over 1969 and 1970 is not entirely surprising. In the quarterly tests, the model did better forecasting only 1968(4) than it did forecasting for any longer period.

[9] For these computations, each variable was expressed as a ratio to its corresponding mean forecast, so that variances become coefficients of variation, etc.

TABLE 9

MICHIGAN: HIGH AND LOW VARIANCE TEST STATISTICS BY VARIABLE

	k	High Variance Test Statistic	Low Variance Test Statistic
1. Gross National Product ($1958)	2	1.86	0.17
2. Implicit Deflator for GNP (1958 = 100)*	2	4.18*	1.84
3. Consumption ($)	2	1.40	0.97
4. Corporate Before-Tax Profits ($)	2	0.36	0.53
5. Business Fixed Investments ($)*	1	1.21	2.09*
6. Private Nonfarm Housing Starts (000's)	3	0.35	0.79
7. Corporate AAA Interest Rate (%)*	2	4.34*	1.85
8. Unemployment Rate (%)	2	2.19	0.38
9. Change in Business Inventories ($)	4	0.75	1.44
10. Output Per Manhour Nonfarm Index 1957–1959 = 100	1	5.07*	0.27
11. Employment Rate of Males (20 Years and Over (%))	2	1.62	0.33
12. Residential Construction ($)	3	0.12	1.43

TABLE 10

FRB-MIT: HIGH AND LOW VARIANCE TEST STATISTICS BY VARIABLE

	k	High Variance Test Statistic	Low Variance Test Statistic
1. Gross National Product ($1958)	2	0.92	1.00
2. Implicit Deflator for GNP (1958–100)*	2	24.72*	1.61
3. Consumption ($)	2	0.93	0.21
4. Dividend Price Ratio (%)	3	0.79	1.47
5. Commercial Paper Interest Rate (%)*	4	3.91*	1.28
6. Corporate AAA Interest Rate (%)*	3	9.03*	3.24*
7. Deposits at S&Ls ($)	2	0.04	0.80
8. Corporate Before-Tax Profits ($)	4	1.05	0.64
9. Residential Construction ($)	3	0.35	2.89*
10. Product Durables ($)	2	2.07	0.99
11. Nonresidential Structures ($)	2	8.92*	1.82
12. Change in Business Inventories ($)	4	0.32	0.54
13. State and Local Purchases ($)*	2	12.43*	1.60
14. Employed Civilian Labor Force (mil.)	2	0.97	2.00
15. Unemployment Rate (%)	2	1.50	0.93
16. Federal Taxes ($)	3	3.54*	0.68

V. OTHER PROPERTIES OF THE FORECAST DISTRIBUTIONS

A. Nonstochastic Point Forecasts and Their Relationship to Mean Forecasts

We computed nonstochastic point forecasts, those minus mean forecasts, and the standard errors of the mean forecasts, which we take to be the standard errors of forecast in Tables 2 and 3 divided by the square root of 299—299 is the number of Monte Carlo replications minus one. The nonstochastic point forecasts for each model are obtained from a single endogenous simulation over the forecast period

TABLE 11

ANNUAL JOINT TEST RESULTS

Michigan		FRB-MIT	
Forecast Span	F	Forecast Span	F
1968	1.67	1969	8.26*
1968–69	1.26	1969–70	8.41*
1968–70	2.12*		

with parameters and residuals set at their means: the parameters at their point estimates, the residuals at zero. (These data are available upon request.)

For both models, there are some large discrepancies between points and means. A single joint test for each model—to determine whether all the discrepancies could arise from sampling error attributable to the Monte Carlo experiment—yields an F statistic equal to 4.85 for the Michigan model, and one equal to 5.55 for the FRB–MIT model, in each case exceeding the relevant 5 percent critical value. In a statistical sense, at least, points do not adequately represent means, which is what one expects to find for any model other than one consisting of estimated linear reduced-form equations. Of course, despite the high values of the test statistics, one might still want to use the nonstochastic estimates because they can be obtained more cheaply. The important point, though, is that such a judgment would be hard to make before appraising the kind of discrepancies that result for each model.

B. A Sequence of One-Quarter Forecast Distributions

The variation over time of the standard errors of forecast in Tables 2 and 3 could, in principle, be traced to two different sources. One involves the presence in both models of lagged endogenous variables: the greater the forecast span, the greater the number of lagged random disturbances affecting forecasts by way of their effects on the values of lagged endogenous variables. The other involves changes in average initial conditions: each standard error of forecast is a function of the fixed values of the predetermined variables conditional on which the forecast is being made. By analogy with linear models, we expect standard errors of forecast to be larger the more distant are the values of the predetermined variables from their sample period means. And since most variables in these models are stated in terms of levels, deviations of predetermined variables from their means can be expected to increase with time during the forecast period.

In order to draw some inferences about the importance of each source of variation, we computed standard errors of forecast from sequences of one quarter simulations in which lagged endogenous variables are each quarter set equal to actual values.[10] These standard errors of forecasts vary only because average initial conditions change. Unlike those in Tables 2 and 3, they tend to increase only

[10] These data are available upon request. Because we were missing data for many of the endogenous variables for the FRB-MIT model for the period 1969–(2) through 1970–(4), we performed one-period simulations for that model only for the first three quarters of the forecast period.

slightly as a function of time. Thus our suggestion that most of the increase in variance in Tables 2 and 3 is attributable to the presence of lagged endogenous variables seems largely correct.

C. Residual Standard Errors

For a single linear equation, the forecast variance can be split into a sum consisting of the residual variance and the variance of the mean forecast, where the latter is attributable entirely to parameter estimate variance which approaches zero as the sample size increases. The forecast variances we have computed cannot be split up in this way because structural parameters and residuals enter the reduced form nonlinearly. Thus, if we had computed the variances of mean forecasts from a set of simulation experiments in which only parameters were drawn randomly and added them to the corresponding residual variances computed from experiments in which only residuals were drawn randomly, we would not expect the sum to equal the forecast variance. Nevertheless, it is of interest to examine the residual variance, because it provides an estimate of the part of the forecast variance that, in principle, is independent of the amount of data available and that can be reduced only by altering the specification of the model.

Therefore, we computed the ratio of each residual standard error to the corresponding standard error of forecast from Tables 2 and 3.[11] For both models, the ratios tend to decline with time although the pattern is more consistent and far more pronounced for the Michigan model. For example, consider the results for real GNP in the ninth quarter of the forecast period for both models. While the standard error of forecast is about 25 billion for both models (see Tables 2 and 3), for the Michigan model only about 50 percent is directly attributable to the structure of the model and would remain no matter how large a data set had been available; for the FRB–MIT model about 75 percent is attributable to the structure of the model. The models differ more in this respect than in almost any other we have examined.

VI. Concluding Remarks

As we hope is evident, our goal has not simply been to "test" two models. Rather, it has been to illustrate the range of tests one can perform on an estimated simultaneous equation model and the kinds of implications one can draw. It is also our goal to provide something of a rationale for those tests and, hence, to convince others they are worth performing. Since we have dealt throughout with the situation of a post-sample period too short to allow for separate post-sample estimation of the parameters, the only data requirement is that there be *some* post-sample data.

In closing, we would like to add one last caution about interpreting the results of the kinds of tests we have performed. Passing such tests is more impressive the more different is the forecast period from the base period in terms of the regimes

[11] The residual variances were calculated from a set of simulation experiments similar in all respects to those underlying the statistics in Tables 2 and 3, except that parameters were held fixed at their point estimates. The data are available upon request.

generating the variables taken to be exogenous. If there are no grounds for supposing those regimes to be different, the fact that a model passes such tests does not imply a similar validity of its policy evaluation implications.

APPENDIX I. PROPERTIES OF THE STATISTICS

A. With our particular models and small sample sizes, there are no available tests with known optimality properties (in terms of power). Therefore, in choosing both the general form of the test statistic, and the particular estimates and modifications used, along with the distribution used to define the critical region, we have been guided by known results for more simple models and asymptotic results for a general class of models which include ours.

We have chosen to use a test which is a (modified) special case of a general class of tests for which the rejection region is given by

$$\hat{g}'\hat{\Sigma}^{-1}\hat{g} \geq rF_\alpha(r, s)$$

where

(i) g is an r-vector of estimable parameters of the joint distribution of the endogenous variables y,

(ii) the null hypothesis can be stated as $g = 0$, and

(iii) $\hat{g}$ and $\hat{\Sigma}$ are sufficiently good estimates of g and Σ, respectively, made without the null restrictions.

For the normal linear model with scalar covariance matrix, this is the classical F-test which is a UMP invariant test. For the normal linear multivariate regression model, a statistic of this type can be derived as the sum of invariant statistics; however, only in special cases does a UMP invariant test exist. Lehmann [7] and Rao [8] provide thorough treatments of these linear finite-sample-size models.

Wald [9] has shown that in an asymptotic sense, the above test statistic yields a UMP invariant test for a wide variety of models and hypotheses including the multivariate model. The main requirements are that the estimates of g be maximum likelihood (or asymptotically equivalent to m.l.) and that $\hat{\Sigma}$ be the estimated asymptotic covariance matrix. Normality as such is not required. In the limit, the statistic has a $F(r, \infty)$ or $\chi^2(r)/r$ distribution under the null hypothesis which is used to set up the critical region.

Despite these results, we are painfully aware that our models are too distant from the simple models (for which there are finite sample results) and our samples too small to provide any rigorous justification at this time. Nevertheless, in our judgment there are enough favorable indications to justify use of a modified form of the above statistic if the alternative is a nonstatistical test. We now describe the modified form.

B. The general hypothesis we wish to test is that the structural coefficient values are the same in the "estimation" and "forecast" periods. The alternate hypotheses are that those values are different in the estimation and forecast periods. In both cases we assume, and we emphasize this, that the distribution of the structural disturbances remains the same. Thus, we would like to test the hypothesis $g \equiv \beta_2 - \beta_1 = 0$, where β_1 and β_2 are the structural parameters valid in the estimation and forecast periods, respectively. However, in our case the forecast period is too short to allow estimation of the whole set of coefficients for that period. Only certain functions of the parameters can be estimated.

We would like to find the maximum set of (functionally) independent functions of the coefficients (including perhaps some parameters of the distribution of the residuals) which are estimable. By estimable we mean that the equations obtained by setting the first derivatives of the likelihood function (for the forecast period) with respect to the new functions equal to zero have a solution and a negative definite Hessian.

We know of no systematic and practical way of determining such a maximal set of estimable functions. However, three possibilities come to mind. They are the conditional expected values of the endogenous variables in the forecast period conditioned on the values of the endogenous variables (a) before the estimation period, or (b) before the forecast period, or (c) before each time for which a mean is sought. These correspond to three different "reduced forms," different because we are dealing with a system of difference equations. We chose (b) for reasons we will discuss below.

We can now restate the hypothesis we wish to test as $g \equiv h(\beta_2) - h(\beta_1) = 0$ where h is the conditional mean vector computed on the basis of (b). Defined this way, the functions in $h(\beta)$ are estimable and if functionally independent are maximal since they make use of all the forecast-period data. We suspect that functional independence hinges largely on whether any of the structural equations can be estimated on forecast-period data alone. If none can be, the functions in $h(\beta)$ are functionally independent; if some can be, there are dependencies.[12]

A reasonable, in fact, unbiased estimate of $h(\beta_2)$ is the vector of actual values in the forecast period so that $\hat{g} = y_2 - h(\hat{\beta}_1)$, where $h(\hat{\beta}_1)$ is simply the vector of reduced-form "mean forecasts" made on the basis of (b) using the estimation period parameter estimates.[13]

A further consequence of the shortness of the forecast period is that we cannot estimate the covariance of $\hat{g}$. The covariance matrix of $\hat{h}(\beta_2)$, which is a reduced-form covariance, depends on the structural covariance matrix of the residuals (which we assume is unchanged and which can be estimated from the estimation period) *and* on values of the coefficients β_2 which are not estimable. We sidestep this problem by computing the covariance matrix using the null hypothesis. This gives us a type of prediction interval test.

There is one important consequence of using the null hypothesis to compute the covariance of $\hat{g}$. Let Σ be the covariance under the null hypothesis and V that under the alternate hypothesis $\beta_2 \neq \beta_1$. Also let $\Delta = h(\beta_2) - h(\beta_1)$ and assume for the moment that $h(\hat{\beta}_1)$ is an unbiased estimate of $h(\beta_1)$ with negligible variance.

[12] For a linear structure, we can prove the second part of the statement. For any structural equation that can be estimated (with sufficient degrees of freedom to make it "worthwhile"), the elements of $g(\beta)$ corresponding to the L.H.S. variable for that equation should be replaced by the parameters themselves, with, of course, the required changes made in the covariance matrix.

[13] As described in section II, we actually use as an estimate of $h(\beta_1)$ an expected value of $h(\beta_1)$ where the expectation is over the distribution of β_1. This differs from $h(\hat{\beta}_1)$ only because in our case h is non-linear in β. Given this difference, the expectation seems more consistent with our view of the test as a type of prediction-interval test.

Then, the expectation of our test statistic,

$$\begin{aligned} &E(y_2 - h(\hat{\beta}_1))' \textstyle\sum^{-1}(y_2 - h(\hat{\beta}_1)) \\ &= E[\operatorname{tr} \textstyle\sum^{-1}(y_2 - h(\beta_2) + \Delta)(y_2 - h(\beta_2) + \Delta)'] \\ &= \operatorname{tr} \textstyle\sum^{-1}(V + \Delta\Delta'). \end{aligned}$$

Since β_2 is not estimable, there are a large number of $\beta_2 \neq \beta_1$ for which $\Delta = 0$. Among these are almost certainly some for which V is very small, thus, implying $\operatorname{tr} \sum^{-1} V \ll r$. This means that our test is biased in that there are values of $\beta_2 \neq \beta_1$ for which the null hypothesis is more likely to be accepted than for $\beta_2 = \beta_1$. However, the presence of the $\Delta\Delta'$ term means that values of β_2 for which the test is biased and which lie in the direction of increasing Δ are in a bounded region of β_1 (independent of V). For values of β_2 which lie in a direction which leaves Δ unchanged, no bound can be given.

With regard to the choice among estimable functions conditioned on values of the endogenous variables (a) before the estimation period, or (b) before the forecast period, or (c) before each date for which a mean is sought, since the test statistic is essentially a probability density function, if it were based on (a), it could be written as (an integral of) a product of density functions: one based on (b), i.e., conditional on the estimation period; the other giving the distribution of the initial conditions at the end of the estimation period. Thus, it would seem that basing the test on (a) rather than (b) would only add noise and lower the power of the test. And, since we view the test as a prediction interval test, computing it based on (b) seemed easier and more consistent.

Our choice of a distribution to compute the critical region was based on the behavior of the simple and asymptotic cases mentioned above. An F distribution for finite samples is consistent with a χ^2 asymptotically. In addition the simple models indicate that an F might be an appropriate way to take account of the fact that the covariance matrix must be estimated.[14]

As an approximation to the "denominator" degrees of freedom we use a rough average number of degrees of freedom for our equations in the "estimation" period. Because of our assumption of independence of residuals across equations, we did not attempt to subtract the additional degrees of freedom due to inverting an estimated covariance matrix as suggested by the multivariate simple model.

C. Computing our prediction interval test involves solving a large system of simultaneous equations many times. Could not the test be done more simply equation by equation? We think not.

[14] We could have proceeded, in a sense, nonparametrically, by generating a distribution of the statistic D under the null hypothesis, finding the 0.95 percentile point of the distribution, and rejecting if D computed at the actual value of y exceeds that point. For each different test, this would require computations using the entire distribution of the solved-for y's rather than simply the mean and covariance matrix.

At the prompting of our colleague, Professor Sims, we did examine certain one-variable-at-a-time distributions, those plotted in Figures 1–4, to determine whether 5 percent critical regions determined nonparametrically are very different from those based on normality and whether any conclusions would be different. We found no systematic differences between critical regions and in none of the cases examined would our conclusions have been different.

Consider a single equation from a set of simultaneous equations, which, to simplify the exposition, we assume to be linear and without lags:

$$f_t(\beta, \gamma) \equiv y_{1t} + \sum_{k=2}^{n} \beta_k y_{kt} + \sum_{k=1}^{K} \gamma_k x_{kt} = u_t.$$

Let $\hat{\beta}$, $\hat{\gamma}$ be the estimation period coefficient estimates, $\hat{u}_t = f_t(\hat{\beta}, \hat{\gamma})$, and $\hat{u}$ be the vector of $\hat{u}_t$'s from the forecast period, and consider a test statistic of the form

$$\hat{u}' \sum^{-1} \hat{u}$$

with an appropriate $\sum$.

In order for a test based on this statistic to have good properties, we might require $E\hat{u} \cong 0$ and $\sum \cong E\hat{u}\hat{u}'$, with the expectations conditioned on some appropriate observations.

If we condition on the current values of the endogenous variables other than y_1, then, in general, $E\hat{u} \neq 0$. The idea is that conditioning on some endogenous variables implies conditioning on some reduced-form residuals, which, in general, are functions of u.

If, alternatively, we condition on the values of the endogenous variables in the estimation period, we do have $E\hat{u} \cong 0$. But then we must estimate $\sum \cong E\hat{u}\hat{u}'$ with the expectation conditioned on estimation-period values not current values. Assuming, as an approximation, that $\hat{\beta}$ and $\hat{\gamma}$ are unbiased and independent of u, the (t, t') element of $E\hat{u}\hat{u}'$ is, then,

$$\sigma(t, t') = \delta(t, t')Eu_t^2 + \mathrm{tr}[V(Ez(t)z(t')')]$$

where $z(t) = (y_{2t}, \ldots, y_{nt}, x_{1t}, \ldots, x_{Kt})$, V is the covariance matrix of $(\hat{\beta}, \hat{\gamma})$ and δ is the Kronecker delta. But, since $Ez(t)z(t')'$ is, then, an expectation conditional on estimation period values not current values, it depends on the whole system of equations. We are thus led back to the same computations we set out to avoid.

Appendix II. Additional Computational Details

A. *Check for Strange Runs*

The models we deal with are nonlinear. The solution procedure, the Gauss–Seidel iterative routine, finds a solution, but it may not be the only solution. As illustrated by Friedman [4], there is no guarantee that quarter by quarter the solution is not switching, say, between alternative roots of a quadratic equation. The procedure outlined below is designed to discover such anomalies. It identifies runs in which the path over time of any variable exhibits unusually large jumps or oscillations.

Let $y_i(t)$ be the solved-for value of the i-th variable at date t in a particular simulation run. Let $x_i(t) = y_i(t) - \bar{y}_i(t) - [y_i(t-1) - \bar{y}_i(t-1)]$ where $y_i(0) = \bar{y}_i(0)$ —the actual value of the i-th variable in the last quarter of the estimation period—and where for $t > 0$, $\bar{y}_i(t)$ is the mean forecast of the i-th variable at the t-th quarter. The variance of $x_i(t)$ is $V_i(t) = S_i(t, t) + S_i(t-1, t-1) - 2S_i(t, t-1)$, where $S_i(a, b)$ is the covariance of the i-th variable between quarters a and b. We compute the ratio

$$R(i, t) = |x_i(t)|/[V_i(t)]^{1/2}$$

which we expect to be large for runs for which the solution routine is oscillating quarter by quarter between different multiple solutions.

Since for Michigan we examine results for 12 variables over twelve quarters and for FRB–MIT results for 16 variables over nine quarters, and since for each model we performed 300 simulations, there are 43,200 observations on R for each model. The distribution of R for each model is summarized below along with what would be implied by normality for R.

Interval	Frequency		
	Michigan	FRB-MIT	Normality
0–2.0	41,272	41,237	41,271
2.0–3.0	1,811	1,814	1,853
3.0–4.0	103	140	74
4.0–5.0	8	9	2
5.0–6.0	6	0	0
6.0–	0	0	0

Since the results are closely in accord with what we would expect from a normal distribution for x, we concluded that there were no "strange runs" among our simulations.

B. Coding Checks

Since the computer programs that were written to solve the Michigan and FRB–MIT models were not designed for our computations, it was necessary to add a significant amount of new coding. Our computations required two major programming additions: the first was to include a stochastic residual in each structural equation which was consistent with the form of the estimated equation; the second was a subprogram that generated random coefficients and residuals consistent with the distributions implied by estimation.

To check our residual coding and the randomization procedure, a program was written to generate for the estimation period 100 sets of stochastic predictions of the dependent variables and a nonstochastic set. For each equation we generated predictions using actual values of right-hand side endogenous variables and then calculated two statistics: a residual variance

$$\hat{\sigma}^2 = (\hat{y} - y)'(\hat{y} - y)/(N - k)$$

and the ratio

$$R = \frac{1}{100N} \sum_{i=1}^{100} (\hat{y}_i - \hat{y})'(\hat{y}_i - \hat{y})/\hat{\sigma}^2$$

where y is the $(N \times 1)$ vector of actual values of a dependent variable, over the estimation period, $\hat{y}$ the corresponding vector of nonstochastic single-equation predicted values, and $\hat{y}_i$ the vector of stochastic single-equation predicted values generated using the i-th set of random coefficients and residuals. N and k are the number of observations used in estimating the equation in question and the number of independent variables, respectively. If the original coding was correct, $\hat{\sigma}^2$ should equal the residual variance reported in estimation. If our new coding is correct, the ratio R can be treated as $F[(100)N, N - k]$.

In both models these statistics proved helpful in detecting and locating numerous errors that were bound to occur in a project of this size. For example, in a number of equations the residuals were improperly coded causing R to range as high as 1000.

C. Random Values of the Serial Correlation Coefficient

In the FRB–MIT model, a number of equations were corrected for serial correlation by taking partial first differences using an estimated first-order autocorrelation coefficient. Therefore, just as with all other estimated parameters, it was necessary to pick values of the autocorrelation coefficient consistent with the distribution implied by estimation.

Hildreth [5] has shown that the maximum likelihood estimator, $\hat{\rho}$, is asymptotically uncorrelated with all other estimated parameters, is asymptotically unbiased, and has asymptotic variance $(1 - \hat{\rho}^2)/N$, where N is the number of observations. Based on that result and on the constraint that ρ lies in the interval $(0, 1)$, we constructed an approximate distribution for ρ as follows.

Define

$$\rho^* = \frac{1}{1 + e^{A+BX}}$$

where X is distributed normally with mean zero and variance one. Clearly, ρ^* is confined to the interval $(0, 1)$. The problem is to find values of A and B such that, $E(\rho^*) = \hat{\rho}$ and $V(\rho^*) = (1 - \hat{\rho}^2)/N$. To approximate such values, we used a series approximation to ρ^*, denoted r^*; where r^* consists of the first two terms of a Taylor expansion of ρ^* about the mean of X:

$$r^* = \frac{1}{1 + e^A} - \frac{Be^A}{(1 + e^A)^2}X + \frac{B^2e^A(e^A - 1)}{2(1 + e^A)^3}X^2$$

Since X is normal,

$$E(r^*) = \frac{1}{1 + e^A} + \frac{B^2e^A(e^A - 1)}{2(1 + e^A)^3}$$

$$V(r^*) = \frac{(Be^A)^2}{(1 + e^A)^4} + 2\left[\frac{B^2e^A(e^A - 1)}{2(1 + e^A)^3}\right]^2$$

Setting $E(r^*)$ equal to the estimated mean, $\hat{\rho}$, and $V(r^*)$ equal to the estimated variance, $(1 - \hat{\rho}^2)/N$, the resulting equations can be solved for A and B.

The approximation was checked for different $\hat{\rho}$'s by drawing samples of 500 ρ^*'s and calculating sample means and variances. It was found that for $\hat{\rho}$ close to one, the approximation was poor; for $\hat{\rho}$'s greater than 0.9, the sample variances exceeded $(1 - \hat{\rho}^2)/N$ by more than 20 percent. That led us to try a third-order Taylor expansion for ρ^*. With the third-order approximation, for $\hat{\rho}$ less than 0.98, sample means and variances differed from the actuals by less than 5 percent. However, for $\hat{\rho}$'s greater than 0.98, the approximation was still poor. Therefore, for the two equations with $\hat{\rho}$'s in excess of 0.98, we assumed zero variance as one would if first differences had been taken.

References

[1] Chow, Gregory C., "Tests for Equality between Subsets of Coefficients in Two Linear Regressions," *Econometrica*, 28 (1960) pp. 591–605.

[2] DeLeeuw, Frank and Gramlich, Edward, "The Federal Reserve-MIT Econometric Model," *Federal Reserve Bulletin*, 54 Part I (1968), pp. 11–40.

[3] ———, "The Channels of Monetary Policy: A Further Report on the Federal Reserve-M.I.T. Model," *The Journal of Finance*, 24 (1969), pp. 265–89.

[4] Friedman, Benjamin M., "Econometric Simulation Difficulties," Discussion Paper Number 160, Harvard Institute of Economic Research, January 1971.

[5] Hildreth, Clifford, "Asymptotic Distribution of Maximum Likelihood Estimators in a Linear Model with Autoregressive Disturbances," *Annals of Mathematical Statistics*, 40 (1969), pp. 583–94.

[6] Hymans, Saul H. and Shapiro, Harold T., *The DHL-III Quarterly Econometric Model of the U.S. Economy*, Research Seminar in Quantitative Economics, University of Michigan, Ann Arbor (1970).

[7] Lehmann, E. L., *Testing Statistical Hypotheses*, John Wiley and Sons Inc.

[8] Rao, C. R., *Linear Statistics Inference and its Applications*, John Wiley and Sons Inc., New York (Chapter 8).

[9] Wald, Abraham, "Tests of Statistical Hypotheses Concerning Several Parameters When the Number of Observations is Large," *Transactions of the American Mathematical Society*, 54 (1943), pp. 426–82.