NBER WORKING PAPER SERIES

ECONOMETRIC MODELING AS INFORMATION AGGREGATION

Ray C. Fair

Robert J. Shiller

Working Paper No. 2233

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
May 1987

The research reported here is part of the NBER's research program in Economic Fluctuations. Any opinions expressed are those of the authors and not those of the National Bureau of Economic Research.

NBER Working Paper #2233
May 1987

Econometric Modeling as Information Aggregation

ABSTRACT

A forecast produced by an econometric model is a weighted aggregate of predetermined variables in the model. In many models the number of predetermined variables used is very large, often exceeding the number of observations. A method is proposed in this paper for testing an econometric model as an aggregator of the information in these predetermined variables relative to a specified subset of them. The test, called the "information aggregation" (IA) test, tests whether the model makes effective use of the information in the predetermined variables or whether a smaller information set carries as much information. The method can also be used to test one model against another.

The method is used to test the Fair model as an information aggregator. The Fair model is also tested against two relatively non theoretical models: a VAR model and an "autoregressive components" (AC) model. The AC model, which is new in this paper, estimates an autoregressive equation for each component of real GNP, with real GNP being identically determined as the sum of the components. The results show that the AC model dominates the VAR model, although both models are dominated by the Fair model. The results also show that the Fair model seems to be a good information aggregator.

Ray C. Fair
Cowles Foundation
Box 2125, Yale Station
New Haven, CT 06520

Robert J. Shiller
Cowles Foundation
Box 2125, Yale Station
New Haven, CT 06520

ECONOMETRIC MODELING AS INFORMATION AGGREGATION

by

Ray C. Fair and Robert J. Shiller

I. Introduction

Structural econometric models often make use of large information sets in forecasting a given variable. The information sets used in large-scale macroeconometric models are typically so large that the number of predetermined variables exceeds the number of observations available for estimating the model. Estimation can proceed effectively only because of the large number of a-priori restrictions imposed on the model, restrictions that do not work out to be simple exclusion restrictions on the reduced form equation for the variable forecasted. The a-priori restrictions make the model an aggregator of information, an aggregator that could not have been produced without the restrictions.

Are these restrictions basically right in producing derived reduced forms that depend on so much information? Is the large amount of information being applied usefully, for the purpose of forecasting, or is most of it extraneous? With enough observations, it is possible to test all the overidentifying restrictions of a model. One need only compare the unrestricted estimate of the reduced form against the restricted reduced form. The problem with this approach in practice is that with large-scale econometric models there are not enough observations to estimate the unrestricted reduced form, even if this can in principle be done.[1] Also,

---

[1] In practice most structural macroeconometric models are nonlinear, which means that analytic expressions for the reduced form equations are generally not available.

for present purposes we are not interested in testing models against such general alternatives. We would like to test models against much simpler alternatives. We cannot, of course, compare a structural model against all possible simpler alternatives, but parsimony and a sense of relative importance of predetermined variables may lead to a specification of an information set that would be in most simple models.

What variables should one include in a simple model to test a more complicated model against? Advocates of autoregressive forecasting models are one possible source of such an information set. Nelson (1972) and Cooper and Nelson (1975) argued that a simple univariate autoregression would be a good forecaster of real GNP, in which case the information set consists only of lagged values of real GNP. Vector autoregressive techniques restrict the elements of the vector of explanatory variables by excluding all but the most "important" variables. Importance is apparently judged intuitively. Thus, for example, Sims (1980) confines the elements in his vector to real GNP, the GNP deflator, the unemployment rate, the nominal wage, the import price deflator, and the money stock (all these variables except the unemployment rate are in logs). Litterman (serial) before August 1984 confined the elements of his vector to real GNP, the GNP deflator, the unemployment rate, real nonresidential fixed investment, the money stock, and the three-month Treasury bill rate. In August 1984 Litterman (1984) added the value of the trade weighted U.S. dollar and the Standard and Poor's 500 stock price index. This is the VAR model that is now being used by Sims (serial) for periodic forecasts. In these and other cases the list of variables included sounds like the list of variables in a simple textbook macroeconomic model.

The model-free forecasting methods are motivated by a great mistrust of the overidentifying restrictions imposed by macroeconometric modelers -- see for example Liu (1960), Sims (1980), (1986). No case, however, has ever been made that the overidentifying restrictions are so inaccurate that one is better off restricting the forecasting equations as done in the vector autoregressive methodology. We examine this issue below.

In Section II we describe our methodology for testing models. The models that are tested are discussed in Section III. In this section we introduce an "autoregressive components" (AC) model, which is a nontheoretical simple model that is based on the idea that there may be important information in the components of GNP. The results of the tests are presented in Section IV.

II. Tests of Models as Information Aggregators

Consider an econometric model forecast $\hat{Y}_t$ made with information through period t-1. Let Z_t denote a small information vector of dimension k (small relative to the number of predetermined variables in the model), where the variables in Z_t are known as of the end of period t-1. Consider the regression equation:

$$(1) \qquad Y_t = \hat{Y}_t\gamma + Z_t\delta + u_t , \quad t = T_1, \ldots, T ,$$

where $E\hat{Y}_t u_t$ and $EZ_t' u_t$ equal zero and where γ is a scalar and δ is a k x vector of coefficients.[2] We will estimate such regression equations below.

[2]We refer to equation (1) as a regression 'equation' rather than as a regression 'model' to avoid confusion with the econometric model that gave rise to $\hat{Y}_t$. The equation may be considered as just the projection or

There are four possible outcomes when this equation is estimated: 1) neither the estimate of γ nor the estimate of δ is statistically significant, 2) both estimates are significant, 3) the estimate of γ but not of δ is significant, and 4) the estimate of δ but not of γ is significant. We first explore the meaning of each of these outcomes.

Consider first the extreme case in which the true model is simply $Y_t = Z_t\delta + u_t$, so that $E(Y_t|\hat{Y}_t, Z_t) = Z_t\delta$. If the model under consideration uses only information in Z_t or in a subset of Z_t and if $\hat{Y}_t$ is simply a linear combination of some or all of the variables in Z_t, then $\hat{Y}_t$ and Z_t will be perfectly correlated and the coefficients in equation (1) cannot be defined uniquely. If $\hat{Y}_t$ is (incorrectly) based on variables not in Z_t, then the perfect collinearity is broken, and γ (the true γ in the theoretical regression (1)) will be zero. Therefore, if the true model is very simple -- remember that Z_t is meant to be a small subset of the predetermined variables in a typical macroeconometric model -- γ will either be zero or not capable of being determined. This implies that rejecting the null hypothesis H_{01} that $\gamma = 0$ can be construed as showing that the true model contains variables other than those in Z_t and that the model under consideration captures at least some of these additional variables.

Now relax the assumption that the true model is simple. Consider the case where $\hat{Y}_t = E(Y_t|I_t)$, where I_t is a vector of information variables that includes as elements the elements of Z_t and other variables. In this case $\hat{Y}_t$ is not perfectly collinear with Z_t, and in the theoretical regression γ

---

theoretical regression of Y_t onto $\hat{Y}_t$ and Z_t. That is, the coefficients γ and δ in the theoretical regression are defined as coefficients that minimize Eu_t^2.

is one and δ is zero. This implies that rejecting the null hypothesis H_{02} that $\delta = 0$ can be construed as showing that the model under consideration is missing some information included in Z_t.

Although the hypotheses H_{01} and H_{02} are fairly straightforward to test, they are extreme in that they disregard estimation error in the model that produced $\hat{Y}_t$.[3] We now turn to the realistic case in which the forecast $\hat{Y}_t$ is based only on estimates of the coefficients of the model, not the true values themselves.

Since we do not want to take account of the estimation method used to produce the model that gave rise to $\hat{Y}_t$, it is essential that the model be estimated with information only through period t-1.[4] In order for the forecast $\hat{Y}_t$ to be based only on information through period t-1, the model generating the forecast must not be estimated beyond period t-1. One possibility, which we will call "full sample rolling regression," is to use the maximum available number of observations for each period to estimate the model each period: for the forecast for period T_1 the model is estimated

---

[3] In classical statistics, the forecast $\hat{Y}_t$ from a model is not regarded as the conditional expectation $E(Y_t|I_t)$, but rather as an <u>estimate</u> of this conditional expectation. In Bayesian statistics, the forecast $\hat{Y}_t$ may indeed be regarded as $E(Y_t|I_t)$, but in this case the expectation operator E is construed as operating over parameter values as well as values of the noise term u_t. Thus, the parameters γ and δ have their hypothesized values under H_{02} only if the Eu_t^2 is minimized over the prior distribution as well. In real world data, we observe only one drawing from the prior distribution.

[4] For example, suppose the sample period used to estimate the model extends through the sample used to estimate (1). Suppose the estimation method for the model uses up all degrees of freedom and obtains a perfect fit in its forecasts. Our estimates of (1) would show a spurious domination of the model over the information set Z_t. As long as we regard the estimation method for the model as a black box, there is no way to take account of the degrees of freedom used up in estimation of the model.

using data from T_0 through T_1-1, for the forecast for period T_1+1 the model is estimated using data from T_0 through T_1, and so on. (Given the data set, T_0 is meant to be the first observation that can be used in the estimation, after accounting for lags.) This is the procedure followed for the empirical results in Section IV. Another possibility is a "fixed sample size rolling regression," where the starting observation is increased by one each time the ending observation is. A third possibility is simply to estimate the model once through period T_1-1 and use this version for all the future forecasts. We have chosen the full sample rolling regression because it makes maximum use of the available data. Unless otherwise noted, "rolling regression" in what follows will refer to full sample rolling regression.

Having a forecast be based only on information through the previous period may lead to a situation in which, paradoxically, a model seems to be dominated by its own information set. Assume again that the true model is simply $Y_t = Z_t\delta + u_t$ and that the model under consideration is specified correctly. Assume that $\hat{Y}_t = Z_t\hat{\delta}_{t-1}$, where $\hat{\delta}_{t-1}$ is estimated by rolling regressions. Even though the structural model makes use of the same information as in Z_t, we will not observe perfect collinearity between $\hat{Y}_t$ and Z_t, as we would if there were no estimation error, since the parameter $\hat{\delta}_{t-1}$ used to define $\hat{Y}_t$ is stochastic. In this case γ in the theoretical regression will be zero. Thus, the rolling regression forecast <u>is</u> dominated by its own information set in the sense that there is a linear combination of the information variables that does a better job forecasting than does the rolling regression. When data are used to estimate γ and δ, Z_t has an advantage over $\hat{Y}_t$ because the regression can choose the coefficients of Z_t

to fit the current observations, while the rolling-regression coefficients that determine $\hat{Y}_t$ are determined by past observations.

Let us again relax the assumption that the true model is simple. We saw above that when $\hat{Y}_t = E(Y_t|I_t)$, where I_t includes Z_t and other variables, then $\gamma = 1$ and $\delta = 0$ in equation (1). What can we say in the presence of estimation error? In this case γ tends to be less than one even though the model uses more information than that contained in Z_t. To see this, it is useful to regard $\hat{Y}_t$ as equal to $E(Y_t|I_t) + \eta_t$, where η_t is the "estimation error." Assume that the estimation error is uncorrelated with current information I_t and with the current residual u_t in (1). This assumption seems likely to be a good approximation. In the case where the econometric model is a linear regression model of Y_t onto a vector I_t with i.i.d. errors independent of all past, present, and future values of I_t, we can prove the assumption. In this case $\eta_t = I_t(\hat{\beta}_t - \beta)$, where $(\hat{\beta}_t - \beta)$ is a linear function of lagged error terms in the regression over the entire estimation period for $\hat{\beta}_t$. It follows that η_t is uncorrelated with I_t and u_t. With this assumption, we can use standard errors-in-variables results (as applied to theoretical regressions rather than regression estimates) to assess the impact of estimation error on equation (1).

So long as I_t includes relevant information not in Z_t (so that a regression of Y_t on $E(Y_t|I_t)$ and Z_t would produce $\gamma = 1$ and $\delta = 0$), then we would expect to see a positive coefficient on $\hat{Y}_t$. In fact, the coefficient γ of $\hat{Y}_t$ will equal $1 - \nu^2\chi_{11}/(1+\nu^2\chi_{11})$, where ν^2 is the variance of η_t, χ_{11} is the upper-left corner element of $E(X_t'X_t)^{-1}$, and $X_t = [E(Y_t|I_t) \;\; Z_t]$. Thus, the coefficient of $\hat{Y}_t$ must lie between zero and one. Moreover, if one then expands Z_t to include another element also in I_t, γ can never increase and will generally fall.[5] We will see this happening in the tables below, by comparing rows with different numbers of Z_t variables.[6] As one continues to add other variables to Z_t that are in I_t, the coefficient γ will tend to fall until, when all are included, γ will equal zero as long as ν^2 is not zero. We test the hypothesis $\gamma = 0$ and not $\gamma = 1$, because only the former hypothesis tells us something important even in the presence of estimation error.

If ν^2 is small, then, other things being equal, γ will be close to one. We may thus interpret a coefficient of $\hat{Y}_t$ close to one as consistent with the notion that the model has small estimation error. If ν^2 is large, γ will be close to zero.[7]

Under the above assumptions, δ equals $-\nu^2/(1+\nu^2\chi_{11})\chi_{21}$, where χ_{21} is

---

[5]Adding another variable to a random vector can never decrease the diagonal elements, corresponding to the original variables, of the inverse of the covariance matrix. If we begin with an h×h positive definite symmetric matrix A and add another row and column to produce an expanded positive definite symmetric matrix B whose first h rows are [A c] and whose last row is [c′ e], (e a scalar) then the upper-left h×h submatrix C of B^{-1} is $(A-ce^{-1}c')^{-1}$. Using the rule for inverting the sum of two matrices, $C = A^{-1} + A^{-1}c\{e-c'A^{-1}c\}c'A^{-1}$. By Schwartz's inequality $c'A^{-1}c \le e$, so the second term in the expression for C has nonnegative diagonal elements.

[6]Of course, the results for the theoretical regression do not imply that <u>estimated</u> γ must fall in all samples as variables are added.

[7]This result is an application of the simple errors in variables results for multiple regressions. Recall that if there is an error in a single variable in a multiple regression, the coefficient of that variable is biased towards zero. This result is not normally useful in a multiple regression context, since usually errors in variables are not confined to a single variable. If there are errors in more than one variable, then the direction of bias for any variable cannot be predicted. In our application, we need only assume that the we observe the variables in the information set I_t without error, not that they measure correctly what they purport to measure.

the $k \times 1$ vector consisting of the second through k+1th element of the first column of $(E(X'X))^{-1}$. This coefficient is small if ν^2 is small, and it approaches $-\chi_{21}\chi_{11}^{-1}$ as ν^2 is increased.[8] Thus, the larger is the variance of the estimation error in $\hat{Y}_t$, the more likely it is that Z_t will be significant in our tests of H_{02} even when Z_t is used properly in the model.

We thus see that if the true model contains more variables than those in Z_t and if the model under consideration captures at least some of these additional variables, then the estimate of γ is likely to be significant with a big enough sample. We will refer to the test of the hypothesis H_{01} that γ equals zero as the "information aggregation" (IA) test. If the hypothesis is rejected, i.e. the estimate of γ is significant, this is evidence that the model is a useful information aggregator. The IA test is more important than a test of H_{02} because only H_{01} is not affected by estimation error: if $\gamma = 0$ with the true model, then $\gamma = 0$ with the estimated model. The estimate of δ may also, of course, be significant, since even if the model uses correctly the information in Z_t, the forecast has the disadvantage of not being based on the model estimated through period T. If the estimate of δ is not significant, this may mean that the variables in Z_t affect Y_t nonlinearly and that the model adequately captures this nonlinearity. It may also mean that the variables in Z_t simply do not affect Y_t after the model has accounted for all the variables in its

---

[8]Note the identity $\chi_{21}\chi_{11}^{-1} \equiv - M_{22}^{-1}M_{21}$ where $M = E(X_t'X_t)$ is partitioned conformably to $E(X_t'X_t)^{-1}$. Thus, as ν^2 is increased, δ approaches the vector of projection of $E(Y_t|I_t)$ onto Z_t. This is as we would expect from Theil's specification error theorem if $\hat{Y}_t$ were omitted from the regression. Indeed, putting a measurement error onto $E(Y_t|I_t)$ with a very large variance is equivalent to dropping it from the regression.

information set.

We can summarize the implications of this discussion for estimation of equation (1) as follows. If both the estimates of γ and δ are insignificant, then either there is a collinearity problem among $\hat{Y}_t$ and Z_t or neither the variables in the model nor those in Z_t seem to affect Y_t. If the estimate of δ but not of γ is significant, the model may still be correctly specified if the truth is simply $Y_t = Z_t\delta + u_t$ and the model correctly captures this. On the other hand, this result does suggest that the truth is not very complicated and that a simple model is all that is needed. If both the estimates of γ and δ are significant, this is evidence in favor of the proposition that the model incorporates relevant information not in Z_t. This conclusion is also true if the estimate of γ but not of δ is significant. The only difference if the estimate of δ is insignificant is that the variables in Z_t do not seem to affect Y_t in a linear way and may in fact not affect Y_t at all.

Estimation Methods

For the estimation of (1), $\hat{Y}_t$ and Z_t are assumed to be uncorrelated with u_t. In the empirical work below we have been careful to make sure that $\hat{Y}_t$ is based only on information through period t-1 and thus not to be based on variables that may be correlated with u_t. We are also assuming that u_t is serially uncorrelated. We shall include in Z_t a constant term and lagged values of Y_t as well as current and lagged values of other variables. If there are enough lags included in Z_t, then under H_{01} lagged values of u_t may be regarded as in Z_t, so that since u_t is uncorrelated with Z_t, u_t is uncorrelated with its own lagged values as well. Presumably the structural model is good enough that the forecast errors $Y_t - \hat{Y}_t$ are not themselves serially correlated, so that under H_{02} the residuals u_t are serially uncorrelated.[9]

Lack of serial correlation of u_t and lack of correlation between $\hat{Y}_t$ and u_t and between Z_t and u_t does not, however, imply that u_t is independent of $\hat{Y}_t$ and Z_t or even that u_t is uncorrelated with future values $\hat{Y}_{t+j}$ and Z_{t+j}, $j > 0$. Thus, the traditional assumptions of the regression model, which would assure that ordinary least squares gives unbiased estimates of the coefficients γ and δ, are not assured here. Moreover, there is no implication of these assumptions that the error term u_t should be homoskedastic.[10] To take account of possible heteroskedasticity, our hypothesis tests will be Wald tests of the restrictions using an estimate of the asymptotic variance of $\mathbf{X}'\mathbf{u}/\sqrt{n}$, along the lines described in White (1982), where $\mathbf{X}$ is the $k+1 \times T-T_1-1$ matrix whose tth row is $(\hat{Y}_t, Z_t)$, $\mathbf{u}$ is the $(T-T_1-1)$-element vector whose tth element is u_t, and $T-T_1-1$ is the number of observations. We are assuming that $E\mathbf{uu}'$ is diagonal, and so the estimate of the variance of $\mathbf{X}'\mathbf{u}/\sqrt{n}$ is

---

[9] Granger and Newbold (1986, p. 281) have warned that serial uncorrelation of u_t is not valid for many real world models. Uncritical application of our method to models with serially correlated errors could thus lead to a spurious rejection of H_{02}.

[10] In fact, it seems likely that the error term will be heteroskedastic. If, for example, in equation (1) $\gamma = 1$ and $\delta = 0$, then the error term is simply the forecast error from the model, and in general forecast errors are heteroskedastic. For the case in which the model is the classical linear regression model, the nature of the heteroskedasticity can be inferred from the X matrix. Ramsey's (1969) RESET specification test deals with heteroskedasticity by regressing Theil's (1965) BLUS residuals on information variables (in his case, variables related to powers of the fitted values). We cannot use such a procedure here because we are regarding the model as a black box for which BLUS residuals cannot be calculated. In addition, we are using rolling regressions, which does not fit into the procedure.

(2) $$\hat{V}_n = n^{-1}\sum_{t=1}^{n} X_t' e_t e_t' X_t \quad ,$$

where $e_t = Y_t - X_t\hat{\beta}$ $[\hat{\beta} = (\mathbf{X'X})^{-1}\mathbf{X'Y})]$ is the residual in an ordinary least squares regression of Y on X. The Wald test statistic for a hypothesis that $\mathbf{R}\beta = 0$ (where **R** is a $q \times k+1$ matrix) is:

(3) $$W = n\hat{\beta}'\mathbf{R}'\{\mathbf{R}(\mathbf{X'X}/n)^{-1}\hat{\mathbf{V}}_n(\mathbf{X'X}/n)^{-1}\mathbf{R}'\}^{-1}\mathbf{R}\hat{\beta} \quad .$$

Other assumptions are also needed for the asymptotic distribution theory for the Wald test to imply that W is asymptotically χ^2_q. Such assumptions (see, for example, White [1984], p. 125) do not seem unreasonable in the present context, except in special cases.[11] The assumptions required concern such things as the asymptotic independence of $[X_t \; u_t]$ from $[X_{t-n} \; u_{t-n}]$ (so that, for example, the relation of u_t to future X_{t+n} dies out appropriately with n), the convergence of EX'X/n to a positive definite matrix (so that, for example, all of our observations of independent variables do not become zero after a certain time period), and the finiteness of certain moments. The assumptions rule out unit roots for the time-series representations of the processes. We assume for the most part that all processes are stationary around a trend, although results will also be reported below for differenced information sets Z_t.

Note that the use of the asymptotic distribution theory in connection

---

[11] In the special case noted above in which $\hat{Y}_t$ is generated by a full-sample rolling regression on Z_t, the X'X/n matrix will not have a nonsingular probability limit.

with (1) is straightforward in connection with the fixed sample length rolling regression case, but not the full-sample rolling regression case that we actually use. In the former, we might assume that the variables discussed above in connection with the theoretical regression equation (1) are jointly stationary. On the other hand, the case we actually apply, that of full-sample rolling regression, would generally imply if the true model is unchanging that ν declines with time and that $\hat{Y}_t$ is getting better and better through time as a forecast. Ultimately, the estimation error disappears completely, so that asymptotic distribution theory would predict that if $\hat{Y}_t$ makes use of some information not in Z_t, then plim$\hat{\gamma}$ = 1.0. Some Monte Carlo experiments where the true model generating Y_t is a simple linear regression model do confirm that nonetheless in finite samples the coefficient γ does tend to lie between 0 and 1 when Z_t includes some but not all the variables used in a rolling regression to determine $\hat{Y}_t$. The experiments also show that γ tends to fall as more variables are added, in accordance with our theoretical results above. The only anomaly that was found was that the estimated γ showed a tendency to become negative when Z_t included all variables used to determine $\hat{Y}_t$.[12] We suspect that this

---

[12] One experiment, with 10,000 replications, was a follows. In each replication a new 20x2 matrix Z was generated, where the elements of the first column are i.i.d. standard normal variables and the elements of the second column are all 1.0. An independent 20-element column vector u of i.i.d. standard normal variables was created. Y was defined as Z[1 0]' + u. Rolling regressions of Y on Z were run with samples 1,...,j, j = 10,..., 19, and the estimated coefficient vector was denoted β_j. A 10-element vector $\hat{Y}$ was generated whose ith element is $Z_{10+i}\beta_{9+i}$. When the bottom 10-element subvector of Y was regressed on $\hat{Y}$ together with the bottom 10-element submatrix of Z, the average value of the coefficient of $\hat{Y}$ was -3.3 and of the random variable in Z was +4.3. The coefficient of $\hat{Y}$ was negative in 91%

anomalous result is analogous to the small sample bias that tends to produce a negative coefficient when white noise is regressed on its own lagged value.

Comparisons of Different Forecasts

Nothing that we have said so far precludes Z_t being replaced by a forecast from a second model. If the coefficient estimate for one forecast is significant and the coefficient estimate for the other forecast is not, one forecast can be said to "dominate" the other. Under the hypothesis that the coefficient of one forecast is one and the coefficient of the other forecast is zero, the one forecast is said to "encompass" the other in the terminology of Chong and Hendry (1986, p. 677).[13] When equation (1) consists of two forecasts, our procedure of comparing models is similar to that of Nelson's (1972) and Cooper and Nelson (1975). These studies, however, allowed the sample period used to estimate the model to overlap with the sample period used to estimate equation (1). The forecasts they used for period t were not based only on information through period t-1. In the comparisons below we use rolling regressions for all models, and the forecasts from all models are based only on information through the previous period. The studies also did not account for the likely heteroskedasticity

---

of the replications, although a conventional t-statistic on the coefficient of $\hat{Y}$ was significant at the 5% level in only 17% of the replications.

[13] Testing whether a model encompasses another is not the same as comparing the size of the forecast errors. Hendry and Richard (1982, p. 19) emphasize that an encompassing model will variance dominate but a variance dominating model need not encompass.

of the error term in equation (1), which we do here.[14]

III. The Models

In order to carry out the above tests, we need forecasts from models that are based only on information through the period prior to the forecast period (through period t-1 for a forecast for period t). There are four ways in which future information can creep into a current forecast. The first is if actual values of the exogenous variables for period t are used in the forecast. The second is if the coefficients of the model have been estimated over a sample period that includes observations beyond t-1. The third is if information beyond t-1 has been used in the specification of the model even though for purposes of the tests the model is only estimated through period t-1. The fourth is if information beyond period t-1 has been used in the revisions of the data for periods t-1 and back, such as revised seasonal factors and revised benchmark figures.

The way we have handled the exogenous-variable problem is to add autoregressive equations for the exogenous variables to the model. For each exogenous variable in the model an eighth-order autoregressive equation (with a constant term and time trend included) has been postulated. When these equations are added to the model, the model effectively has no exogenous variables in it. This method of dealing with exogenous variables in structural models was advocated by Cooper and Nelson (1975) and McNees (1981). McNees, however, noted that the method handicaps the model: "It is

[14]Cooper and Nelson (1975) emphasized that their forecasts using structural models were "ex ante" predictions. By this, however, they meant only that exogenous variables were forecast using autoregressions and thus that actual future values of the exogenous variables were not used. Their forecasts were not ex ante in the estimation error.

easy to think of exogenous variables (policy variables) whose future values can be anticipated or controlled with complete certainty even if the historical values can be represented by covariance stationary processes; to do so introduces superfluous errors into the model solution." (McNees 1981, p. 404). The Fair model is thus to some extent handicapped in the following tests.

For the coefficient-estimate problem, we use rolling regressions. For the forecast for period t, we estimate the model through period t-1; for the forecast for period t+1, we estimate the model through period t; and so on. By "model" in this case we mean the model inclusive of the exogenous-variable equations. The beginning observation is not changed for the regressions, and so in the terminology of the previous section we are doing "full sample rolling regressions."

The third problem -- the possibility of using information beyond period t-1 in the specification of the model -- is more difficult to handle. Models are typically changed through time, and model builders seldom go back to or are interested in "old" versions. We have, however, attempted to account for this problem in this paper regarding the Fair model. We consider two versions of the Fair model, the current version and the version that existed as of the second quarter of 1976. By comparing the results for the two versions, we can see in some sense how important the specification changes that have been made since 1976 are.

We have done nothing about the data-revision problem in this paper. The data that have been used are the latest revised data. It would be extremely difficult to try to purge these data of the possible use of future information, and we have not tried. Note that it is not enough simply to use data that existed at any point in time (say period t-1) because data on the one-period-ahead value (period t) are needed to estimate equation (1). We would have to try to construct data for period t that are consistent with the old data for period t-1.

We now discuss the various models used for the tests in this paper. The models consist of the two versions of the Fair model, an "autoregressive components" model, and three versions of a VAR model.

<u>The Fair Model -- Current Version (FAIR-CUR)</u>

The Fair model as it existed in 1984 is described in Fair (1984). A few changes have been made to the model over time as it has been updated and reestimated. The version used here is based on data through 1986 II. The model consists of 30 stochastic structural equations and 98 identities. For purposes of this paper two stochastic equations had to be changed. The interest rate reaction function, an equation explaining the behavior of the Federal Reserve, has a dummy variable in it to pick up a possible change in Fed behavior between 1979 IV and 1982 III. This dummy variable was dropped because it contained future information for any sample period that ended prior to at least 1983 I. Likewise, the equation explaining capital consumption has a number of dummy variables in the 1980's to try to pick up the changing effects of depreciation laws. These dummy variables were also dropped.

Dropping the above-mentioned dummy variables left the model with 97 exogenous variables. For each of these variables an eighth order autoregessive equation was postulated with a constant term and time trend included.[15] When these equations are added to the model, there are 127 stochastic equations, and this is the version that was used.

For the results below the model was estimated 57 times. For each of the estimation periods the beginning observation was always 1954 I. The first estimation period ended in 1972 I, the second in 1972 II, and so on through the 57th in 1986 I.[16] (The estimation techniques are two stage least squares for the 30 structural equations and ordinary least squares for the exogenous-variable equations.) This allowed 57 one-quarter-ahead forecasts to be made, starting in 1972 II, each forecast based only on estimates through the end of the previous quarter.

The Fair Model -- Old Version (FAIR-OLD)

The first version of the Fair model was presented in Fair (1976). This version was based on data through 1975 I. One important addition that was made to the model from this version was the inclusion of the interest rate reaction function in the model. This work is described in Fair (1978), which is based on data through 1976 II. Some changes have been made to the

---

[15]Simpler equations were estimated for four exogenous variables. For three of the variables the equations merely consisted of a constant term, and for the fourth variable the equation was a fourth-order rather than an eighth-order autoregressive equation. This was done because of collinearity problems. For the early sample periods there were not enough non-zero observations to allow eighth-order equations to be estimated. The four variables are a dummy variable for 1971 IV, a dummy variable for 1972 I, wage accurals less disbursements of the state and local government sector, and housing investment of the financial sector. The latter variable is the one for which a fourth-order equation was estimated.

[16]The import equation contains a number of dummy variables to pick up the effects of dock strikes. The last dummy variable is for 1972 I, and this is the main reason the first estimation period was chosen to end in 1972 I. The last estimation period ended in 1986 I because the overall sample period ended in 1986 II.

model since 1976 II, and it is of interest to know how important these changes have been. Fortunately, we can work with this early version of the model to examine this question.

The version of the model in Fair (1976) consists of 26 structural stochastic equations. With the addition of the interest rate reaction function, there are 27 stochastic equations.[17] There are 106 exogenous variables, and for each of these variables an eighth order autoregressive equation with a constant and time trend was added to the model. This gave a model of 133 equations, and this is the version that was used.

The first estimation period ended in 1976 II, which is the quarter in which the model could definitely be said to exist. This allowed the model to be estimated 40 times (through 1986 I).

To conclude, the forecasts from FAIR-OLD can be said to be forecasts that are truly based only on information through the previous period (except for the data revision problem).[18] This may be the first time that a model this old has been tested.

---

[17] This compares to 30 structural stochastic equations for the current version of the model. In other words, three more variables are endogenous in FAIR-CUR than are endogenous in FAIR-OLD.

[18] This statement needs to be qualified slightly. Although the structural stochastic equations used for FAIR-OLD here are exactly as in Fair (1976) and (1978) -- same left hand side and right hand side variables -- the data revisions in the National Income Accounts since 1976 have required slight modifications to some of the identities in the model. Also, the identities in Fair (1976) for the government sector are for the total government sector, whereas in FAIR-OLD here there are separate identities for the federal government sector and the state and local government sector. This disaggregation of the government sector does not affect anything except that it means that there are more exogenous variables (and thus more exogenous-variable equations) in FAIR-OLD than there were in Fair (1976).

The Autoregressive Components Model (AC)

Time series models like VAR models typically ignore the components of GNP. For example, the current VAR model used by Sims (serial) includes only nonresidential fixed investment among the various components. Including many components in a VAR model rapidly uses up degrees of freedom, and this is undoubtedly one of the main reasons the components are seldom used. A possible alternative to the VAR approach, but one that also does not use much economic theory, is to model each of the components of real GNP by a simple autoregressive equation and then determine GNP as the sum of the components.

For present purposes we have used a slightly more sophisticated version of what we will call the "Autoregressive Components" (AC) model. Each equation for a component is an eighth order autoregressive equation with a constant and time trend added and with the first four lagged values of real GNP added. The components are three consumption categories, eight investment categories, imports, exports, and four government spending categories. All the 17 components are in real terms. (No logs were taken for the AC model.) Real GNP is determined by the GNP identity. Each stochastic equation of the AC model has only 14 coefficients to estimate, and so there is no serious degrees of freedom problem. The reduced form equation for GNP has 136 lagged components in it, but this equation is never estimated and so there is no problem.

The AC model was estimated 57 times using the same sample periods as were used for FAIR-CUR. The model was then used to make 57 forecasts of real GNP.

The AC model is of interest in two respects. First, if the Fair model turns out to dominate the VAR models (which it does), it is of interest to know if this is due simply to the fact that the Fair model is dealing with the components of GNP. If this is the case, then the AC model should do better than the Fair model, and this can be tested. The AC model carries the components idea even further by including eight lags of the components, which the Fair model does not. Second, the AC model is to some extent a competitor of the VAR model within the class of non theoretical models, at least regarding the predictions of GNP. Both models are based on very little economic theory. It is thus of interest to see if one model dominates the other.

The VAR Models (VAR4, VAR2, and VAR1)

We consider three VAR models in this paper. The first, VAR4, is the same as the model used in Sims (1980) except that we have added the three-month Treasury bill rate to the model. There are seven variables in the model: real GNP, the GNP deflator, the unemployment rate, the nominal wage rate, the price of imports, the money supply, and the bill rate. All but the unemployment rate and the bill rate are in logs. Each equation consists of each variable lagged one through four times, a constant, and a time trend, for a total of 30 coefficients to estimate.

The second VAR model, VAR2, uses only the first two lags of each variable, for a total of 16 coefficients in each equation. The third model, VAR1, uses only each variable lagged once, for a total of 9 coefficients. Litterman (1984) points out that large VAR models seem to suffer from overparameterization, and this is the reason we have tried VAR1 and VAR2 in addition to VAR4.[19]

The same sample periods and procedures were used for the VAR models as were used for the AC and FAIR-CUR models.

The Information Set: the Variables in Z_t

We have chosen to use the 28 variables in the VAR4 model as potential variables in the information set, i.e., as potential variables in Z_t. One of the main questions we are interested in is whether the forecasts from the Fair model contain useful information not in these variables.

IV. The Results

We first consider the results for FAIR-CUR, which are presented in Table 1. All the regressions in Table 1 are for the 1972 II - 1986 II period, for a total of 57 observations. The dependent variable is the log of real GNP. All the equations in the table were estimated by ordinary least squares with the White correction for heteroskedastity. The W-statistic is the Wald statistic for the test of the hypothesis that all the coefficients except the coefficient of the forecast are zero. In other words, the test is a test of the hypothesis that δ in equation (1) is zero. δ is taken to include the constant term.

The first equation includes the log of the forecast, whose coefficient is denoted γ, and the constant term. The estimate of γ is .984 with a t-statistic of 123.73. The estimate of the constant term is not quite

[19]Another possibility is to use Bayesian priors for VAR4 to lessen the overparamaterization problem, which is what Litterman (1984) does. Our preference is to exclude lags rather than use priors, but priors could easily be used within the present procedure. This is clearly of interest to try in future work.

TABLE 1

Estimates of Equation (1) and Tests of H_{01} and H_{02}.

Results for FAIR-CUR. Dependent Variable is log Y.
Sample period = 1972 II - 1986 II, 57 observations.

Equation	log $\hat{Y}$	const.	Other Variables	WALD	SE	R^2	DW
1	.984 (123.73)	.124 (1.94)		3.78	.00835	.99284	2.00
2	.808 (5.92)	.636 (1.66)	T, $\log Y_{-1}, \ldots, \log Y_{-4}$	10.09	.00789	.99361	2.18
3	.966 (18.65)	.266 (0.80)	T, $\log P_{-1}, \ldots, \log P_{-4}$	9.26	.00802	.99340	2.12
4	.788 (9.99)	1.578 (2.70)	T, $U_{-1}, \ldots, U_{-4}$	17.73**	.00709	.99485	2.28
5	.963 (22.49)	.258 (0.84)	T, $\log PM_{-1}, \ldots, \log PM_{-4}$	11.69	.00804	.99338	2.07
6	.975 (19.88)	−.020 (0.07)	T, $\log W_{-1}, \ldots, \log W_{-4}$	6.64	.00817	.99315	1.99
7	.990 (20.56)	.357 (0.94)	T, $\log M1_{-1}, \ldots, \log M1_{-4}$	8.49	.00817	.99315	1.96
8	.839 (19.59)	1.181 (3.75)	T, $r_{-1}, \ldots, r_{-4}$	37.24**	.00724	.99463	2.15
9	.902 (4.01)	2.494 (1.26)	T, all_{-1}	13.81	.00755	.99415	2.04
10	.761 (5.55)	6.503 (2.09)	T, $all_{-1,-2}$	62.15**	.00596	.99636	2.00
11	.607 (3.20)	10.747 (3.20)	T, $all_{-1,-2,-3}$	110.58**	.00557	.99682	1.93
12	.590 (3.02)	9.199 (2.09)	T, $all_{-1,-2,-3,-4}$	201.59**	.00503	.99740	1.92

$\hat{Y}$ = forecast of real GNP from FAIR-CUR.

Notes: The test statistic (IA test) for H_{01} is the t-statistic for the coefficient of log $\hat{Y}$.
The test statistic for H_{02} is the WALD test statistic.
* = significant at 5 percent level.
** = significant at 1 percent level.
t-statistics in absolute value in parentheses.
Y = real GNP.
T = time trend, 1952 I = 1.
P = GNP deflator.
U = unemployment rate.
PM = import price index.
W = nominal wage rate.
M1 = money supply.
r = three-month Treasury bill rate.
all = log Y, log P, U, log PM, log W, log M1, r.

Estimation technique = Ordinary least squares with White correlation for heteroskedasticity.

significant at the 5 percent confidence level.[20] The equation shows no signs of serial correlation of the error terms.

For the second equation in Table 1 the time trend and the first four lagged values of the log of real GNP have been added. The estimate of γ is .808 and is significant. The additional six variables are not significant as a group as revealed by the W-statistic. In other words, the lagged values of GNP do not contribute significantly to the explanation of GNP once the FAIR-CUR forecast is included in the equation.

Equations 3 through 8 in Table 1 have added to them the time trend and the first four lagged values of one of the variables in the information set. The GNP deflator, the price of imports, and nominal wage, and the money supply are not significant. The unemployment rate and the bill rate are significant at the 1 percent level. In all cases the coefficient estimate of the FAIR-CUR forecast is significant. In equation 9 the time trend and all seven variables lagged once are included. They are not as a group significant. In equation 10 the time trend and all seven variables lagged once and twice are included. Equation 11 adds one more lag of each variable, and equation 12 adds yet one more. In all three of these cases the variables as a group are significant and the FAIR-CUR forecast is significant. In going from equation (9) to (12) the estimate of γ falls, as is expected from the theoretical results in Section II.

The overall results in Table 1 are quite supportive of FAIR-CUR being a useful aggregator of information. The coefficient estimate of γ is always significant. The IA test has strongly rejected the hypothesis that γ is

---

[20] By "significant" in what follows is meant significant at the 5 percent confidence level unless stated otherwise. A variable is said to be significant if its coefficient estimate is significant.

zero.

We next want to compare FAIR-CUR with FAIR-OLD to see if the good results for FAIR-CUR are due to information used after 1976 in the specification of the model. The results in Table 2 are for FAIR-CUR over the shorter sample period beginning in 1976 III. Table 2 is the same as Table 1 except that the regressions are over 40 observations rather than 57. The results in Table 2 are similar to those in Table 1 except for the last three equations, where the estimates of γ are not significantly different from zero. These last three equations have only 23, 16, and 9 degrees of freedom, respectively, and so it is not clear that much confidence should be put on the results for these three equations.[21] We discussed above that in small samples the estimate of γ will tend to be much less than one or even negative even when the theoretical model used to produce $\hat{Y}_t$ is absolutely correct. It is difficult to generalize from the few specific Monte-Carlo experiments we ran, but it seems likely that when degrees of freedom are very small, there may be a tendency for a small or negative value of γ.

Table 3 is the same as Table 2 except that the results are for FAIR-OLD rather than FAIR-CUR. The results for FAIR-OLD are quite similar to those for FAIR-CUR. If anything, FAIR-OLD does slightly better. The changes that have been made to the model since 1976 thus do not seem to be very important, at least regarding the forecasting accuracy of the model.

The results in Table 4 are for the AC model. The sample period in Table 4 is the same as in Table 1, namely the longer period of 57

---

[21] To give an example of how fickle the results are when the number of degrees of freedom is small, when equation 10 is reestimated with the wage rate and the price of import variables dropped, the estimate of γ goes from an insignificant -.447 to a significant .504.

TABLE 2

Estimates of Equation (1) and Tests of H_{01} and H_{02}.

Results for FAIR-CUR for Shorter Sample Period. Dependent Variable is log Y.
Sample period = 1976 III - 1986 II, 40 observations.

Equation	$\log \hat{Y}$	const.	Other Variables	WALD	SE	R^2	DW
1	.969 (63.96)	.247 (2.01)		4.04*	.00886	.98216	2.00
2	.862 (3.82)	.734 (1.72)	T, $\log Y_{-1}$, ..., $\log Y_{-4}$	9.57	.00836	.98413	2.34
3	.867 (6.72)	.778 (1.08)	T, $\log P_{-1}$, ..., $\log P_{-4}$	8.30	.00831	.98430	2.11
4	.574 (3.56)	3.203 (2.65)	T, U_{-1}, ..., U_{-4}	37.94**	.00626	.99110	2.50
5	1.014 (14.63)	−.084 (0.17)	T, $\log PM_{-1}$, ..., $\log PM_{-4}$	7.54	.00809	.98513	2.30
6	.879 (10.48)	−.092 (0.28)	T, $\log W_{-1}$, ..., $\log W_{-4}$	7.18	.00818	.98479	1.92
7	.975 (17.35)	.471 (0.90)	T, $\log M1_{-1}$, ..., $\log M1_{-4}$	5.03	.00874	.98265	1.97
8	.821 (17.14)	1.327 (3.73)	T, r_{-1}, ..., r_{-4}	47.33**	.00742	.98750	2.31
9	1.358 (3.22)	3.507 (1.68)	T, all_{-1}	15.42	.00763	.98680	1.94
10	−.447 (0.95)	7.113 (2.46)	T, $all_{-1,-2}$	167.50**	.00478	.99480	2.45
11	−.133 (0.34)	14.548 (3.63)	T, $all_{-1,-2,-3}$	263.61**	.00400	.99637	2.18
12	.225 (0.87)	26.608 (5.46)	T, $all_{-1,-2,-3,-4}$	1318.81**	.00266	.99840	2.47

Notes: $\hat{Y}$ = forecast of real GNP from FAIR-CUR.
See Notes, Table 1.

TABLE 3

Estimates of Equation (1) and Tests of H_{01} and H_{02}.

Results for FAIR-OLD. Dependent Variable is log Y.
Sample period = 1976 III - 1986 II, 40 observations.

Equation	log $\hat{Y}$	const.	Other Variables	WALD	SE	R^2	DW
1	.978 (70.25)	.177 (1.56)		2.44	.00870	.98281	2.06
2	.879 (4.66)	.560 (1.49)	T, $\log Y_{-1}, \ldots, \log Y_{-4}$	10.82	.00820	.98472	2.43
3	.928 (6.75)	.480 (0.63)	T, $\log P_{-1}, \ldots, \log P_{-4}$	4.90	.00849	.98363	2.13
4	.629 (4.14)	2.793 (2.45)	T, $U_{-1}, \ldots, U_{-4}$	43.84**	.00623	.99119	2.47
5	1.011 (13.24)	−.051 (0.10)	T, $\log PM_{-1}, \ldots, \log PM_{-4}$	4.45	.00841	.98395	2.14
6	.928 (11.38)	.040 (0.12)	T, $\log W_{-1}, \ldots, \log W_{-4}$	3.63	.00839	.98403	1.91
7	.969 (15.45)	.328 (0.69)	T, $\log M1_{-1}, \ldots, \log M1_{-4}$	4.15	.00855	.98342	2.00
8	.834 (18.75)	1.226 (3.72)	T, $r_{-1}, \ldots, r_{-4}$	49.79**	.00737	.98766	2.33
9	1.111 (2.69)	2.202 (1.01)	T, all_{-1}	6.56	.00813	.98498	1.87
10	−.042 (0.14)	7.144 (2.28)	T, $all_{-1,-2}$	226.21**	.00485	.99466	2.45
11	.348 (1.04)	15.747 (3.60)	T, $all_{-1,-2,-3}$	311.57**	.00395	.99645	2.23
12	.391 (1.68)	28.198 (5.55)	T, $all_{-1,-2,-3,-4}$	1328.21**	.00260	.99847	2.42

Notes: $\hat{Y}$ = forecast of real GNP from FAIR-OLD.
See Notes, Table 1.

TABLE 4

Estimates of Equation (1) and Tests of H_{01} and H_{02}.

Results for AC. Dependent Variable is log Y.
Sample period = 1972 II - 1986 II, 57 observations.

Equation	log $\hat{Y}$	const.	Other Variables	WALD	SE	R^2	DW
1	1.000 (93.95)	−.002 (0.02)		0.00	.00942	.99090	2.18
2	.563 (2.10)	.760 (1.64)	T, log Y_{-1}, ..., log Y_{-4}	4.57	.00908	.99154	2.09
3	.922 (16.86)	.423 (1.19)	T, log P_{-1}, ..., log P_{-4}	7.20	.00906	.99158	2.12
4	.751 (7.38)	1.849 (2.44)	T, U_{-1}, ..., U_{-4}	15.47*	.00837	.99282	2.15
5	.932 (21.52)	.455 (1.47)	T, log PM_{-1}, ..., log PM_{-4}	21.22**	.00883	.99201	2.16
6	.933 (22.11)	−.234 (0.71)	T, log W_{-1}, ..., log W_{-4}	8.77	.00894	.99181	2.04
7	.922 (21.75)	.228 (0.59)	T, log $M1_{-1}$, ..., log $M1_{-4}$	9.38	.00913	.99147	2.14
8	.840 (14.98)	1.166 (2.82)	T, r_{-1}, ..., r_{-4}	20.08**	.00839	.99279	2.14
9	.585 (1.86)	3.400 (1.43)	T, all_{-1}	15.75	.00824	.99304	2.12
10	.065 (0.21)	5.967 (1.93)	T, $all_{-1,-2}$	59.60**	.00680	.99526	1.97
11	.153 (0.58)	9.830 (2.55)	T, $all_{-1,-2,-3}$	156.11**	.00596	.99635	2.05
12	.248 (0.89)	7.515 (1.50)	T, $all_{-1,-2,-3,-4}$	235.78**	.00536	.99706	2.02

Notes: $\hat{Y}$ = forecast of real GNP from AC model.
See Notes, Table 1.

observations. Comparing equation 1 in Tables 1 and 4, it can be seen that the AC model is not as accurate as FAIR-CUR. The estimated standard error is .00942 for AC and .00835 for FAIR-CUR. Also, the estimates of γ are insignificant in equations 9-12 for AC, which is not the case for FAIR-CUR. It is interesting, however, that the estimates of γ for the first 8 equations in Table 4 are significant, which means that the AC model does carry information not in the Z_t vector and suggests that part of the success of the Fair model in forecasting might come from this disaggregation.

The best of the VAR models was VAR2, and so the results for VAR2 will be emphasized here. The results for VAR2 are presented in Table 5, which is the same as Tables 1 and 4 except for a different model. The results for VAR2 are not as good as those for AC (and thus a fortiori for FAIR-CUR). For example, the estimate of γ is not significant in equation 2, where the lagged values of GNP are added. Also, the standard error of equation 1 is greater for VAR2 than it is for AC. Equations 11 and 12 are not presented for VAR2 because for these equations the number of variables in the information set exceed the number of variables in the VAR2 model. The model is quite likely to be dominated in these cases (which it was). In equation 10 the variables in the information set and the variables in the model are the same. We know in this case that the forecast is likely to be dominated by the information variables, and our monte-carlo experiments suggest that the coefficient of the forecast is likely to be negative. This is what we indeed observe in Table 5, equation 10. The estimate of γ is -.487, with a t-statistic of 1.16 in absolute value.

TABLE 5

Estimates of Equation (1) and Tests of H_{01} and H_{02}.

Results for VAR2. Dependent Variable is log Y.
Sample period = 1972 II - 1986 II, 57 observations.

Equation	log $\hat{Y}$	const.	Other Variables	WALD	SE	R^2	DW
1	.954 (101.18)	.377 (4.96)		24.65**	.00963	.99049	1.66
2	.417 (1.49)	.703 (1.44)	T, log Y_{-1}, ..., log Y_{-4}	35.68**	.00922	.99129	2.00
3	.952 (14.63)	.295 (0.74)	T, log P_{-1}, ..., log P_{-4}	30.20**	.00939	.99095	1.71
4	.845 (6.39)	1.180 (1.21)	T, U_{-1}, ..., U_{-4}	32.62**	.00924	.99124	1.75
5	1.016 (16.72)	−.047 (0.11)	T, log PM_{-1}, ..., log PM_{-4}	38.06**	.00943	.99089	1.83
6	.969 (19.43)	−.017 (0.05)	T, log W_{-1}, ..., log W_{-4}	32.05**	.00918	.99136	1.82
7	.995 (16.13)	.422 (1.05)	T, log $M1_{-1}$, ..., log $M1_{-4}$	32.43**	.00946	.99083	1.67
8	.830 (11.99)	1.278 (2.52)	T, r_{-1}, ..., r_{-4}	48.65**	.00870	.99223	1.73
9	.435 (1.52)	2.195 (1.03)	T, all_{-1}	55.32**	.00849	.99262	2.06
10	−.487 (1.16)	5.416 (1.84)	T, $all_{-1,-2}$	105.53**	.00673	.99536	1.93

Notes: $\hat{Y}$ = forecast of real GNP from VAR2 model.
See Notes, Table 1.

Comparisons of the Forecasts

The results in Table 6 compare the forecasts from the various models. For each equation the log of real GNP is regressed on a constant term and the logs of two or more forecasts. (These equations are also estimated using the White correction for heteroskedasticity.) One forecast is said to dominate another if its coefficient estimate is significant and the other's is not. Equations 1 through 8 in Table 6 are for the longer sample period of 57 observations, and equations 9 through 14 are for the shorter period of 40 observations. Remember that all the forecasts are based on rolling regressions, and so all of them use only information through the previous period for the forecast of the current period.

Equations 1 through 4 show that FAIR-CUR dominates the AC, VAR1, VAR2, and VAR4 models.[22] Equations 5 and 7 show that AC dominates VAR1 and VAR4. In equation 6 both the AC and VAR2 forecasts are significant, although the AC forecasts has a larger coefficient estimate and a higher t-statistic. In equation 8 FAIR-CUR dominates both AC and VAR2. It is interesting to note in this case that even though AC is better than VAR2 in equation 6, VAR2 has a larger coefficient estimate in equation 8 than does AC (although both estimates are insignificant). This is an indication that the AC forecast is correlated more with the FAIR-CUR forecast than is the VAR2 forecast. This is as expected since both FAIR-CUR and AC estimate equations for the components of GNP.

Equations 9 through 13 show that FAIR-OLD dominates AC, VAR1, VAR2, and

---

[22]Cooper and Nelson (1975) also found coefficients near one for a structural forecast of real GNP from the FMP model in analogous regressions were the autoregressive forecasts were univariate autoregressions for real GNP.

TABLE 6

Regression of the log of actual GNP on the log of the forecasts from two or more models.

Results for All Models. Dependent Variable is log Y

Equation	const.	FAIR-CUR ($\log \hat{Y}$)	FAIR-OLD ($\log \hat{Y}$)	AC ($\log \hat{Y}$)	VAR1 ($\log \hat{Y}$)	VAR2 ($\log \hat{Y}$)	VAR4 ($\log \hat{Y}$)	SE	DW	Sample
1	.093 (1.26)	.779 (4.43)		.209 (1.16)				.00827	2.10	722-862
2	.154 (1.98)	.836 (4.48)			.144 (0.80)			.00828	2.06	722-862
3	.189 (2.44)	.697 (4.43)				.279 (1.85)		.00807	2.05	722-862
4	.179 (2.15)	.775 (5.32)					.202 (1.46)	.00815	1.99	722-862
5	.097 (0.80)			.704 (3.15)	.284 (1.34)			.00922	2.12	722-862
6	.161 (1.43)			.546 (3.09)		.434 (2.62)		.00894	2.02	722-862
7	.125 (1.04)			.678 (3.87)			.306 (1.87)	.00905	2.05	722-862
8	.186 (1.84)	.690 (3.92)		.013 (0.07)		.274 (1.57)		.00807	2.06	722-862
9	.103 (0.59)		.726 (2.44)	.261 (0.83)				.00861	2.16	763-862
10	.178 (1.37)		.973 (3.96)		.004 (0.02)			.00870	2.06	763-862
11	.276 (2.30)		.703 (3.60)			.263 (1.39)		.00851	2.03	763-862
12	.253 (2.03)		.744 (4.11)				.225 (1.30)	.00850	1.91	763-862
13	.230 (1.00)		.633 (2.20)	.113 (0.30)		.225 (0.96)		.00850	2.08	763-862
14	.197 (1.85)	.330 (0.57)	.645 (1.10)					.00864	2.06	763-862

VAR4. The conclusion regarding FAIR-OLD is thus the same as the conclusion regarding FAIR-CUR. In equation 14 both the FAIR-CUR and FAIR-OLD forecasts are included. Neither coefficient estimate is significant, but FAIR-OLD has a coefficient estimate that is about twice the size of the FAIR-CUR coefficient estimate. FAIR-OLD is thus slightly better than FAIR-CUR in this sense. This is encouraging in that it shows that the changes in structure of the Fair model made between 1976-II and 1986-II did not contribute to the success it has in forecasting over this period.

First Differenced Results

Note that the regressions in Table 6 do not include a time trend. If the forecasts must be differenced to induce stationarity, then the regressions are in the form of "cointegrating regressions" (Engle and Granger [1987]), for which the usual asymptotic distribution theory does not apply (Phillips and Durlauf [1986]). We reestimated the equations using differenced data, and the results are presented in Table 7. Table 7 is the same as Table 6 except that the dependent variable is $\log Y_t - \log Y_{t-1}$ and the forecast variables $\log \hat{Y}_t$ are replaced by $\log \hat{Y}_t - \log Y_{t-1}$. The results in Table 7 are very similar to those in Table 6. The main difference is that in Table 7 the VAR models do not do quite as well against the others as they do in Table 6.

As a final test for FAIR-CUR, we estimated equation (1) in first differenced form. We regressed $\log Y_t - \log Y_{t-1}$ on $\log \hat{Y}_t - \log Y_{t-1}$, a constant, and the first differences of the various variables. The results are presented in Table 8. Table 8 is the same as Table 1 except for the different functional forms. The results in Table 8 are similar to those in

TABLE 7

Regression of actual log GNP changes on forecasted changes from two or more models.

Results for All Models. Dependent Variable is $\Delta \log Y$.

Equation	const.	FAIR-CUR	FAIR-OLD	AC	VAR1	VAR2	VAR4	SE	DW	Sample
		$\log \hat{Y} - \log Y_{-1}$								
1	-.0027 (1.46)	.736 (4.93)		.284 (1.54)				.00835	2.12	722-862
2	-.0018 (1.14)	.889 (5.31)			.030 (0.17)			.00847	1.95	722-862
3	-.0014 (0.93)	.804 (5.24)				.174 (1.22)		.00837	1.96	722-862
4	-.0018 (1.17)	.834 (5.22)					.109 (1.00)	.00841	1.92	722-862
5	-.0004 (0.19)			.755 (4.09)	.183 (1.08)			.00927	2.11	722-862
6	.0001 (0.07)			.669 (3.95)		.292 (2.29)		.00908	2.06	722-862
7	-.0006 (0.30)			.711 (3.88)			.214 (1.84)	.00913	2.04	722-862
8	-.0022 (1.27)	.679 (4.19)		.242 (1.43)		.142 (1.07)		.00828	2.12	722-862
9	-.0044 (1.44)		.751 (2.91)	.427 (1.57)				.00860	2.31	763-862
10	-.0032 (1.20)		1.047 (4.70)		-.097 (0.40)			.00880	2.02	763-862
11	-.0027 (0.99)		.880 (3.60)			.137 (0.71)		.00877	2.02	763-862
12	-.0030 (1.06)		.856 (3.68)				.144 (0.92)	.00874	1.92	763-862
13	-.0040 (1.43)		.683 (2.39)	.405 (1.53)		.101 (0.56)		.00856	2.28	763-862
14	-.0035 (1.28)	.257 (0.41)	.769 (1.24)					.00879	2.05	763-862

TABLE 8

Estimates of Equation (1) and Tests of H_{01} and H_{02}.

Results for FAIR-CUR. Dependent Variable is $\Delta \log Y$.
Sample period = 1972 II - 1984 II, 57 observations.

Equation	$\log \hat{Y} - \log Y_{-1}$	const.	Other Variables	WALD	SE	R^2	DW
1	.904 (6.85)	−.0029 (1.16)		1.34	.00847	.40840	1.93
2	.873 (8.05)	−.0033 (2.10)	$\Delta\log Y_{-1}, \ldots, \Delta\log Y_{-4}$	8.00	.00819	.44708	2.16
3	.880 (6.59)	.0037 (1.09)	$\Delta\log P_{-1}, \ldots, \Delta\log P_{-4}$	6.47	.00819	.44626	2.03
4	.746 (5.68)	−.0003 (0.18)	$\Delta U_{-1}, \ldots, \Delta U_{-4}$	9.40	.00766	.51558	2.31
5	.809 (4.83)	−.0001 (0.03)	$\Delta\log PM_{-1}, \ldots, \Delta\log PM_{-4}$	10.91	.00825	.43897	1.96
6	.912 (6.99)	.0025 (0.64)	$\Delta\log W_{-1}, \ldots, \Delta\log W_{-4}$	6.61	.00804	.46667	1.92
7	.919 (6.28)	−.0009 (0.33)	$\Delta\log M1_{-1}, \ldots, \Delta\log M1_{-2}$	6.39	.00833	.42739	1.87
8	.786 (5.25)	−.0008 (0.52)	$\Delta r_{-1}, \ldots, \Delta r_{-4}$	22.14**	.00804	.46667	1.93
9	.933 (8.03)	−.0003 (0.11)	all_{-1}	23.09**	.00749	.53736	1.85
10	.709 (4.87)	−.0012 (0.35)	$all_{-1,-2}$	64.97**	.00704	.59098	2.07
11	.725 (4.52)	−.0025 (0.37)	$all_{-1,-2,-3}$	88.88**	.00651	.65069	1.95
12	.903 (5.54)	−.0012 (0.14)	$all_{-1,-2,-3,-4}$	244.25**	.00602	.70113	2.10

Notes: $\hat{Y}$ = forecast of real GNP from FAIR-CUR.
See Notes, Table 1.
all = $\Delta\log Y$, $\Delta\log P$, ΔU, $\Delta\log PM$, $\Delta\log W$, $\Delta\log M1$, Δr.

Table 1. The coefficient estimate of the forecast variable is always significant. First differencing thus makes little difference.

IV. Conclusion

A general method has been proposed in this paper for examining models as information aggregators (the IA test) and for comparing alternative models. The IA test results show that the Fair model is a useful aggregator of information. The overall results also show that the Fair model dominates the nontheoretical AC and VAR models and that the AC model tends to dominate the VAR models. Since the AC model resembles the Fair model in its use of lagged values of the components of GNP, this suggests that some but not all of the information that the Fair model uses to dominate the VAR models is in the lagged values of the components.

References

Chong, Yock Y., and David F. Hendry, "Econometric Evaluation of Linear Macro-Economic Models," Review of Economic Studies, 53: 671-90, August 1986.

Cooper, J. Phillip, and Charles R. Nelson, "The Ex-Ante Prediction Performance of the St. Louis and FRB-MIT-Penn Econometric Models and Some Results on Composite Predictions," Journal of Money, Credit, and Banking, 7: 1-32, February 1975.

Engle, Robert F., and C.W.J. Granger, "Cointegration and Error Correction: Representation, Estimation and Testing," Econometrica, 55: 251-76, March 1987.

Fair, Ray C., A Model of Macroeconomic Activity, Vol. 2, The Empirical Model, Ballinger Publishing Co., 1976.

Fair, Ray C., "The Sensitivity of Fiscal Policy Effects to Assumptions About the Behavior of the Federal Reserve," Econometrica, 46: 1165-79, 1978.

Fair, Ray C., Specification, Estimation, and Analysis of Macroeconometric Models, Harvard University Press, 1984.

Granger, C.W.J., and Paul Newbold, Forecasting Economic Time Series, 2nd ed., Academic Press, 1986.

Hendry, David F., and Jean-Francois Richard, "On the Formulation of Empirical Models in Dynamic Economics," Journal of Econometrics, 20: 3-33, October 1982.

Litterman, Robert B., "Forecasting with Bayesian Vector Autoregressions--Four Years of Experience," Federal Reserve Bank of Minneapolis Working Paper 259, August 1984.

Litterman, Robert B., "Economic Forecasts From a Vector Autoregression," xeroxed, serial.

Liu, T. C., "Underidentification, Structural Estimation, and Forecasting," Econometrica, 28: 855-65, 1960.

McNees, Stephen K., "The Methodology of Macroeconometric Model Comparisons," in J. Kmenta and J.B. Ramsey, eds., Large Scale Macro-Econometric Models, North-Holland, 1981, 397-422.

Nelson, Charles R., "The Prediction Performance of the FRB-MIT-Penn Model of the U. S. Economy," American Economic Review, 62: 902-17, December 1972.

Phillips, P.C.B., and S.N. Derlauf, "Multiple Time Series Regressions with Integrated Processes," Review of Economic Studies, 53: 473-95, August 1986.

Ramsey, J.B., "Tests for Specification Errors in Classical Least-Squares Linear Regression Analysis," Journal of the Royal Statistical Society, Series B, 31: 350-71, 1969.

Sims, Christopher A., "Macroeconomics and Reality," Econometrica, 48: 1-48, January 1980.

Sims, Christopher A., "Book Review," (of Fair [1984]) Journal of Money, Credit and Banking, 18: 121-6, February 1986.

Sims, Christopher A., "Economic Forecasts From a Vector Autoregression," xeroxed, serial.

Theil, Henri, "The Analysis of Disturbances in Regression Analysis," Journal of the American Statistical Association, 60: 1067-79, December 1965.

Theil, Henri, Principles of Econometrics, John Wiley, New York, 1971.

White, Halbert, Asymptotic Distribution Theory for Econometricians, Academic Press, Orlando Florida, 1984.