Volume Title: Forecasts with Quarterly Macroeconomic Models

Volume Author/Editor: Yoel Haitovsky, George Treyz, and Vincent Su

Volume Publisher: NBER

Volume ISBN: 0-870-14266-6

Volume URL: http://www.nber.org/books/hait74-1

Publication Date: 1974

Chapter Title: Introduction and Summary

Chapter Author: Yoel Haitovsky, George Treyz, Vincent Su

Chapter URL: http://www.nber.org/chapters/c3632

Chapter pages in book: (p. 1 - 22)

Part I

MODELS AND SIMULATIONS

1

Introduction and Summary

1.1 ECONOMETRIC MODELS

Professor Jan Tinbergen laid down the general framework for building macroeconometric models about forty years ago in his pioneering formulation of a model for the U. S. economy.[1] An econometric model of the economy is composed of an interconnected system of equations, each of which describes a sector or a feature of the economy. Some equations are based on the behavior of decision-making units in the economy, such as consumers or investors; some set forth adjustment mechanisms, such as market clearance; and some represent technological or institutional relations, such as production or tax revenue functions. All of these are called "behavioral" equations. They are meant to describe causality in the system to the extent that causal relationships can be formulated, and they all have stochastic components. In addition to the behavioral equations—also termed "structural" inasmuch as they describe the structural characteristics of the economy as depicted by the model builder—the system also includes "definitional" equations, or "identities."

The variables used in the model are divided into two broad categories: "endogenous" and "exogenous." The former are determined

[1] Jan Tinbergen, *Statistical Testing of Business-Cycle Theories, II: Business Cycles in the United States of America, 1919–1932*, League of Nations, Geneva, 1939.

within the model, given the values of the exogenous (determining) variables that are the input variables for the system. Within the set of endogenous variables it is customary to distinguish between lagged and contemporaneous variables. All exogenous variables, together with the lagged endogenous variables, are termed "predetermined."

The endogenous variables typically include the "target" variables in the model. These represent economic quantities which are the subjects of economic policy—the unemployment rate, the price level, the components of expenditure (from the demand side), the amount of disposable income (from the supply side), and some of the other variables determined by the model. The set of exogenous variables is customarily divided into two mutually exclusive subsets: (a) the "controlled" (or instrument) variables and (b) the "uncontrolled" variables. The former are those economic quantities that serve the government as instruments for operating on the targets to achieve economic goals, and the latter comprise all the rest.

Econometric models may serve several purposes. On the one hand, they are of scientific value: they enhance our theoretical understanding of how complex interrelated economic systems operate and aid the economic historian in describing a historical period.[2] On the other hand, they may assist in the governmental decision-making process as a tool in projecting the economic consequences of alternative policy measures. Econometric models are also used for unconditional forecasting. In this capacity they not only help the business community to make decisions but may also help the policy maker by occasionally producing forecasts that imply the need for a change in government policy.

1.2 FORECASTING

Forecasts with econometric models are normally made by estimating the coefficients (or weights) in the model and then solving the system of equations for given values of the predetermined variables. This can be done only if the equation system is a "closed" model. A model cannot be closed unless there is a structural equation corresponding to each endogenous variable. If the model is linear and closed, each endogenous

[2] See Marc Nerlove, "Notes on the Production and Derived Demand Relations Included in Macro-econometric Models," *International Economic Review*, Vol. 8, No. 2, June 1967, pp. 223–42, for the advantages and shortcomings of this use.

variable can be expressed as a function of the predetermined variables. This form of the model is called the "reduced form." Thus, the reduced form derived from the structural model can be used for forecasting from a closed linear model.[3] The discrepancy between the forecast value of an endogenous variable and its realized value is called the "forecasting error." If, on the other hand, we insert the realized values for both the endogenous and the exogenous variables in the structural equation, we get the "structural equation residual" (*SER*). The latter concept will prove useful in the decomposition of the forecasting errors. If the model is nonlinear, as most econometric models now are, including those under consideration here, the model solution is achieved by an iterative procedure.[4]

Typically, however, the forecasting process is not purely mechanical and devoid of the forecaster's judgment. The model's forecast can be influenced by the introduction of judgmental factors that operate directly on the model coefficients. The most common of these adjustments, made to the constant terms (intercept) of the structural equations, is called "adjustment to the constant term of the equation" or, in short, "constant adjustment."

The rationale behind this type of adjustment can be easily explained. Each behavioral equation includes an additive disturbance term to capture the randomness of the economic relationships expressed by that equation. This disturbance term is not observable in principle but can be estimated as a residual (*SER*) when all other quantities in the equation are observed or estimated. When the forecaster regards these disturbances as random or does not have any knowledge of future disturbances,

---

[3] The parameters of the reduced form referred to here are estimated by methods that reflect the restrictions on the system implied by the structural equations. An alternative approach to forecasting would be to directly estimate the relationship between certain exogenous variables and the endogenous variable in question without specifying the structural relationships of the system. We call the latter model a "reduced form model" or an "unrestricted reduced form." The issues involved in choosing between these alternative forecasting procedures are discussed by Ta-Chung Liu in "Underidentification, Structural Estimation, and Forecasting," *Econometrica*, Vol. 28, No. 4, October 1960, pp. 855–65.

[4] See M. K. Evans and L. R. Klein, *The Wharton Econometric Forecasting Model*, Studies in Quantitative Economics No. 2, Economics Research Unit, Wharton School of Finance and Commerce, Philadelphia, 1967, Chapter 4. This method of solution may converge to a "wrong" root under certain circumstances. See Benjamin Friedman, "Econometric Simulation Difficulties: An Illustration," *Review of Economics and Statistics*, Vol. 53, No. 4, November 1971, pp. 381–84.

they might be set equal to zero for the forecast. This is not the case when a systematic pattern of *SER*s is observed and expected to continue into the future, or when the forecaster predicts that factors which are not included in his behavioral equation will change the dependent variable in the equation in a predictable way. An example of the latter is a dock strike, which reduces imports below the expected "normal" level during the strike and increases imports above this level in the poststrike period to absorb the backlog. If the forecaster predicts a strike, its length and severity, and then its recovery period, he may adjust the appropriate equations accordingly. This calls for inserting nonzero disturbances, which, in turn, become constant adjustments.

Another typical use of constant adjustments is to compensate for data revisions. As a first approximation, equations are shifted upward or downward in accordance with the *SER*s resulting from the data shifts. In addition, forecasters will quite often change slope coefficients. For instance, a change in tax laws usually leads to changes in the coefficients of the tax revenue equations.

1.3 EX POST VERSUS EX ANTE FORECASTS

Econometric models are conditional in nature. They are designed to yield accurate solutions for the endogenous variables conditional on the correct values of the exogenous variables. However, the values of the exogenous variables contemporaneous with the endogenous variables are not available at the time of the forecast and hence guesses about their future values must be provided by the forecasters. When the correct values of the exogenous variables are inserted "after the fact" the resulting forecasts are called "ex post" forecasts, while the forecasts that use the guessed values of the exogenous variables are called "ex ante" forecasts. Therefore, ex post forecasts are conditional forecasts, while ex ante forecasts are unconditional in the sense that the forecaster's judgment about the future development of the exogenous variables is an integral part of the forecasting process.

Both conditional and unconditional forecasts are typically made with a model that has been altered by the insertion of adjustments to the constant terms in the stochastic equations. However, the extent of such equation adjustment may vary greatly. Some econometricians use no adjustments at all, or only mechanical adjustments, to account for the cyclical patterns in the structural equation residuals. Others use equation adjustments to introduce exogenous factors that are not included in the specification of the model. Finally, some forecasters allow complete interaction between the model and their guesses. The latter view is expressed forcefully by P. J. Verdoorn of the Dutch Central Planning Bureau:

In practice, model forecasts are but seldom uniquely based on a straightforward solution of the system. Usually the straightforward or "provisional" output is used as the input for the formation of expert opinion. Feed-back of this opinion, together with other relevant independent information into the model results, after a process of interaction, in a new set of values for the predictions. This new set, then, is at the same time compatible with the existing expert opinion on future developments, and consistent with the restrictions and observed behavior pattern of the social and economic system as reflected by the structural equations. Apart from being a mere mathematical forecasting tool, an econometric model, therefore, serves too as a vehicle for the consistent allocation and processing of such available information as was not originally contained in the model.[5]

Although it would be most interesting to analyze forecast development using the interaction procedure, it is impossible to obtain forecasting data appropriate for it. On the other hand, the constant adjustments that were actually employed by the forecasters were obtained for relatively long periods for the two models under investigation here. When these constant adjustments are used in conjunction with the realized values of the exogenous variables included in the model, the forecast is called the "ex post *OR*" (original) adjustment forecast. When they are used with the forecaster's guesses about these exogenous values, the prediction is the "ex ante *OR*" forecast.

We take the position in this monograph that the true test of the accuracy of a model's forecasts is the accuracy of its ex post *OR* forecasts. We accept the *OR* adjustments, despite their involving some judgment on the part of the forecasters, because we recognize the fact that there are unusual occurrences in the economy and that the effect of these events could not be consistently and reliably estimated by reference to past data. The forecaster wants to incorporate the effect of these probable future events in his forecasts and does so by making

[5] P. J. Verdoorn, "The Feasibility of Long-Term Multi-Sector Forecasts of Manpower Requirements by Econometric Models," in O.E.C.D. Conference on Forecasting Manpower Requirements, unpublished, May 1970, p. 1.

equation adjustments.[6] We choose the ex post rather than the ex ante forecast as the true test because failure by a model to make accurate predictions with the forecaster's bad guesses as to the explicit exogenous variables is by no means a proof of the inaccuracy of the model. One can even envisage a situation, admittedly a rare one, in which the monetary or fiscal authorities would markedly change their policy as a result of a forecast based on their intended policy. Use of the ex ante forecast as a criterion in this situation would have made the model seem inaccurate.

It should be noted, however, that the ex post forecast cannot always serve as a criterion in comparing the relative accuracy of different forecasting models. For example, models might differ from each other by the size of their exogenous variables set. In this case the ex post forecasts would tend to favor a model with a large and important exogenous variables set, since, by definition, all exogenous variables will appear at the correct values.

When we consider that an important use of models is aiding policy makers, it is evident that the twin requirements of accurate ex post forecasts and reliable evaluation of alternative policy measures are closely related. The policy maker must first be able to forecast reliably future economic events before determining whether a new policy is desirable. He then needs to forecast the possible results of the new policy. Given that experimentation in the economy is impossible, the ultimate test of an econometric model as a policy aid is its accuracy in predicting the course of the economy, conditional on the exogenous values chosen by the policy makers.

1.4 MECHANICAL CONSTANT ADJUSTMENTS

The *OR* (original) adjustment forecasts can be contrasted with some mechanical methods of adjusting the constants of the behavioral equations. These are mainly designed to take account of the cyclical pattern of the *SER*s (structural equation residuals). We have tried two

[6] Marschak argues that a primary reason for building a structural model is to be able to make economic predictions when a coming structural change can be anticipated. It is then that forecasting is impossible without some knowledge of structure. See Jacob Marschak, "Economic Measurements for Policy and Prediction," in W. C. Hood and T. C. Koopmans (eds.), *Studies in Econometric Method*, Cowles Commission for Research in Economics Monograph 14, New York, John Wiley and Sons, 1953, pp. 1–26.

such mechanical constant adjustments. The first, *AR* (average residual adjustment), which is the kind of mechanical constant adjustment made by Wharton (see below), simply subtracts the average of the last two *SER*s from the forecasting equations; the adjustment does not change with the forecasting span. Any systematic improvement of the *OR* constant adjustment over the latter type is attributable to the Wharton forecaster's judgment. The second type of mechanical constant adjustment, *GG* (for Goldberger-Green adjustment), originated at OBE (see below) and follows Goldberger,[7] who proves that, when the cyclical pattern of the residuals can be specified, an optimal forecasting procedure will assign nonzero values to these terms in the forecasting period. The specific constant adjustment adopted by Green[8] is

$$\rho^T \frac{(e_t + \rho e_{t-1})}{2},$$

where T is the forecasting span, e_t and e_{t-1} are the last two observed *SER*s, and ρ is the estimated autocorrelation coefficient of the *SER*s of the equation in question when the cyclical pattern of these *SER*s can be characterized by a first order Markov scheme. Here the observed *SER*s carry geometrically declining weights as the forecasting span becomes longer. Finally, in addition to these two mechanical adjustments, the model has also been solved with no constant adjustments (*NO*).[9]

We have used these constant adjustments to facilitate our analysis of the models under consideration here, in the full knowledge that some econometricians prefer adding lagged values to their model and using estimating techniques that reduce autocorrelation instead of using constant adjustments in forecasting. However, the introduction of extra lagged values may cause distributed lagged bias in the coefficient estimates, and this may have an important impact on the validity of the model. Procedures for reducing autocorrelation in the sample period may

[7] A. A. Goldberger, "Best Linear Unbiased Prediction in the Generalized Linear Regression Model," *Journal of the American Statistical Association*, Vol. 57, No. 2, June 1962, pp. 369–75.

[8] George R. Green, in association with Maurice Liebenberg and Albert A. Hirsch, "Short- and Long-Term Simulations with the OBE Econometric Model," in Bert G. Hickman, ed., *Econometric Models of Cyclical Behavior*, Vol. 1, New York, NBER, 1972, p. 32.

[9] The only exception to this was the labor force equation for Wharton, where the forecast of unemployment without adjustment was often wrong by several percentage points. Therefore, we adopted (for *NO*, *AR*, and *GG*) a mechanical approximation of the adjustment that was made in all ex ante forecasts (see Chapter 5, footnote 5.)

not eliminate the need for adjustment in the forecast period, when the need for constant adjustments arises from such factors as shifts in data series, structural shifts in equations, or exogenous events that influence an endogenous variable but are not included in the model.

1.5 DECOMPOSITION OF FORECAST ERROR

The ability to create ex post and ex ante forecasts with various constant adjustments can help one understand the nature of the forecasting performance observed. These insights can be augmented by decomposing the forecast error. This can be done by tracing error in ex post forecasts to the *SER*s. One can also examine the extent to which the *SER* error is mitigated by mechanical constant adjustment and by the forecaster's judgmental equation adjustments. For each endogenous variable, the forecast error can be separated into the direct error caused by unadjusted *SER* error in the equation for this variable and the part of the error attributable to the rest of the system, including the reverberations of the direct error throughout the system. The decomposition of ex post error also allows us to determine which part of the error can be attributed to errors in lags in multiperiod forecasts. By tracing the effect of the difference between the guessed-at values of specific exogenous variables and their ex post values we can explain the difference between ex post and ex ante forecast error.

1.6 STANDARDS OF COMPARISON

We have stated before that it is desirable that econometric models yield reliable ex post forecasts. This is particularly important in the case of policy models, for, if they do not yield accurate ex post forecasts, the model multipliers may not represent the "true world" multipliers accurately enough and thus cannot be reliably used as policy guides. In general, policy models must pass more stringent tests than short-run forecasting models. While the latter can draw more heavily on historical regularities, the former must be able to estimate the effects of policy changes in situations in which the policy aim is to depart from historical regularities when these proceed in an undesirable direction. This distinction calls for testing additional properties that might be crucial for policy models but not necessarily for short-run forecasting models. For instance, a test of the stability properties (that is, whether or not the stability properties of the model conform to the model builder's preconception of what they ought to be) is very important for policy models and less so for short-run forecasting models.[10] Or, one might find a model useful only if the policy instruments one wishes to investigate appear in the model in a manner appropriate to one's particular aims.[11] However, in this monograph we confine the scope of our evaluation to forecast performance; structural properties as such will not be explicitly investigated.

In order to evaluate the econometric forecast record we need standards of comparison. Judgmental ex ante forecasts are one of the standards of comparison that we can use for ex ante forecast performance. Naive model extrapolations that are based solely on the past behavior of the series we wish to extrapolate can be used as performance references for both ex post and ex ante forecasts.

We have defined three types of naive models: (a) a fourth order autoregression, i.e.,

$$Y_{t+1} = a_0 + a_1 Y_t + a_2 Y_{t-1} + a_3 Y_{t-2} + a_4 Y_{t-3},$$

where t is the last observed value (i.e., the period before the first quarter of the forecast); (b) a "no change" naive model (or "*Naive 1*") in which the "forecast" assumes the observed value in the jump-off period, i.e.,

$$a_1 = 1,\ a_0 = a_2 = a_3 = a_4 = 0$$

in the equation above; (c) a "same change" naive model (or "*Naive 2*"), in which the "forecast" is derived by adding to the last observed value the last observed change, i.e.,

$$a_1 = 2,\ a_2 = -1,\ a_0 = a_3 = a_4 = 0.$$

[10] Studies on the stability properties of econometric models were reported by H. Theil and J. C. G. Boot in "The Final Form of Econometric Equation Systems," *The Review of the International Statistical Institute*, Vol. 30, 1962, pp. 136–52, and more recently by G. C. Chow and R. E. Levitan in "Nature of Business Cycles Implicit in a Linear Economic Model," *The Quarterly Journal of Economics*, Vol. 83, 1969, pp. 504–17, as well as E. P. Howrey in "Dynamic Properties of a Condensed Version of the Wharton Model," in Hickman, ed., *Econometric Models of Cyclical Behavior*, Vol. 1.

[11] For instance, the coefficients in the equations relating income and corporate profits to tax revenues may be estimated in such a way as to yield some average effective tax coefficient. This procedure would be unsatisfactory for the policy maker who might find it hard to convert a proposed new tax schedule to an average effective tax coefficient.

In performing multiperiod extrapolations with the naive models, the predicted values simply replace, in succession, the observed values on the right hand side of each equation. This is equivalent to multiplying the last observed change by the forecasting span and adding it to the last observed value in the "same change" model. It means using the last observed value for all successive periods in the "no change" version.

Another interesting yardstick for comparison is the reduced form model proposed by Leonall C. Andersen and Jerry L. Jordan in the *Federal Reserve Bank of St. Louis Review* in November 1968.[12] Andersen and Jordan make conditional forecasts with a single equation in which nominal GNP is a function of the money supply and the difference between high employment government revenue and expenditure. The values in the current and three last quarters are used for both variables. The structure of the lags presented in the regression was estimated by the Almon lag technique restricted to a fourth degree polynomial. This GNP equation, with an additional lagged value for each variable, is included in the more elaborate model proposed in the *Review* in November 1970.[13] In order to minimize the bias inherent in ex post model specification, we have used the earlier model in this monograph. The coefficient values for both the fourth order autoregressive and the St. Louis equation are estimated over sample periods that match the sample periods of the structural models in our comparison.

We have also carried out some sample period simulations over a period of trend growth as well as over a period of fluctuation to see whether the performance of the models relative to the standards of comparison was strongly influenced by the recession-free nature of our forecast period.

In keeping with current forecasting practice, we use point estimates for the parameters of the econometric models in this study. Since the coefficients of the models are estimated on the basis of a sample, these parameters are only known in a probabilistic sense. Therefore, forecasts of the distribution of possible outcomes for the endogenous variables might be more appropriate and informative than the point projections

[12] Leonall C. Andersen and Jerry L. Jordan, "Monetary and Fiscal Actions: A Test of Their Relative Importance in Economic Stabilization," *Federal Reserve Bank of St. Louis Review*, Vol. 50, No. 11, pp. 11–23.

[13] Leonall C. Andersen and Keith M. Carlson, "A Monetarist Model for Economic Stabilization," *Federal Reserve Bank of St. Louis Review*, Vol. 52, No. 4, pp. 7–25.

currently made.[14] If we had, for endogenous variables, ex ante forecasts that included not only point predictions but also confidence interval estimates, we could use these probability distributions as a standard for accepting or rejecting the structural specifications set forth in the model. The constant term and other adjustments described above would complicate this line of investigation; nevertheless, we would urge econometric forecasters to make probability estimates of future outcomes instead of point predictions as soon as this becomes technically feasible.

1.7 MEASURING FORECAST INACCURACIES

In order to evaluate forecasting performance, it is necessary to have some measure of forecasting inaccuracy. Ideally, this measure should be based on a loss function that reflects the welfare cost of an incorrect decision resulting from forecast errors. In the absence of such a function we use various simpler alternative measures, each implying a particular mathematical form of the loss function. The two most commonly used are the "average absolute forecasting error" (*AAFE*) and the "root mean square error" (*RMS*).

The Average Absolute Forecasting Error (*AAFE*)

This measure is defined as

$$AAFE = 1/N \sum | F_t - R_t |.$$

F_t and R_t are, respectively, the forecast and realized values in period t. The quantity between the two vertical lines is the absolute value of the difference, and N is the number of such forecasts (i.e., $t = 1 \ldots N$). This inaccuracy measure implies a linear loss function, symmetric around the optimal decision—i.e., as the error doubles in absolute value, the loss doubles.

[14] See Y. Haitovsky and N. Wallace, "A Study of Discretionary and Nondiscretionary Fiscal and Monetary Policies in the Context of Stochastic Macroeconometric Models," in Victor Zarnowitz, ed., *The Business Cycle Today*, Fiftieth Anniversary Colloquium I, NBER, 1972; J. Kareken, T. Muench, T. Supel, and N. Wallace, "Determining the Optimum Monetary Instrument Variable," unpublished paper; G. Schink, "An a priori Measure of the Forecast Error in the Wharton Model," paper presented at Wharton Econometric Seminar, June 24, 1971; G. Treyz, "Effects of Alternative Fiscal Policies on the National Economy: A Flexible Econometric Approach," Ph.D. Dissertation, Cornell University, 1967.

The Root Mean Square Error (*RMS*)

This inaccuracy measure is defined as

$$RMS = [1/N \sum | F_t - R_t |^2]^{1/2}.$$

It[15] implies a quadratic loss function. As the error doubles in absolute value, the loss will more than double.

The *RMS* measure has the advantage that it, or rather its square, the mean square error (*MSE*), lends itself to a meaningful and helpful decomposition suggested by Theil.[16] We have

$$(RMS)^2 = MSE = UM + US + UC,$$

where $UM = (\bar{F} - \bar{R})^2,$

and $US = (S_F - S_R)^2$

$$UC = 2(1 - r_{FR})S_F S_R.$$

$\bar{F}$, $\bar{R}$, S_F, S_R and r_{FR} are, respectively, the means and standard errors of the simulated (or forecast) and realized values, and r_{FR} is the correlation between them. *UM*, *US*, and *UC* are called by Theil the "partial coefficient of inequality due to *unequal central tendency*, to *unequal variation*, and to *imperfect covariation*, respectively."

In addition to Theil's decomposition, we suggest a procedure whereby *MSE* for GNP can be separated into parts that are related to structural and stochastic components. The structural component comes from the interdependencies in the model as specified by the model builder with the estimated coefficients. The stochastic component comes from errors in individual equations as well as interdependencies among the disturbance terms.

Two Variants of the *RMS* Measure

In this monograph we use two additional variants of the *RMS* measure:

---

[15] A general measure which includes both *AAFE* and *RMS* as special cases:

$$[1/N \sum | F_t - R_t^k |]^{1/k} = 1, 2, \ldots$$

When $K = 1$ we have *AAFE*, when $K = 2$ we have *RMS*. See Christopher A. Sims, "Evaluating Short-Term Macro-economic Forecasts: The Dutch Performance," *Review of Economics and Statistics*, Vol. 49, 1967, p. 226.

[16] H. Theil, *Economic Forecasts and Policy*, Amsterdam, North-Holland, 1961, p. 35.

1. *RMS* divided by *RMS* of the "no change naive model" forecast, designated in the tables by *RMS/RMS* of *Naive 1*. The purpose of this statistic is twofold. First, the performance of the forecasting method under investigation can be easily compared to that of the "no change" forecast; a value larger than unity for the new statistic immediately cautions the reader that the model forecast performance was inferior to the simplest of all extrapolations. Second, the division by the "no change" *RMS* can be viewed in some sense as a normalization process, normalizing for the erratic behavior of the various series in the period under investigation.

2. *RMS* of per cent error, designated in the tables by *RMS* per cent error, in which the forecasting error in period t is defined as a ratio of $F - R$ to the preceding realized value: $(F_t - R_t)/R_{t-1}$.[17] The reason for doing so is the heteroscedastic nature of many economic series (increase or decrease of residual variance with the increase of the series level). Heteroscedasticity in economic series often can be reduced, if not eliminated altogether, by taking these ratios.

We feel that the last measure is appropriate for most economic series.[18] Since it is particularly important when long series are analyzed or compared, we use it as a preferred measure when comparing the sample period with the post-sample-period error. However, for short forecast periods we have chosen a simple measure—the *AAFE* (average absolute forecasting error). In cases where it conveys all of the information needed we use it as our inaccuracy measure because it is simple.

Turning Point and Acceleration Analysis

The ability of a forecasting procedure to predict turning points or significant acceleration and deceleration is important. An inaccuracy measure that centered on these aspects of forecast error would be appropriate for a loss function where great weight is given to predicting deviations from trend growth or changes in the economy's direction.

---

[17] This measure is extensively used by the Netherlands Central Planning Bureau for evaluating their forecast accuracy. See, for example, Central Planning Bureau, *Forecast and Realization: The Forecasts by the Netherlands Central Planning Bureau* 1953–1963, C. P. B. Monograph No. 10, The Hague, 1965.

[18] Obvious exceptions are series which are already in difference form, such as "change in stocks," or series computed as differences of other series, such as "net foreign balance."

However, to be meaningful, these measures should be based on a substantial number of turning points and significant accelerations and decelerations. Since our aggregate series, such as *GNP* and *GNP58*, were trend-dominated in the forecast period covered, and since our forecast period is short, we have not included summary statistics on turning points or acceleration forecasting error.

1.8 DATA REVISIONS

The frequent revisions in the national account series complicate the evaluation of forecast inaccuracies, since the forecaster uses preliminary data releases for his forecast. Accordingly, the preliminary values will be different for each set of forecasts, and often markedly different from the corresponding revised data set.

Since these preliminary values are the base from which predictions start, and since the forecasters must rely on the data available at the time of the forecast, it would be incorrect to compare the ex ante forecast based on preliminary data with the revised realized values. In order to minimize possible unjust penalties on forecasters, we compare the predicted change in the variable in question with the realized change, by defining the realized value set as the revised realized change added to the lag values actually used by the forecaster. That is, we define

$$R_{t+T} = P_t + A_{t+T} - A_t = P_t + \Delta_T A_t.$$

where R denotes the realized value defined to take account of data revisions, t is the jump-off period (the last period for which data were available and used as point of departure), T stands for the forecasting span, P denotes preliminary values of the variables under investigation known to the forecaster at the time of forecasting, A denotes the corresponding revised values, and Δ_T is the differentiating operator:

$$\Delta_T A_t = A_{t+T} - A_t.$$

The realized values so defined are used to make comparisons with the forecast values and to substitute for the exogenous variable values in ex post forecasts.[19]

[19] George Green has demonstrated that this procedure can result in inconsistent series when price, nominal, and real-value variables are all computed in this way. For example, if, in the

We prefer this procedure to other alternatives because it makes it possible to use the original constant adjustments in the *OR* ex post forecast. This would have been impossible if we had used the revised data, since the *OR* constant term adjustments were made in part to reflect structural equation residuals that were calculated by the forecaster on the preliminary data set.[20]

1.9 SUMMARY

The models and forecasts analyzed and compared in Chapter 2 are the products of the Office of Business Economics (OBE) of the U.S. Department of Commerce and the Wharton School of Finance and Commerce (Wharton) of the University of Pennsylvania. They each contain about fifty behavioral equations and are similar in general structure, with the OBE models emphasizing the government sector and the Wharton models, the private sector.

Our first simulations in Chapter 3 deal with single versions of the OBE and Wharton models over their respective sample periods. We find that the econometric model projections for one year ahead of GNP in current dollars (*GNP*) and in constant dollars (*GNP58*) are superior to either "no change" (*Naive 1*) or "same change" (*Naive 2*) projections. The OBE simulations show a similar result for the unemployment rate, but, surprisingly, the Wharton Model simulations for this variable are inferior to the "no change" (*Naive 1*) projections. The same comparisons for projections one quarter ahead are less favorable to econometric models. The OBE first quarter predictions are superior to the naive projections for *GNP* and *GNP58*, but they are only equivalent for the unemployment rate, while the Wharton results are inferior to the naive projections for the unemployment rate and only about the same for *GNP* and *GNP58*. Use of constant adjustments to the econometric equations

---

revised series, real GNP changes from 1000 to 1020 and prices change from 100 to 102, then nominal GNP changes from 1000 to 1040.4. This is consistent with multiplying 102 × 1020. But if the preliminary value of real GNP was 990 and the preliminary value of the price index was 100, then 1010 × 102 = 1030.2, which is not equal to 1030.4. We obtain this latter number by adding the revised value of the change in nominal GNP to its value in the preliminary series. This inconsistency can be avoided by computing the price series as a ratio of the nominal to the real series (as defined in the above formula) instead of computing all three values by the formula.

[20] In Chapter 3 we do not use the *OR* adjustments. Thus, we can use the revised values for the lags and, in this special case, use our procedure with the A_t value substituted for P_t.

calculated by either one of the two formulas mentioned above leave our observations unchanged. Considering the ease with which we might expect econometric models to outperform naive projections in their sample period, the results are very poor in reproducing quarterly fluctuations (especially for Wharton) and only moderately encouraging in reproducing the annual movements in the economy.

Examining the characteristics of the econometric sample period error, we find that model error in tracking quarterly fluctuations is caused primarily by unequal covariation (*UC*) rather than unequal central tendency (*UM*) or unequal variation (*US*). We also find that the error in GNP is larger than one might expect on the basis of the error for the individual components of GNP. This is evidence of error aggregation over variables. Another type of error aggregation is over time. We find that the errors in consecutive quarters, rather than systematically reinforcing each other, show slight evidence of some error offset.

When we turn from sample period findings to ex post forecast results (with the same models that we use in the sample period) we can expect to observe larger econometric errors. One reason for this is that the statistical expectation of error is always smaller in the period of fit than in any other period. A less obvious but probably more important reason is that an econometrician cannot respecify an equation that shows "structural shift" in the forecast period as he can in the sample period. Thus, evaluating a model with data that were not available when the model was specified and estimated is a much more rigorous performance test than evaluating it on the basis of data available but left out of the sample period when the model was estimated. We subject the Wharton and OBE models to this more difficult test.

It is disappointing that in the forecast period the "same change" (*Naive 2*) projections for one quarter ahead are superior to the unadjusted model projections for *GNP*, *GNP58*, and the unemployment rate for both models. The performance of the econometric models relative to naive projections improves with the length of the forecasting span, but even for one year ahead, their predictions are superior for only one (*GNP58*) out of the three variables. While adjustments to the constant terms in the equations generally improve results more in the forecast than in the sample period, they don't make enough difference to alter any of the above results.

In the forecast period the unequal central tendency (*UM*) component of mean-squared error becomes very important (especially for Wharton) relative to either of the other sources of error or to its importance in the sample period. This result appears to be caused by "structural shifts" in some of the equations after the sample period. The constant adjustments have the general effect of reducing the *UM* component for those equations that would otherwise have persisting structural equation residuals (*SER*s) of the same size and sign. However, the adjustments increase the unequal covariation (*UC*) component in some cases. In most instances the adjustments improve performance because they reduce the *UM* component more than they increase the *UC* component. The improvement from adjustment in the first quarter is most striking in the Wharton forecast, where the *UM* component in the unadjusted forecast is extremely large. The increase in error due to aggregation over GNP's components and the slight error offset in aggregation over time periods that we observe in the sample period carry over into the forecast period.

In the last part of Chapter 3 we direct our attention to the effect of model specification and estimation on error aggregation over GNP components, since this seems to be a major cause of macroeconometric forecast error. By theoretical analysis and experimentation with Wharton models, we show that error aggregation over the component variables of GNP comes from interdependencies among the structural equation residuals (*SER*s) and from error propagation through the estimated structure of the model. The *SER*s in the GNP components of the standard version of the Wharton-EFU model tend to offset each other somewhat, whether they are adjusted by the average residual (*AR*) adjustment or left unadjusted. The major explanation for aggregation error over variables in the Wharton *AR* forecasts is the structural interdependence within the system.

Estimated structural interdependence is reduced if we add anticipations variables to some of the equation specifications in the Wharton model (anticipations version). It is also reduced if we obtain our coefficient estimates by finding the least squares coefficients over the Wharton sample period, using the first-quarter *AR* simulated values from the standard two-stage least squares (*TSLS*) Wharton model for the contemporaneous explanatory variables that appear in the equations for aggregate demand. In both cases the first quarter mean-squared error is reduced for *GNP58*. Almost the entire explanation for the improved performance comes from reduced structural interdependency. Thus, even though the adjusted *SER*s are indistinguishable as to size or interdependency among the three methods, improved first-quarter forecast performance is achieved with the latter methods because there is less interdependence in their estimated structures than in the standard version *TSLS* structure. While this suggests a possible strategy for model builders in the future, the improved Wharton forecasts for all three variables in the first quarter of forecast are still inferior to the *Naive 2* projections.

In Part II we turn our attention from testing specific models used without the benefit of subjective judgment to the use of models in an actual forecasting situation. This latter procedure requires the retrieval of the specific model used for each forecast as well as the values of all the adjustments and exogenous variables that were used. Even though the models we employ for some quarters differ from those of Chapter 3, and we deal with only a subset of the forecast period of Chapter 3, the econometric model performance—when the models are used with the correct values of the exogenous variables and with no equation adjustments—is as poor as it was in that chapter. However, this is not conclusive evidence on the value of econometric models. Even if econometric models per se cannot explain adequately short-term movements away from trend, they can serve as useful forecasting aids if they provide a system into which the additional information necessary for accurate predictions can be introduced. Our major interest in Part II is to see how well econometric models perform when they are used in this way. We begin by looking at individual forecasts in detail, and then scrutinize the summary results.

In Chapter 4 we develop an error decomposition procedure that enables us to trace forecast error back to its sources. If we take the estimated slope coefficients of the system as given, our decomposition procedure allows us to see how a particular adjustment influences the forecast, what effect errors in exogenous variables have on the forecast, where structural equation residuals (*SER*s) are large, how error reverberates through the system, and what effect errors in lagged variables have in a multiperiod forecast.

Chapter 5 traces the error in individual Wharton forecasts back to its sources. This allows us to draw the appropriate lesson from errors in past forecasts beyond the simple observation that the model forecast for a particular variable is either too high or too low. For example, a forecast of the consumption of nondurables and services (*CNS*) may be very accurate because an error in the model forecast of disposable income (one of the explanatory variables in the *CNS* equation) offsets a structural equation residual (*SER*) in the *CNS* equation. In this case, it would certainly be inappropriate to interpret the accurate *CNS* forecast as evidence of the ability of the *CNS* equation to explain the consumption of nondurables and services. In Chapter 6 the same procedure is applied to the OBE forecasts. These decompositions provide useful information for model builders and users as well as for economic historians. For example, we found that Wharton adjustments to particular equations in the consumption sector tend to improve the equation being adjusted. However, the improvement in the entire consumption sector is smaller than the individual equation's improvement would lead us to expect because the adjustments that were made systematically reduce offsetting *SER* error in the consumption sector.

Chapter 6 concludes with an examination of the summary results for the OBE and Wharton forecasts decomposed in Chapters 5 and 6. We find that these results are consistent with the hypothesis that there was forecast improvement from interaction between the model forecast and the forecaster through both his selection of exogenous values and equation adjustments. Since conditional forecasts are meant to be conditional on the values shown for the exogenous variables, this evidence is not encouraging for those interested in accurate conditional forecasts. However, our evidence is also consistent with the hypothesis that there is forecast improvement from equation adjustments based on the introduction of information about both (a) past equation residuals and (b) events that are external to the model but not included in the exogenous variable set or in the equations specifications. Such evidence would indicate that conditional econometric forecasts can indeed be more accurate in practice than one would expect from the forecasting model's ability to trace movements in the economy without the aid of adjustments by an econometrician. This finding means that conditional macroeconometric forecasts cannot be rejected as unreliable solely on the basis of mechanistic tests of the reliability of the econometric model used for making the forecasts.

When we compare the ex post econometric record using the original (*OR*) adjustments with other forecasting approaches, we see econometric performance by a model being used to its best advantage. In this case, the correct values are used for the exogenous variables, and the adjustments bring in the best additional information available at the time of the forecast. From this evidence we find it difficult to recommend that policy makers rely on conditional forecasts with econometric models at this juncture. The St. Louis equation, in particular, has produced better forecasts for GNP than the structural econometric models in our forecast period. This suggests that structural specification beyond the structural tax functions that are used to construct a high-employment budget variable for the St. Louis projection may hurt conditional predictions. Evidence of this sort is consistent with the arguments of T. C. Liu, who points out the possible deleterious effects of structural restrictions for forecasting in a world where the true model may be underidentified.[21] The ex post econometric forecasts for one year ahead with the original adjustments (*OR*) are slightly better on the average than the "same change" (*Naive 2*) projections for *GNP* and *GNP58*, but they are inferior for the unemployment rate. The first-quarter results show *Naive 2* superiority to both OBE and Wharton for all three variables, with the single exception of the OBE forecasts of *GNP58*. We might find that future models will yield ex post *OR* forecasting records that have consistently smaller forecast error than all other prediction techniques. Until that time it would probably be wise for econometricians not to oversell the reliability of forecasts made with structural quarterly macroeconometric models in preference to predictions resulting from other forecasting techniques.

[21] Ta-Chung Liu, "Underidentification, Structural Estimation, and Forecasting," *Econometrica*, Vol. 28, No. 4, October 1960, pp. 855–65.