The Uses of Complex Models

CHAPTER 4:

from Ascher. W & Overholt WH (1983)
T Planning & Forecasting: Political Risk & Econ Opportunity
(N. York: John Wiley & Sons)

Judging by the volume of research devoted to the development of complex models and the money spent on subscriptions to forecast modeling services, these models have become enormously attractive to many forecast users. This attractiveness, which warrants a separate chapter on complex models, is understandable despite the fact that complex models used in forecasting are generally expensive, cumbersome, time-consuming, and often barely manageable.

TRENDS IN METHODS

In choosing methods, analysts and policy makers alike are naturally swayed by what seem to be the directions of scientific developments in general. Thus, the principal trend in forecasting for at least the past two decades has been the development of "sophisticated" methods, reflecting the rise in the allied sciences of econometrics, sociometrics, cliometrics, politometrics, and so on.

The pursuit of methodological "sophistication" in forecasting has had five related aspects.

Quantification

Numerical and mathematical approaches have been supplanting qualitative and judgmental approaches. It is significant, however, that quantification is not necessary for explicitness or rigor. It became fashionable during the recent era in the social and behavioral sciences to presume an equivalence between quantification and rigorous science. In fact, rigorous explorations (for example, through explicit scenarios) can be undertaken without recourse to quantification.

Explicitness

Explicit statistical techniques have been overlaid on numerous forecasting methods. Whereas extrapolation used to be done with a ruler, it is now

done with a regression formula. Correlating energy growth with projections of GNP used to be done implicitly; now it is more often done with the formal apparatus of correlation and regression.

Disaggregation

Trends summarizing several somewhat different kinds of activity or different types of units can be broken down into separate trends for each, with no theoretical limitation on how fine the disaggregation can be. For example, while population growth was once projected as a unitary phenomenon, now it is projected through fertility and mortality rates applied to different age cohorts for different racial components of the population. Energy demand projections for a particular fuel, like petroleum, can be made for each different end use and the results then reaggregated to determine total consumption. The rationale of disaggregation is that each separate, more homogeneous trend will behave in a simpler and therefore more predictable way than the overall trend treated as a single entity. (Sometimes, however, this rationale is false, as indicated by the analogies with quantum mechanics in Chapter 1.) When such disaggregation is employed, the elaborateness of method comes not only in having to project a larger number of more specific trends, but also in summing them up again. Elaborate models for tracking and combining these components often appear to be, and are sold as, the most impressive methodological devices of forecasting. In fact, they are little more than highly detailed but not theoretically complex accounting mechanisms.

Complex Models

All of these developments have set the stage for the shift to explicit behavioral models that go beyond these accounting mechanisms by embodying actual theoretical propositions and assumptions. A complex model has been defined (see Chapter 3) as a set of two or more explicit propositions connected by their sharing of at least one factor or variable. The resulting interconnectedness of propositions can, in theory, capture the complexity of real situations, represent intricate relationships such as mutual causation and feedback, and yield surprise implications that may not have been apparent to the analyst examining each relationship in isolation from the others. Yet complex models can also be intimidating to both their creators and their users, and they are no more valid than the (often unexplored) assumptions that their formidable equations actually express.

Bigger Models

Within this modeling movement there has been a dramatic trend toward bigger and bigger models. While this reflects disaggregation to some extent, it also reflects the tendency to include more complex relationships in these models. With the exception of a few small monetarist models designed more for theoretical explorations than for practical forecasting, econometric models have been growing from the 10-to-30 structural equation average around 1960 to the more-than-100 equation size of today.

THE PERFORMANCE AND POTENTIAL OF COMPLEX MODELING

There is no question that complex models do have unique properties as forecasting tools, just as they do for analysis in general. Multiple, interconnected propositions (outputs of one become inputs of another, or solutions must satisfy several propositions or equations simultaneously) can explicitly and systematically express mutual causation, feedback phenomena, and other intricate relationships. The question is whether these properties can be harnessed to improve prediction and explanation.

To some extent, this potential can be revealed by the existing performance record of forecasting models. Yet it may be argued that complex modeling is such a recent development in practical forecasting that its track record is not an adequate basis of evaluation, because the accuracy of many recent forecasts produced by complex models cannot yet be determined, and current models will presumably be further refined and elaborated through efforts to improve them. But even if the current record of complex models is an insufficient basis of appraisal, the record of forecasting in general reveals much about the nature of forecasting and the nature of the world that the forecasts address. In addition, we may also consider the logical and structural properties of complex models and evaluate those properties with respect to the task of forecasting. Thus, if indeed it is premature to judge the forecasting potential of modeling solely on the basis of the record to date, other bases are available. They are: the nature of complex modeling, the nature of forecasting, and the nature of the practical application of modeling to forecasting. After reviewing the performance record of complex models in forecasting, each of these topics will be considered in turn. Finally, some conclusions regarding the optimal use of complex models in forecasting will be drawn from these considerations.

Judging the Forecasting Performance of Complex Models

Complex models have been prominent in projecting only a few policy-relevant trends. For many years now, econometric models have been used in short-term economic forecasting, and more recently complex models have been employed for energy demand projections. In population forecasting, however, the impressive-seeming models utilized by the Census Bureau are not complex models at all according to our definition, since they lack connected propositions.[1] They are, in reality, elaborate accounting devices, tracking birth and death rates implied by component and cohort-specific fertility and mortality rates, but without any of the interactive effects required by our definition of the complex model. Similarly, travel demand forecasting models, also often formidable in appearance, are nonetheless usually single-equation regression models treating the demand for a particular travel mode strictly as a dependent variable.[2] No matter how many independent variables are included in such an equation, they are not explained or affected by the single dependent variable, but rather projected independently of the models (that is, taken exogenously from extrapolations of other models). Consequently, the performance record of complex modeling rests largely on the economic and energy forecasts, although complex modeling *could* be applied to forecasting in any of the other areas.

In appraising the complex models used in forecasting, there are two questions we may ask of the performance record. The first is how well complex models forecast, either in absolute terms or in comparison with other forecasting methods. The second, equally important question is whether complex forecasting models, subject to considerable refinement and elaboration over the past two decades, have been improving in terms of their accuracy.

To answer these questions, it is useful to take advantage of the fact that complex models do compete with numerous other techniques as forecasting tools. The forecasting performance of other methods provides not only a comparative basis for evaluating the record of complex modeling, but also a benchmark for establishing the intrinsic difficulty of predicting trends of a given period. For example, projecting energy demand trends after 1973 (that is, predicting the demand in particular target years beyond 1973) has been a more difficult task, both for models and for other forecasting techniques such as extrapolation or judgment, than projecting these trends for target years before 1973.[3] Thus, it is very useful to regard the average error of judgmental forecasts as an indicator of the inherent (rather than method-specific) unpredictability or uncertainty of a specific era. By tracking the level of judgmental error over time (that is, of forecasts made in years t_1, t_2, t_3, etc., for target years $t_1 + {}_N$, $t_2 + {}_N$, $t_3 + {}_N$, etc., where N is a fixed length of time, say five or ten years), we can determine whether changes in the accuracy levels of more explicit methods such as modeling can be attributed to changes in predictability. Thus, for example, if the judgmentally based five-year petroleum demand forecasts made during the period 1969–73 (for the target years 1974–78) have a higher average error than the judgmental five-year projections made in, say, 1960–67 (for the target years 1965–72), then our assessment of less accurate complex-model projections made in 1969–73 for 1974–78 ought to take into account the greater uncertainty of the post-1973 period as indicated by the decline in judgmental accuracy.

The Record of Economic Forecasting Models

Econometric models for short-term economic forecasting have been evaluated periodically by the model operators themselves (who generally developed econometric forecasting models initially as theoretical experiments, but increasingly as practical commercial enterprises) and more systematically by an ongoing survey sponsored jointly by the American Statistical Association and the National Bureau of Economic Research.[4] This survey provides the most telling appraisal of how well econometric models forecast.

The ASA-NBER survey indicates that econometric models have not—and still do not—forecast quite as well as the judgmental approach relying on no explicit routines. The most comprehensive ASA-NBER comparison of a large number of economic forecasting sources covered forecasts produced from 1968 through 1973, and found that judgmental forecasting efforts slightly outperformed econometric approaches for quarterly GNP (both nominal and real) and were roughly even in forecasting annual GNP changes.[5]

Although there has been no attempt to repeat so comprehensive a comparison for more recent forecasts, there are some up-to-date comparisons of selected models and noneconometric efforts. The median forecast of the ASA-NBER survey has been used as one benchmark for evaluating the accuracy of prominent modeling operations; the accuracy of forecasts by the Council of Economic Advisors is another noneconometric benchmark. For very short-term forecasts (one or two quarters), Vincent Su found that for the period 1968–77 the ASA-NBER median forecast still outperforms the Wharton model for most variables, but there is little difference for longer horizons.[6] Victor Zarnowitz found that for 1969–76

the annual forecasts of real GNP and the inflation rate by two modeling operations (Wharton and the University of Michigan model) and the two noneconometric sources do not establish the superiority of one method over another. Zarnowitz concludes: "This is in agreement with earlier findings, which strongly suggests that the search for a consistently superior forecaster is about as promising as the search for the philosophers' stone."[7] Finally, Stephen McNees found the same two noneconometric sources to be roughly equal in accuracy to several econometric operations (Chase Econometrics and DRI in addition to Wharton and Michigan) in annual forecasting for the 1961–76 period.[8] Thus, there seems to be no dissent from the conclusion that even the most recent econometric modeling has not offered the advantage of greater accuracy over judgment in short-term economic forecasting, despite the refinement and elaboration of the models.

Does this mean, though, that the models per se forecast with the same general level of accuracy as expert opinion? If so, this would be no mean accomplishment. Yet it is untrue. The prominent econometric modeling services all operate with a considerable amount of judgmental input, introduced either by choosing exogenous variables so that the model's predictions seem plausible to the model operator, or by tinkering with the model specifications or parameters to achieve the same result.[9] Zarnowitz points out that "the genuine ex ante forecasts here considered are all to a large extent 'judgmental.' "[10]

How do we know that pure models (the mechanical operation of the models, insulated from judgment) would not do as well or better than the models as they are run today? There are three pieces of evidence. First, the fact that the modelers leave themselves this leeway indicates their appreciation for adding the polish of judgment. Second, a valiant attempt by economist Ray Fair to operate an econometric forecasting model without judgmental adjustments led to a very poor forecasting record; parallel attempts to operate the Wharton model strictly according to its published specifications led to inferior performance compared with the judgmentally modified forecasts actually released by Wharton.[11] Third, the actual forecasting errors of the models based on runs with parameter values that had to be guessed are lower than the errors produced by operating them with the *actual* values of exogenous variables once these values are known.[12] This superiority of so-called ex ante predictions over ex post predictions is quite startling; knowing the true values of the parameters actually *reduces* the accuracy of model forecasts. This demonstrates that the convergence of plausible results provided by the forecasters' selection of exogenous values (even if they are, in retrospect, incorrect) is needed to offset errors attributable directly to the models' specifications.

The Record of Energy Forecasting Models

Appraising the performance record of energy models is complicated by the widespread use of conditional forecasts in this area, as well as by the fact that most energy forecasting models are designed to project long-range trends, so that their accuracy cannot yet be evaluated retrospectively. A brief methodological digression may clarify the appraisal problems.

When models (such as those used for short-term economic forecasting) produce unconditional forecasts, their accuracy can be evaluated straightforwardly by measuring the difference between the forecasted and the actual results. However, when models are set up to generate conditional forecasts, their performance can be evaluated in two quite different ways. First, the projection indicated as most likely may be considered as if it were the unconditional forecast nestled among other conceivable but less likely possibilities. This most likely projection may then be evaluated like a standard unconditional forecast: retrospectively, by measuring the discrepancy of predicted trend from actual trend; for current forecasting efforts, by examining the spread of most likely projections from several sources.

However, in recent years, particularly since the 1973–74 jump in oil prices, energy forecasters have been far less willing to presume that a particular energy-supply condition, or policy-choice scenario, is more likely than the others. The set of conditional forecasts (or "policy simulations") is offered as an aid to policy makers rather than as a prediction of what they will do.[13] No single projection can be meaningfully regarded as the forecasters' prediction.

When no scenario is designated as most likely, the scenarios must be regarded as exogenous factors, whose likelihoods are not at issue in the modeling exercise. The model produces a set of projections, each posited as correct *if* the corresponding condition or scenario were to hold, but without implying that any particular one will hold or that some are more likely than others. In this case, the retrospective evaluation of forecast accuracy must proceed by first establishing which condition actually prevailed, and then measure the discrepancy between the projection tied to that condition and the actual level of the predicted trend. If it is still too early to evaluate a set of conditional forecasts retrospectively, the spread of conditional forecasts of the same trend for the same target year can be used as one indication of uncertainty or minimum error, but *only if the conditional is the same for every forecast of the set*. For example, one model's prediction of petroleum demand in a specific target year under a particular scenario can be compared only with the predictions of other models for that year and that scenario. The differences among the petroleum demand predictions for that same scenario in that year necessarily reflect error, since only one (or none) of them can be correct.

This last approach is the only feasible one for judging the performance of complex energy models. They are too recent for their accuracy to be judged retrospectively, and they project several different conditionals without designating one as most likely. Because it makes no sense to compare the projections of several models if they take different scenarios as their conditional elements, many apparent opportunities for comparison and measurement of dispersion are precluded. When modelers work independently, there is no reason to expect that any of the scenarios examined will be precisely the same for two or more of them. One cannot say, then, whether differences in projections reflect these scenario differences rather than indicating disagreement, uncertainty, or error.

Comparability *is* feasible, however, when modelers work together to examine the implications of separate models run with a common set of scenarios. The most comprehensive and careful effort in this vein was conducted by the Modeling Resource Group of the National Research Council's Committee on Nuclear and Alternative Energy Systems, with the participation of four modeling groups projecting energy consumption and prices to the year 2010 and beyond, and with limited participation of two medium-range models.[14] Each of the long-term models projected the outcomes of six different policy scenarios, using the same starting points and common assumptions regarding economic growth. Therefore, differences in outcomes from one model to another, for the same scenario, represent at least the minimum degree of uncertainty or error of the models.

There are, in fact, some serious divergences between different models' results. In forecasting energy consumption, the base case projections for the year 1990 vary up to 19% for total energy, and by more than 50% for electricity generation. Some models project five times the consumption of oil indicated by other models. Even when the two extreme projections (of the five models involved in this particular comparison) are dropped, one remaining projection of oil consumption is less than half the levels forecasted by the other two models. One model projects more than twice the consumption of nuclear energy projected by the others.[15]

Most seriously—because the rationale of such simulation exercises is to determine which policies produce optimal results—shifts from one policy scenario to another bring different changes according to different models, especially for the three models most extensively compared in the Modeling Resource Group experiment. For example, the Brookhaven DESOM model and the Stanford University ETA model foresee little impact of a moratorium on nuclear development and limits on coal and shale oil exploration (compared to the base case), while the Nordhaus model (developed by William Nordhaus of Yale University) projects a 20% decline in energy consumption and domestic production under the moratorium condition.

For energy price projections, there are also major discrepancies. While the ETA and DESOM models project only small differences among different scenarios for the 1990 price of oil, the Nordhaus model foresees a 60%–70% greater price under one scenario (limits on coal and shale oil production). Oil prices for the year 2010 forecasted by the ETA and DESOM models are consistently a third to a half greater than those projected by the Nordhaus model. Under all of the scenarios involving constraints on coal and shale oil development, the price of coal is projected to increase twofold or more between 1990 and 2010 according to the ETA and DESOM models, yet the Nordhaus model projects the price to decline by nearly half. For electricity prices, ETA projects sharp increases from 1990 to 2010 under the scenarios of nuclear moratorium and limits on coal and shale oil production; Nordhaus projects declining prices.[16]

Although there are numerous consistencies across models, there are also many inconsistencies. The unavoidable conclusion is that some of the models must be wrong in quite significant ways on policy-relevant issues. It is also discouraging that even the agreement across models need not be an indication of validity; they could all be wrong. For example, all energy models predicting the 1975 levels of U.S. electricity, petroleum, and total energy consumption projected these levels higher than they actually turned out to be.[17] This confident consensus was no guarantee that the models were correct then; any consensus among models' predictions in the future may be equally misleading.

The Nature of Complex Modeling

The key attributes of complex modeling are explicitness and complexity. While complex modeling has no monopoly on either, the combination is unique. Explicitness and complexity are the virtues of complex models—if and only if the models are valid. When models are misspecified, their explicitness forecloses the possibility of adjusting their results on the basis of judgment and plausibility; the model operator must do what the model says. Complexity becomes a burden rather than a virtue for a misspecified model because complexity makes it difficult to determine just where the errors are.

Explicitness. Complex models are formulated by specifying assumptions and hypothesized relationships as explicit, usually mathematical propositions. While this procedure is often very helpful in uncovering inconsistency and vagueness in the initial ideas or verbal formulations, it cannot establish the correctness of the model's propositions. Models express assumptions but do not validate them. If the modeler tries to ensure the validity of the model's propositions by focusing on disaggregated behavior of presumably greater regularity, the problem of reaggregating these behaviors to model overall patterns becomes another potential source of error. If the modeler only includes relationships proven by past experience, there is no guarantee that they will hold in the future. There is no procedure or format of model specification that guarantees the validity of this specification.

With explicitness and the computational capacity of computers, complex models can combine almost immediate response with the fundamental analysis that an elaborate model—if correct—provides. Other explicit mechanical modes of analysis and forecasting, ranging from trend extrapolation to multiple regression, may be just as quick and automatic but cannot make as bold a claim to represent fundamental understanding of the dynamics of the system under study.[18] If the modeler wishes to convey a detailed image of reality through the model, explicitness will require the model to be large and elaborate. Whereas the analyst relying on judgment keeps the richness of detail and nuance stored in his head, the modeler must commit it all to formal expression. Greater elaborateness has been the most striking trend in model building in the past 15 years.

Yet, as noted in Chapter 3, explicitness also exercises a subtle constraint on the model's capacity to fully represent reality. Explicitness tends to limit the amount of context encompassed by the model, beause the need to formulate explicit, consistent relationships requires the modeler to discard those factors that are of some relevance but without clearcut, consistent relationship to the outcomes. In modeling approaches there is nothing of the notion of a "directed search" to identify contextual factors that may be relevant, but only in particular cases or in idiosyncratic ways. In short, modeling, like extrapolation and other routinized, explicit procedures, tends to simplify and restrict contextual considerations.

The fully explicit model, once formulated, *can* be applied mechanically, independent of any further judgment on the part of the modeler. Although the model represents the model builder's assumptions about how the system operates, mechanical operation can thereafter insulate the results from both the modeler's myopia and his insights about the specific situations being forecast. If the model truly embodies a fundamental understanding of reality with greater sophistication than the casual or less systematically formulated opinion of the model operator, its mechanical operation is a virtue. There are numerous documented instances of analysts incorrectly disregarding the signals of their methods because of personal biases and faulty preconceptions. Yet if the model produces silly results, the lack of intervention by its operators would be a deficiency. Model operators—especially when they are also model builders—are often experts in the substantive area of the model; hence, their judgment might be an important means of screening model results. But this alternative of plausibility testing is correspondingly problematical, as we shall explore later on.

Complexity. The "complexity" of complex forecasting models lies in the existence of several (and often many) interdependent relationships. If some variables (or factors) appear in more than one equation (or proposition), the relationships are "interactive," producing overall patterns of outcomes that would not be obvious from the isolated consideration of each relationship.[19]

This sort of complexity should be distinguished from the elaborateness of the accounting devices included in many models (both complex and not) to recombine disaggregated trends or to keep track of their changes over time. Very often these devices appear to be the most impressive methodological parts of a model; in fact, they represent neither the theoretical complexity of the model nor the behavioral complexity of the phenomena being modeled. For example, an energy model with 50 different equations for the consumption levels of 50 different fuels, and a few more equations to summarize consumption for broader fuel categories, may nonetheless be limited to quite simplistic propositions for each specific trend—for example, that the consumption of a given fuel increases by a fixed proportion each year. In contrast, a model linking the consumption level of one broadly defined fuel category to the levels of a few other broad fuel types (for instance, petroleum consumption limited by a large increase in coal consumption, which in turn is stimulated by lagging nuclear energy generation) may be considerably more complex.

A set of interactive relationships has *emergent properties,* often not apparent to the analyst who examines these relationships one at a time without an explicit model to track the interactions. Though the emergent patterns are indeed nothing more than implications of the basic relationships and the explicitly expressed connections among them, these patterns nonetheless can be surprising, counterintuitive, and inconsistent with expectations.[20]

The Nature of Forecasting

When complex models are applied to the task of forecasting, they enter a field with a track record that antedates complex modeling. Thus, although

modeling itself is a recent addition to the tools of forecasting, some insights can be drawn both from a theoretical examination of what forecasting entails, and from the practical benefits and caveats of different forecasting approaches.

A retrospective appraisal of the forecasting record provides the first major source of insight into whether the trends underlying the emergence of modeling as a favored forecasting method are likely to produce better analysis. The record we compiled to evaluate the accuracy and biases of various forecasting methods and sources consists of all available forecasts, up to 1975, of specific U.S. trends in population, energy demand (consumption of petroleum, electricity, and total energy), transportation (airline passenger volume, general aviation fleet size, motor vehicle registrations), economic growth (real and nominal GNP CHANGE), and technology (computer capability and nuclear energy capacity).[21] This record revealed three general findings.

First, *methodological sophistication contributes very little to the accuracy of forecasts,* as evidenced by the following:

1. The introduction of more sophisticated methods in population forecasting, with the elaborate accounting divisions of components and cohorts, has not resulted in more accurate demographic forecasts.
2. Economic modeling and the explicitly quantitative approach of leading indicators have not improved the accuracy of short-term economic forecasting.
3. For energy demand forecasts, correlation and simple trend extrapolation have had the same general levels of accuracy, though the former is a much more elaborate procedure.
4. Separating electricity or petroleum demand into various end-use components does not add any appreciable accuracy over projecting the demand for each as a unitary trend.
5. In transportation forecasting, the more elaborate regression models recently adopted by the FAA to predict both commercial air traffic and the general aviation fleet size have not improved the accuracy record.

Second, *a forecast's time horizon is the strongest and most consistent correlate of its accuracy.* Though there are some exceptions, the general rule is that shorter forecasts are more accurate, often in a nearly linear relationship:

1. Five-year petroleum consumption forecasts have had a median error of about 6%, and 10-year forecasts have had an error of about 13%.

2. Motor vehicle registration forecasts have median errors in percentages roughly equal to the forecast lengths in years.
3. Technological breakthrough predictions by expert panels show greater dispersion (indicating uncertainty and greater minimum error levels) for more remote breakthroughs than for breakthroughs perceived to be more imminent.

Third, *all evidence points to the essential importance of the validity of core assumptions antecedent to the choice and application of methods.* Behind any forecast, regardless of the sophistication of methods, are irreducible assumptions representing the forecaster's basic outlook on the context within which the specific trend develops. These core assumptions are not derivable from methods; on the contrary, the method is a vehicle for tracing through the consequences or implications of core assumptions originally chosen independently of (and intellectually prior to) the method. The method selected usually reflects and signifies this preconception of the pattern of future change. For example, envelope curves are chosen as a forecasting method only if the forecaster has a preconception that cumulative scientific breakthroughs will cause the technology improvement rate to take the envelope-curve form. Forecasting by analogy is utilized only when the forecaster believes a particular historical pattern to be analogous. Formal models, too, ought to be viewed as the explicit expression of a set of assumptions, and the operation of the model is the tracking of interactions arising from these presumed relationships.

These core assumptions are the principal determinants of forecast accuracy. When the core assumptions are valid, the choice of method is either secondary or obvious. When these assumptions fail to capture the reality of the context, other factors such as method generally make little difference. This helps to explain why correlational forecasts of electricity and petroleum demand do no better than extrapolation: when one trend is projected by correlating it with a presumably more fundamental, contextual trend (say, industrial production or GNP), its accuracy depends not primarily on whether this correlation is appropriate, but rather on how well the more fundamental trend is forecast. The primacy of contextual assumptions is also consistent with the great dispersion of projections of air transportation forecasting models based on regressions on economic growth: the greatest uncertainty is in predicting the economic growth context.[22]

The importance of both recency and core assumptions makes the problem of relying on antiquated core assumptions particularly serious. This "assumption drag" has been the source of the most drastic errors in forecasting. The worst population forecasts prepared in the 1930s and early 1940s were based on a no longer valid assumption of declining birth rates. Similarly, the electricity demand forecasts of the 1960s continued to project fairly low electricity demand growth, even when the actual growth rates were contradicting this assumption.

These core assumptions are often political. Even for technological forecasting, seemingly far removed from political developments, a major source of uncertainty stems from doubt about the political context. If the variation among predictions of a given technological breakthrough is taken as a measure of uncertainty, it can be demonstrated that developments in technological areas for which advances require large-scale official programs (such as health-care systems, medical education, and space exploration) are more uncertain than developments in technologies depending on engineering refinements and the disaggregated market diffusion of such innovations (such as communications, educational technology, and automation).[23] The difficulty in predicting innovations requiring discrete, high-level, "official" (though not necessarily governmental) policy decisions indicates the pivotal role of political assumptions—both because the political context is important to the development of these technologies and because it is difficult to predict. Political forecasting, perhaps because of the discrete and discretionary nature of political decisions, is the weakest link in the whole range of forecasting trends even remotely affected by policy.

The political context is often ignored (on the dubious grounds that, in the long run, relatively short-lived political conditions or policy decisions will have no net effect), or it is considered in a rudimentary fashion. For example, almost all energy projections made before 1973 totally ignored government responses to energy problems, including the possible responses of petroleum-producing nations to the dwindling real earnings of their oil exports. Even now, long-term energy, economic, and transportation forecasts, if they are based on any explicit assumptions about government policy at all, generally pose a single and stable governmental policy choice, exogenous to the forecasting procedure, as a fixed condition.[24] To ignore the possibility of governmental problem-solving decisions is to limit seriously the validity of any forecast.

The parameters found to fit in the past will hold in the future only if the relationships embodied by the model are invariant. Simply finding a set of relationships and parameters that fit past data in no way guarantees future fit. Some theorists argue persuasively that if such relationships describe aggregate behavior (for example, macroeconomic behavior in the case of economic models), invariance is unlikely even in principle. Aggregate specifications are subject to "parameter drift," if only because individuals' decision rules change from one situation to another, from one set of expectations to another.[25] Robert Lucas points out that even the parameters utilized in the large-scale econometric models as if they had been constant have in fact been drifting, and that this is actually acknowledged by the forecasters through their adjustments of model forecasts on the basis of prior model errors.[26]

We come back to the puzzle of why the large-scale econometric models for short-term economic forecasting are under continual development and yet do not improve in accuracy. This line of reasoning suggests an explanation: the "improvement" of substituting new specifications for old ones merely replaces the representation of the previous context with a representation of the newer context, but without coming closer to a more generally valid representation. There is no a priori reason why there should be a set of aggregate-level propositions that remains generally valid over time.

Implications for Modeling

If, indeed, valid up-to-date contextual assumptions, rather than methodological sophistication, have been the primary determinants of forecast accuracy, there is little basis for confidence that the unique capacity of models to explicitly track complex interactions will materialize in consistently greater accuracy.[27] The advantages of complexity and explicitness may well be offset by the built-in assumption drag of elaborate models, as well as by their tendency to restrict contextual considerations to regular, law-like relationships.

Just as importantly, complex modeling, though it certainly does not preclude the incorporation of policy interactions, has failed to do so. The current generation of complex models developed for economic and energy modeling does incorporate sociopolitical factors, including policy choices, but only as givens, usually expressed through conditional scenarios or parameters. Complex modeling has been particularly slow to incorporate policy response. This omission is illustrated by the approach of William Hogan (organizer of the Energy Modeling Forum at the Stanford University Institute for Energy Studies) and Alan Manne (a principal developer of the Stanford ETA model) in an article on economy-energy interactions.[28] They argue:

> The value share of the energy sector determines the incremental effect upon the GNP. If the 4% value share remained constant, this would mean that a 10% reduction in energy inputs would produce only a 0.4% drop in total output. Thus, for small changes in energy availability, there need not be a proportional impact upon the economy as a whole.

They then go on to say that for large changes in energy availability the critical factor in determining economic growth rates becomes the substitutability of energy inputs by nonenergy inputs (such as insulation or energy-efficient machinery) to compensate for energy supply shortfalls.[29] This formulation may be reasonable if the government is passive or acts only to facilitate the substitution of inputs. But it does not encompass the possibility that the government may react to inflation (triggered by higher energy prices) by enacting recessionary economic policies with much greater impact on economic growth. The commonality of recessionary policies adopted by OECD countries after the oil price increase of 1973–74 would indicate that this sort of reaction is regular enough to be modeled with some degree of confidence.

The failure to model policy response is, of course, not a mere oversight. It reflects the almost universal commitment of practical modelers to represent aggregate dynamics with aggregate, macro models in which discrete decisions (if not provided as givens) are assumed to be subsumable by general mechanisms of equilibration or optimization.

In principle, disaggregated models based on decision-making agents as the units of analysis (for instance, based on the perhaps firmer principles of microeconomics) can model both discrete responses and the aggregate patterns. If indeed the aggregate behavior of the system follows a pattern of equilibration or optimization as a result of myriad individual decisions—a point not challenged by macroeconomics—a microlevel model can represent such patterns.

THE PRACTICAL APPLICATION OF MODELING TO FORECASTING

The formal properties of forecasting and of complex modeling do not completely determine how complex models are applied in practice to forecasting tasks. There is a range of discretion in the practice of forecast modeling. Choices within this range can make a huge difference in the utlity of complex models.

First, despite the large number of possible model forms—varying in size, format, level of aggregation, propositional form, and so on—in practice complex forecasting models have become highly uniform. Once practical forecasting in a particular area becomes established, there is a strong tendency for the models to converge. Those showing the earliest success drive out the less successful. From the interesting variety of econometric models in the late 1960s, the surviving practical short-term economic forecasting models are remarkably similar: all large, macroeconomic models in the Klein-Goldberger tradition.[30] This could be justified as "survival of the fittest," except that complex modeling as an intellectual enterprise is not advanced enough for all other lines of development to be rejected as inferior simply on the basis of early falterings. In particular, microlevel models have received surprisingly little support and development, considering their initial promise.[31]

Second, thus far, the application of complex modeling to forecasting has also been marked by the related practices of commitment to the same basic structure for long periods of time, continual superficial revision of such models, and plausibility testing of their outputs. Most methodologically sophisticated forecasting efforts are big, expensive, one-shot affairs. Since rigorous, elaborate analysis is time-consuming and expensive, there has been a natural tendency for forecasters to pour their efforts into grand, once-and-for-all projects, carried out only infrequently and yet used long after they are produced because the immense effort makes them seem definitive. This, of course, poses a very serious problem of obsolescence.

In one important way, it would seem that modeling approaches depart from this pattern of one-shot projects because they are automated, ongoing mechanisms that can be updated and can spew out new forecasts at any time. Yet the sunk costs in the development of an elaborate model force it to be a one-shot effort in a more basic sense. Though forecasters may tinker with their model and reestimate its parameters, the model's basic structure remains intact for long periods of time. It is much easier for the modeler to tinker than to scrap the entire model, even though the core assumptions may reside in the model's basic structure. For example, the central mechanism of some energy models is an optimization routine that establishes supplies, demands, and prices so as to maximize benefits under the presumption of competitive markets.[32] Charles Hitch points out that such models (among others) "assume that the economy and its markets will be permitted to function, that prices are not controlled, and that producers and consumers will not be constrained by regulation from responding to appropriate market signals."[33] Yet, after the modeler has spent years developing optimization routines, apparent violations of such assumptions are more likely to be accommodated by patchwork modifications, or disregarded altogether as short-term aberrations, than they are to trigger the abandonment of the model.

Even though these basic structures endure, complex models in practice undergo continual revision. Although the primary emphasis of forecast modeling is usually on the quality of the forecasts rather than on the development of the models, forecasting models (like complex models in general) still represent efforts at theory building—and these efforts are never finished. Therefore, model revision, which seems to the cynic to be an ad hoc effort to keep a fundamentally misspecified model more or less in line with reality, is generally regarded by the model builder as the normal routine of science. In this view, continual revision complements rather than contradicts the durability of basic model structure, because

altering the same basic model to keep it consistent with the most recent changes in the phenomena it models saves the modelers from scrapping the basic model itself. There is a fundamental tension, ethical as well as practical, in the simultaneous claims of theory building and practical application. Such tension was particularly dramatic in the case of the Club of Rome members, who held press conferences and mailed copies of *The Limits to Growth* to policy makers, demanding urgent policy response to their model's conclusions of imminent doom, but then responded to criticism of their findings by saying that the model was merely a preliminary effort at theory construction.

Third, plausibility testing is an almost universal procedure in the practical application of forecasting models. This results in part from the commitment to continual model development and in part from the modelers' modesty as to the accuracy of their models. Usually, at least at this stage in the development of modeling, when a model generates implausible outcomes the onus is laid on the model rather than on the modeler's perception of what is plausible. As mentioned in Chapter 3, this plausibility testing proceeds not just by directly rejecting implausible model results, but also through the subtler means of providing exogenous information that makes the model's results plausible.

In energy modeling, a good example of the interplay between exogenous value selection and the plausibility of model results is provided by the treatment of energy prices in the RFF/SEAS Modeling System:

> Since the model does not include a mechanism for determining these prices, our procedure is to develop assumptions about price changes, run the model, adjust prices, and iterate if necessary.[34]

What are the combined effects of these practices? The complex model holds an ambiguous status as both practical device and experiment. When model builders come up with new formulations, should forecast users regard such models as the best devices the modelers can offer for forecasting, or as the latest turn—perhaps a dead end—in the theoretical and methodological search? Most importantly, there is an obvious and direct tradeoff between plausibility testing and the capacity of the model to express the counterintuitive implications of its assumptions. These practices drastically weaken the forecast-modeling enterprise's sensitivity to surprising future possibilities. The homogeneity of models simply reinforces this weakness, as similar models undergoing similar judgmental censorship by modelers holding similar outlooks on the future can so easily reassure all parties that the future is seen with certainty.

It is not clear, however, how long these practices will last. Certain models have become institutionalized, with a life and a durability beyond the active role of their originators. Related to this process of institutionalization is a growing distinction between model builders and model operators. Innovators like Lawrence Klein and Otto Eckstein are not preoccupied with the routine operation of the Wharton and DRI econometric models; they have other, more ambitious things to do.

This development presents both drawbacks and advantages. When methods themselves acquire charisma, their operation can take the place of careful reasoning. Model operators who regard themselves as technicians are even less likely than model builders to grapple with the basic structure of an already established model. Hence, the obsolescence of core assumptions embodied in the model's structure will become an even greater risk. If any judgmental input remains in the operation of the model, it will not be from leading economists privy to the discussions at the highest levels of government and industry. Thus, even if the model operators are no less competent as practical economists than prominent model builders, the caliber of inside information used in the remaining practice of plausibility testing is likely to deteriorate.

CONCLUSIONS

The practical issue boils down to this: if complex models as normally operated converge with judgment or, if operated more mechanically, perform no better than judgment, what is the incentive to employ an expensive, formidable model? The current practice of model operation, which is virtually "judgment guided by modeling," is saddled with the predictive weakness of aggregate macrolevel models, without the capacity to reveal surprising outcomes implicit in the component propositions. The possibly emerging practice of mechanically run aggregate models could restore this sensitivity to surprise, but falls prey to the most severe problems of assumption drag, inadequate representation of policy response, and parameter drift.

This dilemma is unresolvable as long as all complex forecasting models are basically alike and are operated alike to pursue the same function. More pluralism in practical forecast modeling is urgently needed. First, the two functions of forecast modeling—to produce consistently reasonable forecasts and to anticipate counterintuitive outcomes—can rarely be pursued by the same model at the same time, but can be as separate enterprises. The former provides credibility for the latter, while surprise

sensitivity, as the unique contribution of modeling, enhances the attractiveness of the entire modeling undertaking. Therefore, the operation of some models in the current mode of plausibility testing, and other models in the surprise-sensitive, unconstrained mode, can be mutually reinforcing.

Second, the homogenization of practical forecasting models should be avoided; it not only cuts off promising avenues for developing better models, but also leaves the field to the large-scale aggregate models which (for all the reasons discussed above) will most likely always require plausibility testing. It should be recognized that the really quite modest success of the macrolevel modeling operations (the models *plus* the judgmentally based interventions of the model operators *plus* the adjustments for parameter drift) in terms of performance is no basis for concluding that the models per se are optimal in either specific detail or general form.

This diversity in operating modes and model forms requires considerable sophistication on the part of the models' audiences and funders. They have to be able to recognize which models' outputs are to be regarded as most consistently and plausibly likely, and which are designed to be ultrasensitive to possible surprises even at the risk of a worse overall record of accuracy. If this is not recognized, forecast users will be disappointed in the first type for not being daring enough, and cynical about the second type for not being accurate enough. Funders must be aware of the widely differing development periods necessary to bring diferent modeling approaches to the point where they start to pay off, and must maintain a broad R&D strategy of cultivating different approaches even when only a few seem superior in the short run. Diversity and a modicum of tolerance will aid in bringing out the potential of complex modeling, without the danger of elevating one complex modeling approach as the panacea for the problems of forecasting.

Most of these comments apply to all forms of complex modeling, and most of the examples of the problems of complex modeling have been drawn from the fields to which complex modeling is most applicable, namely various kinds of economic and energy forecasting. But clearly the problems of complex models are far more severe in political forecasting, where discontinuous leadership changes can frequently change the basic structure of relationships, where changing public perceptions exacerbate the problem of parameter drift, and where the impact of unquantifiable contextual changes is undeniably great. Thus, the political forecaster attempting to use complex models must be doubly wary: he must be somewhat concerned about using complex model forecasts of the economy as inputs to the political forecasting, and he must be particularly skeptical about using complex models to forecast strictly political phenomena.

PROPER USES OF LARGE-SCALE MODELS

The burden of this analysis has been to deflate the ambitious claims frequently made for large-scale, quantitative, computerized predictive models. For most predictive purposes such models simply have not proved more accurate than other methods, and they are far more expensive than competing methods that are easier to keep current. There remain, however, uses for which these models do possess some comparative advantages.

Complex Accounting

While models have not achieved their claims of producing subtle or intuitively surprising results through the interaction of complex substantive assumptions, they do facilitate complicated calculations. That is, when used as calculators rather than as "brains," for accounting rather than analysis, computer models are frequently helpful. For instance, in projecting a balance of payments problem into the future, a model will seldom reveal surprising results, but it will greatly speed up the calculation.

Consistency Checks

Verbal analyses can sometimes conceal internal inconsistencies, particularly when they concern a complex subject or when they represent the combined view of experts with different perspectives. When such complicated analyses can be quantified, or when key parts of them can be quantified, it is frequently useful to check the internal consistency of a paper by formal modeling. For instance, one group analyzing the Chinese economic plans of the late 1970s found that its experts were producing conclusions that were based upon inconsistent premises, and were able to advance their analysis by modeling the inconsistencies.

Sensitivity Estimates

One of the particularly useful forms of accounting that a computer model can perform is the "sensitivity check." For instance, when Japan proposes tariff reductions in response to U.S. demands, a simple computer accounting mechanism can rapidly calculate the approximate effect on the U.S. balance of payments, cutting through the complexity of long lists of detailed changes in rules for large numbers of products. In at least one instance, such a model was used to show very quickly that an apparently responsive proposal was virtually inconsequential.

Policy Simulation

The unique capacity of complex models to trace out the implications of numerous interactions gives these models an advantage over judgment in exploring what would occur if an extensive package of policy changes were to be implemented. Whereas forecasting what *will* happen requires the prediction of policy response—a task that complex models have failed to address—the policy simulation takes a stipulated set of responses as a given, in order to explore its consequences. Human judgment, in contrast, excels in short-term forecasting in a context of limited exogenous change but easily bogs down in calculating the net impact of multiple changes. For example, the consequences of major departures from economic policy or a completely new package of energy regulations can be more easily tracked by a complex model.

Time Estimates

In certain economic situations, models can usefully estimate the time necessary for changes in economic policy or in the economic environment to have a given effect on the economy.

Heuristics

The intellectually most sophisticated use of complex models is the heuristic exploration of an issue. For this use, one creates a model and then manipulates it to see what differences are caused by varying the parameters, the assumptions, and the initial data. For instance, prediction of a country's balance of payments a decade hence is in the strict sense virtually impossible. But by modeling the interactions of principal components of the balance of payments, and then making different assumptions about those components, one can learn much about the dynamics of the balance and the likely limits on change. Such a heuristic use of models is technically just a particularly complicated form of sensitivity testing, but it can be quite enlightening. This use dovetails nicely with the argument of Chapter 2 that analysis-as-heuristics is frequently the optimal contribution the analyst can make to the policy maker.

Finally, it is important to note that the claims for complex modeling techniques are stated both in the present tense and in the future. While complex models have not proved particularly helpful in improving accuracy or insight, it is imaginable that they will improve in the future.

Indeed, modeling efforts that are criticized for failure to improve on other methods almost always assert that the models are in a preliminary phase of development, that one cannot expect immediate results, and that decisive improvements will occur in the future. However, as disussed above, the problems of parameter drift and of critical unquantifiable political and organizational issues impose severe inherent limits on improvements in future models. For the present, most complex, computerized models have created a burdensome expense without corresponding benefits and have detracted from the efforts most vital to successful forecasts: careful analysis of basic assumptions, detailed attention to nuances of context, and frequent updating.

NOTES

1. Except for "identities," equations that aggregate without representing behavioral properties.
2. See Richad Vitek and Nawal Taneja, *The Impact of High Inflation Rates on the Demand for Air Passenger Transportation* (Cambridge, Mass.: Massachusetts Institute of Technology Flight Transportation Laboratory, 1975); and Dale E. McDaniel, "Transportation Forecasting: A Review," *Technological Forecasting and Social Change* 3 (1972): 367–89. J. Scott Armstrong, in *Long-Range Forecasting: From Crystal Ball to Computer* (New York: Wiley, 1978), reports that after extensive searching he was able to find only three transportation models that could be termed "complex" in the sense we are employing.
3. William Ascher, *Forecasting: An Appraisal for Policy-Makers and Planners* (Baltimore: Johns Hopkins University Press, 1978), chapter 5.
4. "Short term" in economic forecasting generally means no more than four to six quarters. The survey of these short-term forecasts is published in various issues of both *The American Statistician* and *Explorations in Economic Research*.
5. Vincent Su and Josephine Su, "An Evaluation of ASA/NBER Business Outlook Survey Forecasts," *Explorations in Economic Research* 2, no. 4 (Fall 1975): 588–618, esp. 600.
6. Vincent Su, "An Error Analysis of Econometric and Noneconometric Forecasts," *Proceedings of the American Economic Association* 68, no. 2 (May 1978): 306–12. It is important to note that the error of the median forecast is not the same as the median error of the set of forecasts.

7. Victor Zarnowitz, "On the Accuracy and Properties of Recent Macroeconomic Forecasts," *Proceedings of the American Economic Association* 68, no. 2 (May 1978): 313–19.
8. Stephen McNees, "An Evaluation of Economic Forecasts: Extension and Update," *New England Economic Review* (September/October 1976): 30–44; and Stephen McNees, "An Assessment of the Council of Economic Advisers' Forecast of 1977," *New England Economic Review* (March/April 1977): 3–7.
9. See the very candid description given by Lawrence Klein in *An Essay on the Theory of Economic Prediction* (Helsinki: Jahnsson Lectures, 1968).
10. Zarnowitz, "Accuracy and Properties," p. 315.
11. See Ascher, *Forecasting* chapter 4; and Klein, *Essay*.
12. Bert G. Hickman, "Introduction and Summary," *Econometric Models of Cyclical Behavior,* Bert G. Hickman, ed. (New York: National Bureau of Economic Research, 1972), p. 17.
13. Congressional Research Service, "Energy Demand Studies—An Analysis and Comparison," *Middle- and Long-Term Energy Policies and Alternatives,* Part 7, Appendix to Hearings before the Subcommittee on Energy and Power, Committee on Interstate and Foreign Commerce, U.S. House of Representatives, March 25–26, 1976 (Washington, D.C.: U.S. Government Printing Office, 1976), p. 2.
14. National Research Council, Committee on Nuclear and Alternative Energy Systems, Synthesis Panel, *Modeling Resource Group, Energy Modeling for an Uncertain Future,* Supporting Paper 2 (Washington, D.C.: National Academy of Sciences, 1978). Two other comparative efforts—reported in Charles Hitch, ed., *Modeling Energy-Economy Interactions: Five Approaches,* Research Paper R-5 (Washington, DC.: Resources for the Future, 1977); and William Hogan, "Reporting on the Energy Modeling Forum," Working Paper (Stanford University, 1977)—did not enforce as much uniformity in scenarios, assumptions, and starting points. In any event, there is much overlap in the models examined.
15. National Research Council, *Modeling Resource Group,* p. 47.
16. *Ibid.*, p. 48.
17. Ascher, *Forecasting,* chapter 5.
18. See the delightful discussion of the fundamentalist-chartist controversy in Martin Shubik, "The Nature and Limitations of Forecasting," *Toward the Year 2000: Work in Progress, Daedalus* (Summer 1967): 945.