A Skeptic's Guide to Computer Models

by John D. Sterman

This article was written by Dr. John D. Sterman, Director of the MIT System Dynamics Group and Professor of Management Science at the Sloan School of Management, Massachusetts Institute of Technology, 50 Memorial Drive, Cambridge, MA 02139, USA; email: jsterman@mit.edu.
 This paper is reprinted from Sterman, J. D. (1991). A Skeptic's Guide to Computer Models. In Barney, G. O. et al. (eds.), *Managing a Nation: The Microcomputer Software Catalog.* Boulder, CO: Westview Press, 209-229. An earlier version of this paper also appeared in Foresight and National Decisions: The Horseman and the Bureaucrat (Grant 1988).

A Skeptic's Guide to Computer Models...

But Mousie, thou art no they lane
In proving foresight may be vain;
The best-laid schemes o' mice an' men
Gang aft a-gley,
An lea'e us nought but grief an' pain,
For promis'd joy!
Robert Burns, "To a Mouse"

The Inevitability of Using Models

Computer modeling of social and economic systems is only about three decades old. Yet in that time, computer models have been used to analyze everything from inventory management in corporations to the performance of national economies, from the optimal distribution of fire stations in New York City to the interplay of global population, resources, food, and pollution. Certain computer models, such as *The Limits to Growth* (Meadows et al. 1972), have been front page news. In the US, some have been the subject of numerous congressional hearings and have influenced the fate of legislation. Computer modeling has become an important industry, generating hundreds of millions of dollars of revenues annually.

As computers have become faster, cheaper, and more widely available, computer models have become commonplace in forecasting and public policy analysis, especially in economics, energy and resources, demographics, and other crucial areas. As computers continue to proliferate, more and more policy debates—both in government and the private sector—will involve the results of models. Though not all of us are going to be model builders, we all are becoming model consumers, regardless of whether we know it (or like it). The ability to understand and evaluate computer models is fast becoming a prerequisite for the policy maker, legislator, lobbyist, and citizen alike.

During our lives, each of us will be faced with the result of models and will have to make judgments about their relevance and validity. Most people, unfortunately, cannot make these decisions in an intelligent and informed manner, since for them computer models are *black boxes*: devices that operate in completely mysterious ways. Because computer models are so poorly understood by most people, it is easy for them to be misused, accidentally or intentionally. Thus there have been many cases in which computer models have been used to justify decisions already made and actions already taken, to provide a scapegoat when a forecast turned out wrong, or to lend specious authority to an argument.

If these misuses are to stop and if modeling is to become a rational tool of the general public, rather than remaining the special magic of a technical priesthood, a basic understanding of models must become more widespread. This paper takes a step toward this goal by offering model consumers a peek inside the black boxes. The computer models it describes are the kinds used in foresight and policy analysis (rather than physical system models such as NASA uses to test the space shuttle). The characteristics and capabilities of the models, their advantages and disadvantages, uses and misuses are all addressed. The fundamental assumptions of the major modeling techniques are discussed, as is the appropriateness of these techniques for foresight and policy analysis. Consideration is also given to the crucial questions a model user should ask when evaluating the appropriateness and validity of a model.

Mental and Computer Models

Fortunately, everyone is already familiar with models. People use models—mental models—every day. Our decisions and actions are based not on the real world, but on our mental images of that world, of the relationships among its parts, and of the influence our actions have on it.

Mental models have some powerful advantages. A mental model is flexible; it can take into account a wider range of information than just numerical data; it can be adapted to new situations and be modified as new information becomes available. Mental models are the filters through which we interpret our experiences, evaluate plans, and choose among possible courses of action. The great systems of philosophy, politics, and literature are, in a sense, mental models.

But mental models have their drawbacks also. They are not easily understood by others; interpretations of them differ. The assumptions on which they are based are usually difficult to examine, so ambiguities and contradictions within them can go undetected, unchallenged, and unresolved.

That we have trouble grasping other people's mental models may seem natural. More surprising, we are not very good at constructing and understanding our own mental models or using them for decision making. Psychologists have shown that we can take only a few factors into account in making decisions (Hogarth 1980; Kahneman, Slovic, and Tversky 1982). In other words, the mental models we use to make decisions are usually extremely simple. Often these models are also flawed, since we frequently make errors in deducing the consequences of the assumptions on which they are based.

Our failure to use rational mental models in our decision-making has been well demonstrated by research on the behavior of people in organizations (e.g., families, businesses, the government). This research shows that decisions are not made by rational consideration of objectives, options, and consequences. Instead, they often are made by rote, using standard operating procedures that evolve out of tradition and adjust only slowly to changing conditions (Simon 1947, 1979). These procedures are determined by the role of the decision makers within the organization, the amount of time they have to make decisions, and the information available to them.

But the individual perspectives of the decision-makers may be parochial, the time they have to weigh alternatives insufficient, and the information available to them dated, biased, or incomplete. Furthermore, their decisions can be strongly influenced by authority relations, organizational context, peer pressure, cultural perspective, and selfish motives. Psychologists and organizational observers have identified dozens of different biases that creep into human decision making because of cognitive limitations and organizational pressures (Hogarth 1980; Kahneman, Slovic, and Tversky 1982). As a result, many decisions turn out to be incorrect; choosing the best course of action is just too complicated and difficult a puzzle.

Hamlet exclaims (perhaps ironically) "What a piece of work is a man, how noble in reason, how infinite in faculties...!" But it seems that we, like Hamlet himself, are simply not capable of making error-free decisions that are based on rational models and are uninfluenced by societal and emotional pressures.

Enter the computer model. In theory, computer models offer improvements over mental models in several respects:

> They are explicit; their assumptions are stated in the written documentation and open to all for review.
>
> They infallibly compute the logical consequences of the modeler's assumptions.
>
> They are comprehensive and able to interrelate many factors simultaneously.

A computer model that actually has these characteristics has powerful advantages over a mental model. In practice, however, computer models are often less than ideal:

> They are so poorly documented and complex that no one can examine their assumptions. They are black boxes.
>
> They are so complicated that the user has no confidence in the consistency or correctness of the assumptions.
>
> They are unable to deal with relationships and factors that are difficult to quantify, for which numerical data do not exist, or that lie outside the expertise of the specialists who built the model.

Because of these possible flaws, computer models need to be examined carefully by potential users. But on what basis should models be judged? How does one know whether a model is well or badly designed, whether its results will be valid or not? How can a prospective user decide whether a type of modeling or a specific model is suitable for the problem at hand? How can misuses of models be recognized and prevented? There is no single comprehensive answer, but some useful guidelines are given on the following pages.

The Importance of Purpose

A model must have a clear purpose, and that purpose should be to solve a particular problem. A clear purpose is the single most important ingredient for a successful modeling study. Of course, a model with a clear purpose can still be incorrect, overly large, or difficult to understand. But a clear purpose allows model users to ask questions that reveal whether a model is useful for solving the problem under consideration.

Beware the analyst who proposes to model an entire social or economic system rather than a problem. Every model is a representation of a system—a group of functionally interrelated elements forming a complex whole. But for the model to be useful, it must address a specific problem and must simplify rather than attempting to mirror in detail an entire system.

What is the difference? A model designed to understand how the business cycle can be stabilized is a model of a problem. It deals with a part of the overall economic system. A model designed to understand how the economy can make a smooth transition from oil to alternative energy sources is also a model of a problem; it too addresses only a limited system within the larger economy. A model that claims to be a representation of the entire economy is a model of a whole system. Why does it matter? The usefulness of models lies in the fact that they simplify reality, putting it into a form that we can comprehend. But a truly comprehensive model of a complete system would be just as complex as that system and just as inscrutable. The map is not the territory—and a map as detailed as the territory would be of no use (as well as being hard to fold).

The art of model building is knowing what to cut out, and the purpose of the model acts as the logical knife. It provides the criterion about what will be cut, so that only the essential features necessary to fulfill the purpose are left. In the example above, since the purpose of the comprehensive model would be to represent the entire economic system, few factors could be excluded. In order to answer all questions about the economy, the model would have to include an immense range of long-term and short-term variables. Because of its size, its underlying assumptions would be difficult to examine. The model builders—not to mention the intended consumers—would probably not understand its behavior, and its validity would be largely a matter of faith.

A model designed to examine just the business cycle or the energy transition would be much smaller, since it would be limited to those factors believed to be relevant to the question at hand. For example, the business cycle model need not include long-term trends in population growth and resource depletion. The energy transition model could exclude short-term changes related to interest, employment, and inventories. The resulting models would be simple enough so that their assumptions could be examined. The relation of these assumptions to the most important theories regarding the business cycle and resource economics could then be assessed to determine how useful the models were for their intended purposes.

Two Kinds of Models: Optimization Versus Simulation and Econometrics

There are many types of models, and they can be classified in many ways. Models can be static or dynamic, mathematical or physical, stochastic or deterministic. One of the most useful classifications, however, divides models into those that optimize versus those that simulate. The distinction between optimization and simulation models is particularly important since these types of models are suited for fundamentally different purposes.

Optimization

The Oxford English Dictionary defines *optimize* as "to make the best of most of; to develop to the utmost." The output of an optimization model is a statement of the best way to accomplish some goal. Optimization models do not tell you what will happen in a certain situation. Instead they tell you what to do in order to make the best of the situation; they are normative or prescriptive models.

Let us take two examples. A nutritionist would like to know how to design meals that fulfill certain dietary requirements but cost as little as possible. A salesperson must visit certain cities and would like to know how to make the trip as quickly as possible, taking into account the available flights between the cities. Rather than relying on trial and error, the nutritionist and the salesperson could use optimization models to determine the best solutions to these problems.

An optimization model typically includes three parts. The *objective function* specifies the goal or objective. For the nutritionist, the objective is to minimize the cost of the meals. For the salesperson, it is to minimize the time spent on the trip. The *decision variables* are the choices to be made. In our examples, these would be the food to serve at each meal and the order in which to visit the cities. The *constraints* restrict the choices of the decision variables to those that are acceptable and possible. In the diet problem, one constraint would specify that daily consumption of each nutrient must equal or exceed the minimum requirement. Another might restrict the number of times a particular food is served during each week. The constraints in the travel problem would specify that each city must be visited at least once and would restrict the selection of routes to actually available connections.

An optimization model takes as inputs these three pieces of information—the goals to be met, the choices to be made, and the constraints to be satisfied. It yields as its output the best solution, i.e., the optimal decisions given the assumptions of the model. In the case of our examples, the models would provide the best set of menus and the most efficient itinerary.

Limitations of Optimization

Many optimization models have a variety of limitations and problems that a potential user should bear in mind. These problems are: difficulties with the specification of the objective function, unrealistic linearity, lack of feedback, and lack of dynamics.

Specification of the Objective Function: Whose Values? The first difficulty with optimization models is the problem of specifying the objective function, the goal that the model user is trying to reach. In our earlier examples, it was fairly easy to identify the objective functions of the nutritionist and the salesperson, but what would be the objective function for the mayor of New York? To provide adequate city services for minimal taxes? To encourage the arts? To improve traffic conditions? The answer depends, of course, on the perspective of the person you ask.

The objective function embodies values and preferences, but which values, whose preferences? How can intangibles be incorporated into the objective function? How can the conflicting goals of various groups be identified and balanced? These are hard questions, but they are not insurmountable. Intangibles often can be quantified, at least roughly, by breaking them into measurable components. For example, the quality of life in a city might be represented as depending on the rate of unemployment, air pollution levels, crime rate, and so forth. There are also techniques available for extracting information about preferences from interviews and other impressionistic data. Just the attempt to make values explicit is a worthwhile exercise in any study and may have enormous value for the clients of a modeling project.

It is important that potential users keep in mind the question of values when they examine optimization models. The objective function and the constraints should always be scrutinized to determine what values they embody, both explicitly and by omission. Imagine that a government employee, given responsibility for the placement of sewage treatment plants along a river, decides to use an optimization model in making the decision. The model has as its objective function the cheapest arrangement of plants; a constraint specifies that the arrangement must result in water quality standards being met. It would be important for the user to ask how the model takes into account the impacts the plants will have on fishing, recreation, wild species, and development potential in the areas where they are placed. Unless these considerations are explicitly incorporated into the model, they are implicitly held to be of no value.

Linearity. Another problem, and one that can seriously undermine the verisimilitude of optimization models, is their linearity. Because a typical optimization problem is very complex, involving hundreds or thousands of variables and constraints, the mathematical problem of finding the optimum is extremely difficult. To render such problems tractable, modelers commonly introduce a number of simplifications. Among these is the assumption that the relationships in the system are linear. In fact, the most popular optimization technique, linear programming, requires that the objective function and all constraints be linear.

Linearity is mathematically convenient, but in reality it is almost always invalid. Consider, for example, a model of a firm's inventory distribution policies. The model contains a specific relationship between inventory and shipments—if the inventory of goods in the warehouse is 10 percent below normal, shipment may be reduced by, say, 2 percent since certain items will be out of stock. If the model requires this relationship to be linear, then a 20 percent shortfall will reduce shipments by 4 percent, a 30 percent shortfall by 6 percent, and so on. And when the shortfall is 100 percent? According to the model, shipments will still be 80 percent of normal. But obviously, when the warehouse is empty, no shipments are possible. The linear relationship within the model leads to an absurdity.

The warehouse model may seem trivial, but the importance of non-linearity is well demonstrated by the sorry fate of the passenger pigeon, *Ectopistes migratorius.* When Europeans first colonized North America, passenger pigeons were extremely abundant. Huge flocks of the migrating birds would darken the skies for days. They often caused damage to crops and were hunted both as a pest and as food. For years, hunting had little apparent impact on the population; the prolific birds seemed to reproduce fast enough to offset most losses. Then the number of pigeons began to decline—slowly at first, then rapidly. By 1914, the passenger pigeon was extinct.

The disappearance of the passenger pigeons resulted from the non-linear relationship between their population density and their fertility. In large flocks they could reproduce at high rates, but in smaller flocks their fertility dropped precipitously. Thus, when hunting pressure was great enough to reduce the size of a flock somewhat, the fertility in that flock also fell. The lower fertility lead to a further decrease in the population size, and the lower population density resulted in yet lower birth rates, and so forth, in a vicious cycle.

Unfortunately, the vast majority of optimizations models assume that the world is linear. There are, however, techniques available for solving certain non-linear optimization problems, and research is continuing.

Lack of Feedback. Complex systems in the real world are highly interconnected, having a high degree of feedback among sectors. The results of decisions feed back through physical, economic, and social channels to alter the conditions on which the decisions were originally made. Some models do not reflect this reality, however. Consider an optimization model that computes the best size of sewage treatment plants to build in an area. The model will probably assume that the amount of sewage needing treatment will remain the same, or that it will grow at a certain rate. But if water quality improves because of sewage treatment, the area will become more attractive and development will increase, ultimately leading to a sewage load greater than expected.

Models that ignore feedback effects must rely on *exogenous variables* and are said to have a narrow boundary. Exogenous variables are ones that influence other variables in the model but are not calculated by the model. They are simply given by a set of numerical values over time, and they do not change in response to feedback. The values of exogenous variables may come from other models but are most likely the product of an unexaminable mental model. The *endogenous variables*, on the other hand, are calculated by the model itself. They are the variables explained by the structure of the model, the ones for which the modeler has an explicit theory, the ones that respond to feedback.

Ignoring feedback can result in policies that generate unanticipated side effects or are diluted, delayed, or defeated by the system (Meadows 1982). An example is the construction of freeways in the 1950s and 1960s to alleviate traffic congestion in major US cities. In Boston it used to take half an hour to drive from the city neighborhood of Dorchester to the downtown area, a journey of only a few miles. Then a limited access highway network was built around the city, and travel time between Dorchester and downtown dropped substantially.

But there's more to the story. Highway construction led to changes that fed back into the system, causing unexpected side effects. Due to the reduction in traffic congestion and commuting time, living in outlying communities became a more attractive option. Farmland was turned into housing developments or paved over to provide yet more roads. The population of the suburbs soared, as people moved out of the center city. Many city stores followed their customers or were squeezed out by competition from the new suburban shopping malls. The inner city began to decay, but many people still worked in the downtown area—and they got there via the new highways. The result? Boston has more congestion and air pollution than before the highways were constructed, and the rush-hour journey from Dorchester to downtown takes half an hour, again.

In theory, feedback can be incorporated into optimization models, but the resulting complexity and non-linearity usually render the problem insoluble. Many optimization models therefore ignore most feedback effects. Potential users should be aware of this when they look at a model. They should ask to what degree important feedbacks have been excluded and how those exclusions might alter the assumptions and invalidate the results of the model.

Lack of Dynamics. Many optimization models are static. They determine the optimal solution for a particular moment in time without regard for how the optimal state is reached or how the system will evolve in the future. An example is the linear programming model constructed in the late 1970s by the US Forest Service, with the objective of optimizing the use of government lands. The model was enormous, with thousands of decision variables and tens of thousands of constraints, and it took months just to correct the typographical errors in the model's huge database. When the completed model was finally run, finding the solution required full use of a mainframe computer for days.

Despite the gigantic effort, the model prescribed the optimal use of forest resources for only a single moment in time. It did not take into account how harvesting a given area would affect its future ecological development. It did not consider how land-use needs or lumber prices might change in the future. It did not examine how long it would take for new trees to grow to maturity in the harvested areas, or what the economic and recreational value of the areas would be during the regrowth period. The model just provided the optimal decisions for a single year, ignoring the fact that those decisions would continue to influence the development of forest resources for decades.

Not all optimization models are static. The MARKAL model, for example, is a large linear programming model designed to determine the optimal choice of energy technologies. Developed at the Brookhaven National Laboratory in the US, the model produces as its output the best (least-cost) mix of coal, oil, gas, and other energy sources well into the next century. It requires various exogenous inputs, such as energy demands, future fuel prices, and construction and operating costs of different energy technologies. (Note that the model ignores feedbacks from energy supply to prices and demand.) The model is dynamic in the sense that it produces a "snapshot" of the optimal state of the system at five-year intervals.

The Brookhaven model is not completely dynamic, however, because it ignores delays. It assumes that people, seeing what the optimal mix is for some future year, begin planning far enough in advance so that this mix can actually be used. Thus the model does not, for example, incorporate construction delays for energy production facilities. In reality, of course, it takes time—often much longer than five years—to build power plants, invent new technologies, build equipment, develop waste management techniques, and find and transport necessary raw materials.

Indeed, delays are pervasive in the real world. The delays found in complex systems are especially important because they are a major source of system instability. The lag time required to carry out a decision or to perceive its effects may cause overreaction or may prevent timely intervention. Acid rain provides a good example. Although there is already evidence that damage to the forests of New England, the Appalachians, and Bavaria is caused by acid rain, many scientists suspect it will take years to determine exactly how acid rain is formed and how it affects the forests. Until scientific and then political consensus emerges, legislative action to curb pollution is not likely to be strong. Pollution control programs, once passed, will take years to implement. Existing power plants and other pollution sources will continue to operate for their functional lifetimes, which are measured in decades. It will require even longer to change settlement patterns and lifestyles dependent on the automobile. By the time sulfur and nitrogen oxide emissions are sufficiently reduced, it may be too late for the forests.

Delays are a crucial component of the dynamic behavior of systems, but—like nonlinearity—they are difficult to incorporate into optimization models. A common simplification is to assume that all delays in the model are of the same fixed length. The results of such models are of questionable value. Policy makers who use them in an effort to find an optimal course of action may discover, like the proverbial American tourist on the back roads of Maine, that "you can't get there from here."

<u>When To Use Optimization</u>

Despite the limitations discussed above, optimization techniques can be extremely useful. But they must be used for the proper problems. Optimization has substantially improved the quality of decisions in many areas, including computer design, airline scheduling, factory siting, and oil refinery operation. Whenever the problem to be solved is one of choosing the best from among a well-defined set of alternatives, optimization should be considered. If the meaning of *best* is also well defined, and if the system to be optimized is relatively static and free of feedback, optimization may well be the best technique to use. Unfortunately, these latter conditions are rarely true for the social, economic, and ecological systems that are frequently of concern to decision makers.

Look out for optimization models that purport to forecast actual behavior. The output of an optimization model is a statement of the best way to accomplish a goal. To interpret the results as a prediction of actual behavior is to assume that people in the real system will in fact make the optimal choices. It is one thing to say, "in order to maximize profits, people should make the following decisions," and quite another to say "people will succeed in maximizing profits, because they will make the following decisions." The former is a prescriptive statement of what to do, the latter a descriptive statement of what will actually happen.

Optimization models are valid for making prescriptive statements. They are valid for forecasting only if people do in fact optimize, do make the best possible decisions. It may seem reasonable to expect people to behave optimally—after all, wouldn't it be irrational to take second best when you could have the best? But the evidence on this score is conclusive: real people do not behave like optimization models. As discussed above, we humans make decisions with simple and incomplete mental models, models that are often based on faulty assumptions or that lead erroneously from sound assumptions to flawed solutions. As Herbert Simon puts it,

> The capacity of the human mind for formulating and solving complex problems is very small compared with the size of the problem whose solution is required for objectively rational behavior in the real world or even for a reasonable approximation to such objective rationality. (Simon 1957, p. 198)

Optimization models augment the limited capacity of the human mind to determine the objectively rational course of action. It should be remembered, however, that even optimization models must make simplifying assumptions in order to be tractable, so the most we can hope from them is an approximation of how people ought to behave. To model how people actually behave requires a very different set of modeling techniques, which will be discussed now.

Simulation

The Latin verb *simulare* means to imitate or mimic. The purpose of a simulation model is to mimic the real system so that its behavior can be studied. The model is a laboratory replica of the real system, a *microworld* (Morecroft 1988). By creating a representation of the system in the laboratory, a modeler can perform experiments that are impossible, unethical, or prohibitively expensive in the real world.

Simulations of physical systems are commonplace and range from wind tunnel tests of aircraft design to simulation of weather patterns and the depletion of oil reserves. Economists and social scientists also have used simulation to understand how energy prices affect the economy, how corporations mature, how cities evolve and respond to urban renewal policies, and how population growth interacts with food supply, resources, and the environment. There are many different simulation techniques, including stochastic modeling, system dynamics, discrete simulation, and role-playing games. Despite the differences among them, all simulation techniques share a common approach to modeling.

Optimization models are prescriptive, but simulation models are descriptive. A simulation model does not calculate what should be done to reach a particular goal, but clarifies what would happen in a given situation. The purpose of simulations may be *foresight (*predicting how systems might behave in the future under assumed conditions) or *policy design* (designing new decision-making strategies or organizational structures and evaluating their effects on the behavior of the system).

In other words, simulation models are "what if" tools. Often such "what if" information is more important than knowledge of the optimal decision. For example, during the 1978 debate in the US over natural gas deregulation, President Carter's original proposal was modified dozens of times by Congress before a final compromise was passed. During the congressional debate, the Department of Energy evaluated each version of the bill using a system dynamics model (Department of Energy 1979). The model did not indicate what ought to be done to maximize the economic benefits of natural gas to the nation. Congress already had its own ideas on that score. But by providing an assessment of how each proposal would affect gas prices, supplies, and demands, the model generated ammunition that the Carter administration could use in lobbying for its proposals.

Every simulation model has two main components. First it must include a representation of the physical world relevant to the problem under study. Consider for example a model that was built for the purpose of understanding why America's large cities have continued to decay despite massive amounts of aid and numerous renewal programs (Forrester 1969). The model had to include a representation of the physical components of the city—the size and quality of the infrastructure, including the stock of housing and commercial structures; the attributes of the population, such as its size and composition and the mix of skills and incomes among the people; flows (of people, materials, money, etc.) into and out of the city; and other factors that characterize the physical and institutional setting.

How much detail a model requires about the physical structure of the system will, of course, depend on the specific problem being addressed. The urban model mentioned above required only an aggregate representation of the features common to large American cities. On the other hand, a model designed to improve the location and deployment of fire fighting resources in New York City had to include a detailed representation of the streets and traffic patterns (Greenberger, Crenson, and Crissey 1976).

In addition to reflecting the physical structure of the system, a simulation model must portray the behavior of the actors in the system. In this context, behavior means the way in which people respond to different situations, how they make decisions. The behavioral component is put into the model in the form of decision-making rules, which are determined by direct observation of the actual decision-making procedures in the system.

Given the physical structure of the system and the decision-making rules, the simulation model then plays the role of the decision makers, mimicking their decisions. In the model, as in the real world, the nature and quality of the information available to decision makers will depend on the state of the system. The output of the model will be a description of expected decisions. The validity of the model's assumptions can be checked by comparing the output with the decisions made in the real system.

An example is provided by the pioneering simulation study of corporate behavior carried out by Cyert and March (1963). Their field research showed that department stores used a very simple decision rule to determine the floor price of goods. That rule was basically to mark up the wholesale cost of the items by a fixed percentage, with the value of the markup determined by tradition. They also noted, however, that through time the traditional markup adjusted very slowly, bringing it closer to the actual markup realized on goods when they were sold. The actual markup could vary from the normal markup as the result of several other decision rules: when excess inventory piled up on the shelves, a sale was held and the price was gradually reduced until the goods were sold; if sales goals were exceeded, prices were boosted. Prices were also adjusted toward those of competitors.

Cyert and March built a simulation model of the pricing system, basing it on these decision-making rules. The output of the model was a description of expected prices for goods. When this output was compared with real store data, it was found that the model reproduced quite well the actual pricing decisions of the floor managers.

<u>Limitations of Simulation</u>

Any model is only as good as its assumptions. In the case of simulation models, the assumptions consist of the descriptions of the physical system and the decision rules. Adequately representing the physical system is usually not a problem; the physical environment can be portrayed with whatever detail and accuracy is needed for the model purpose. Also, simulation models can easily incorporate feedback effects, nonlinearities, and dynamics; they are not rigidly determined in their structure by mathematical limitations as optimization models often are. Indeed, one of the main uses of simulation is to identify how feedback, nonlinearity, and delays interact to produce troubling dynamics that persistently resist solution. (For examples see Sterman 1985, Morecroft 1983, and Forrester 1969.)

Simulation models do have their weak points, however. Most problems occur in the description of the decision rules, the quantification of soft variables, and the choice of the model boundary.

Accuracy of the Decision Rules. The description of the decision rules is one potential trouble spot in a simulation model. The model must accurately represent how the actors in the system make their decisions, even if their decision rules are less than optimal. The model should respond to change in the same way the real actors would. But it will do this only if the model's assumptions faithfully describe the decision rules that are used under different circumstances. The model therefore must reflect the actual decision-making strategies used by the people in the system being modeled, including the limitations and errors of those strategies.

Unfortunately, discovering decision rules is often difficult. They cannot be determined from aggregate statistical data, but must be investigated first hand. Primary data on the behavior of the actors can be acquired through observation of actual decision making in the field, that is, in the boardroom, on the factory floor, along the sales route, in the household. The modeler must discover what information is available to each actor, examine the timeliness and accuracy of that information, and infer how it is processed to yield a decision. Modelers often require the skills of the anthropologist and the ethnographer. One can also learn about decision making through laboratory experiments in which managers operate simulated corporations (Sterman 1989). The best simulation modeling draws on extensive knowledge of decision making that has been developed in many disciplines, including psychology, sociology, and behavioral science.

Soft Variables. The majority of data are soft variables. That is, most of what we know about the world is descriptive, qualitative, difficult to quantify, and has never been recorded. Such information is crucial for understanding and modeling complex systems. Yet in describing decision making, some modelers limit themselves to hard variables, ones that can be measured directly and can be expressed as numerical data. They may defend the rejection of soft variables as being more scientific than "making up" the values of parameters and relationships for which no numerical data are available. How, they ask, can the accuracy of estimates about soft variables be tested? How can statistical tests be performed without numerical data?

Actually, there are no limitations on the inclusion of soft variables in models, and many simulation models do include them. After all, the point of simulation models is to portray decision making as it really is, and soft variables—including intangibles such as desires, product quality, reputation, expectations, and optimism – are often of critical importance in decision making. Imagine, for example, trying to run a school, factory, or city solely on the basis of the available numerical data. Without qualitative knowledge about factors such as operating procedures, organizational structure, political subtleties, and individual motivations, the result would be chaos. Leaving such variables out of models just because of a lack of hard numerical data is certainly less "scientific" than including them and making reasonable estimates of their values. Ignoring a relationship implies that it has a value of zero—probably the only value known to be wrong! (Forrester 1980)

Of course, all relationships and parameters in models, whether based on soft or hard variables, are imprecise and uncertain to some degree. Reasonable people may disagree as to the importance of different factors. Modelers must therefore perform sensitivity analysis to consider how their conclusions might change if other plausible assumptions were made. Sensitivity analysis should not be restricted to uncertainty in parameter values, but should also consider the sensitivity of conclusions to alternative structural assumptions and choices of model boundary.

Sensitivity analysis is no less a responsibility for those modelers who ignore soft variables. Apparently hard data such as economic and demographic statistics are often subject to large measurement errors, biases, distortions, and revisions. Unfortunately, sensitivity analysis is not performed or reported often enough. Many modelers have been embarrassed when third parties, attempting to replicate the results of a model, have found that reasonable alternative assumptions produce radically different conclusions. (See the discussion below of the experiment conducted by the Joint Economic Committee with three leading econometric models.)

Model Boundary. The definition of a reasonable model boundary is another challenge for the builders of simulation models. Which factors will be exogenous, which will be endogenous? What feedbacks will be incorporated into the model? In theory, one of the great strengths of simulation models is the capacity to reflect the important feedback relationships that shape the behavior of the system and its response to policies. In practice, however, many simulation models have very narrow boundaries. They ignore factors outside the expertise of the model builder or the interests of the sponsor, and in doing so they exclude important feedbacks.

The consequences of omitting feedback can be serious. An excellent example is provided by the Project Independence Evaluation System (PIES) model, used in the 1970s by the US Federal Energy Administration and later by the US Department of Energy. As described by the FEA, the purpose of the model was to evaluate different energy strategies according to these criteria: their impact on the development of alternative energy sources, their impact on economic growth, inflation, and unemployment; their regional and social impacts; their vulnerability to import disruptions; and their environmental effects (Federal Energy Administration 1974, p. 1).

Surprisingly, considering the stated purpose, the PIES model treated the economy as exogenous. The economy—including economic growth, interest rates, inflation, world oil prices, and the costs of unconventional fuels—was completely unaffected by the US domestic energy situation—including prices, policies, and production. The way the model was constructed, even a full embargo of imported oil or a doubling of oil prices would have no impact on the economy.

Its exogenous treatment of the economy made the PIES model inherently contradictory. The model showed that the investment needs of the energy sector would increase markedly as depletion raised the cost of getting oil out of the ground and synthetic fuels were developed. But at the same time, the model assumed that higher investment needs in the energy sector could be satisfied without reducing investment or consumption in the rest of the economy and without raising interest rates or inflation. In effect, the model let the economy have its pie and eat it too.

In part because it ignored the feedbacks between the energy sector and the rest of the economy, the PIES model consistently proved to be overoptimistic. In 1974 the model projected that by 1985 the US would be well on the way to energy independence: energy imports would be only 3.3 million barrels per day, and production of shale oil would be 250,000 barrels per day. Furthermore, these developments would be accompanied by oil prices of about $22 per barrel (1984 dollars) and by vigorous economic growth. It didn't happen. In fact, at the time this paper is being written (1988), oil imports are about 5.5. million per day, and the shale oil industry remains a dream. This situation prevails despite the huge reductions in oil demand that have resulted from oil prices of over $30 per barrel and from the most serious economic recession since the Great Depression.

A broad model boundary that includes important feedback effects is more important than a great amount of detail in the specification of individual components. It is worth noting that the PIES model provided a breakdown of supply, demand, and price for dozens of fuels in each region of the country. Yet its aggregate projections for 1985 weren't even close. One can legitimately ask what purpose was served by the effort devoted to forecasting the demand for jet fuel or naphtha in the Pacific Northwest when the basic assumptions were so palpably inadequate and the main results so woefully erroneous.

In fairness it must be said that the PIES model is not unique in the magnitude of its errors. Nearly all energy models of all types have consistently been wrong about energy production, consumption, and prices. The evidence shows clearly that energy forecasts actually lag behind the available information, reflecting the past rather than anticipating the future (Department of Energy 1983). A good discussion of the limitations of PIES and other energy models is available in the appendix of Stobaugh and Yergin (1979).

Overly narrow model boundaries are not just a problem in energy analysis. *The Global 2000 Report to the President* (Barney 1980) showed that most of the models used by US government agencies relied significantly on exogenous variables. Population models assumed food production was exogenous. Agriculture models assumed that energy prices and other input prices were exogenous. Energy models assumed that economic growth and environmental conditions were exogenous. Economic models assumed that population and energy prices were exogenous. And so on. Because they ignored important intersectoral feedbacks, the models produced inconsistent results.

Econometrics

Strictly speaking, econometrics is a simulation technique, but it deserves separate discussion for several reasons. First, it evolved out of economics and statistics, while most other simulation methods emerged from operations research or engineering. The difference in pedigree leads to large differences in purpose and practice. Second, econometrics is one of the most widely used formal modeling techniques. Pioneered by Nobel Prize-winning economists Jan Tinbergen and Lawrence Klein, econometrics is now taught in nearly all business and economics programs. Econometric forecasts are regularly reported in the media, and ready-to use statistical routines for econometric modeling are now available for many personal computers. And third, the well-publicized failure of econometric models to predict the future has eroded the credibility of all types of computer models, including those built for very different purposes and using completely different modeling techniques.

Econometrics is literally the measurement of economic relations, and it originally involved statistical analysis of economic data. As commonly practiced today, econometric modeling includes three stages – specification, estimation, and forecasting. First the structure of the system is specified by a set of equations. Then the values of the parameters (coefficients relating changes in one variable to changes in another) are estimated on the basis of historical data. Finally, the resulting output is used to make forecasts about the future performance of the system.

Specification. The model specification is the description of the model's structure. This structure consists of the relationships among variables, both those that describe the physical setting and those that describe behavior. The relationships are expressed as equations, and a large econometric model may have hundreds or even thousands of equations reflecting the many interrelationships among the variables.

For example, an econometric model of the macroeconomy typically will contain equations specifying the relationship between GNP and consumption, investment, government activity, and international trade. It also will include behavioral equations that describe how these individual quantities are determined. The modeler may expect, for instance, that high unemployment reduces inflation and vice versa, a relationship known as the Phillips curve. One of the equations in the model will therefore express the Phillips curve, specifying that the rate of inflation depends on the amount of unemployment. Another equation may relate unemployment to the demand for goods, the wage level, and worker productivity. Still other equations may explain wage level in terms of yet other factors.

Not surprisingly, econometrics draws on economic theory to guide the specification of its models. The validity of the models thus depends on the validity of the underlying economic theories. Though there are many flavors of economics, a small set of basic assumptions about human behavior are common to most theories, including modern neoclassical theory and the "rational expectations" school. These assumptions are: optimization, perfect information, and equilibrium.

In econometrics, people (economic agents, in the jargon), are assumed to be concerned with just one thing—maximizing their profits. Consumers are assumed to optimize the "utility" they derive from their resources. Decisions about how much to produce, what goods to purchase, whether to save or borrow, are assumed to be the result of optimization by individual decision makers. Non-economic considerations (defined as any behavior that diverges from profit or utility maximization) are ignored or treated as local aberrations and special cases.

Of course, to optimize, economic agents would need accurate information about the world. The required information would go beyond the current state of affairs; it also would include complete knowledge about available options and their consequences. In most econometric models, such knowledge is assumed to be freely available and accurately known.

Take, for example, an econometric model simulating the operation of a firm that is using an optimal mix of energy, labor, machines, and other inputs in its production process. The model will assume that the firm knows not only the wages of workers and the prices of machines and other inputs, but also the production attainable with different combinations of people and machines, even if those combinations have never been tried. Rational expectation models go so far as to assume that the firm knows future prices, technologies, and possibilities, and that it can perfectly anticipate the consequences of its own actions and those of competitors.

The third assumption is that the economy is in or near equilibrium nearly all of the time. If disturbed, it is usually assumed to return to equilibrium rapidly and in a smooth and stable manner. The prevalence of static thinking is the intellectual legacy of the pioneers of mathematics and economics. During the late nineteenth century, before computers or modern cybernetic theory, the crucial questions of economic theory involved the nature of the equilibrium state for different situations. Given human preferences and the technological possibilities for producing goods, at what prices will commodities be traded, and in what quantities? What will wages be? What will profits be? How will a tax or monopoly influence the equilibrium?

These questions proved difficult enough without tackling the more difficult problem of dynamics, of the behavior of a system in flux. As a result, dynamic economic theory—including the recurrent fluctuations of inflation, of the business cycle, of the growth and decline of industries and nations – remained primarily descriptive and qualitative long after equilibrium theory was expressed mathematically. Even now, dynamic behavior in economics tends to be seen as a transition from one equilibrium to another, and the transition is usually assumed to be stable.

The rich heritage of static theory in economics left a legacy of equilibrium for econometrics. Many econometric models assume that markets are in equilibrium at all times. When adjustment dynamics are modeled, variables are usually assumed to adjust in a smooth and stable manner toward the optimal, equilibrium value, and the lags are nearly always fixed in length. For example, most macroeconometric models assume that capital stocks of firms in the economy adjust to the optimal, profit-maximizing level, with a fixed lag of several years. The lag is the same whether the industries that supply investment goods have the capacity to meet the demand or not. (See, for example, Eckstein 1983 and Jorgenson 1963).

Yet clearly, when the supplying industries have excess capacity, orders can be filled rapidly; when capacity is strained, customers must wait in line for delivery. Whether the dynamic nature of the lag is expressed in a model does make a difference. Models that explicitly include the determinants of the investment delay will yield predictions significantly different from models that assume a fixed investment lag regardless of the physical capability of the economy to fill the demand (Senge 1980). In general, models that explicitly portray delays and their determinants will yield different results from models that simply assume smooth adjustments from one optional state to another.

Estimation. The second stage in econometric modeling is statistical estimation of the parameters of the model. The parameters determine the precise strengths of the relationships specified in the model structure. In the case of the Phillips curve, for example, the modeler would use past data to estimate precisely how strong the relationship between inflation and unemployment has been. Estimating the parameters involves statistical regression routines that are, in essence, fancy curve-fitting techniques. Statistical parameter estimates characterize the degree of correlation among the variables. They use historical data to determine parameter values that best match the data themselves.

All modeling methods must specify the structure of the system and estimate parameters. The use of statistical procedures to derive the parameters of the model is the hallmark of econometrics and distinguishes it from other forms of simulation. It gives econometricians an insatiable appetite for numerical data, for without numerical data they cannot carry out the statistical procedures used to estimate the models. It is no accident that the rise of econometrics went hand in hand with the quantification of economic life. The development of the national income and produce accounts by Simon Kuznets in the 1930s was a major advance in the codification of economic data, permitting consistent measures of economic activity at the national level for the first time. To this day all major macroeconometric models rely heavily on the national accounts data, and indeed macroeconomic theory itself has adapted to the national accounts framework.

Forecasting. The third step in econometric modeling is forecasting, making predictions about how the real system will behave in the future. In this step, the modeler provides estimates of the future values of the exogenous variables, that is, those variables that influence the other variables in the model but aren't themselves influenced by the model. An econometric model may have dozens of exogenous variables, and each must be forecast before the model can be used to predict.

<u>Limitations of Econometric Modeling</u>

The chief weak spots in econometric models stem from the assumptions of the underlying economic theory on which they rest: assumptions about the rationality of human behavior, about the availability of information that real decision makers do not have, and about equilibrium. Many economists acknowledge the idealization and abstraction of these assumptions, but at the same time point to the powerful results that have been derived from them. However, a growing number of prominent economists now argue that these assumptions are not just abstract—they are false. In his presidential address to the British Royal Economics Society, E. H. Phelps-Brown said:

> The trouble here is not that the behavior of these economic chessmen has been simplified, for simplification seems to be part of all understanding. The trouble is that the behavior posited is not known to be what obtains in the actual economy. (Phelps-Brown 1972, p. 4)

Nicholas Kaldor of Cambridge University is even more blunt:

> ...in my view, the prevailing theory of value – what I called, in a shorthand way, "equilibrium economics"—is barren and irrelevant as an apparatus of thought... (Kaldor 1972, p. 1237)

As mentioned earlier, a vast body of empirical research in psychology and organizational studies has shown that people do not optimize or act as if they optimize, that they don't have the mental capabilities to optimize their decisions, that even if they had the computational power necessary, they lack the information needed to optimize. Instead, they try to satisfy a variety of personal and organizational goals, use standard operating procedures to routinize decision making, and ignore much of the available information to reduce the complexity of the problems they face. Herbert Simon, in his acceptance speech for the 1978 Nobel Prize in economics, concludes:

> There can no longer be any doubt that the micro assumptions of the theory—the assumptions of perfect rationality—are contrary to fact. It is not a question of approximation; they do not even remotely describe the processes that human beings use for making decisions in complex situations (Simon 1979, p. 510).

Econometrics also contains inherent statistical limitations. The regression procedures used to estimate parameters yield unbiased estimates only under certain conditions. These conditions are known *as maintained hypotheses* because they are assumptions that must be made in order to use the statistical technique. The maintained hypotheses can never be verified, even in principle, but must be taken as a matter of faith. In the most common regression technique, ordinary least squares, the maintained hypotheses include the unlikely assumptions that the variables are all measured perfectly, that the model being estimated corresponds perfectly to the real world, and the random errors in the variables from one time period to another are completely independent. More sophisticated techniques do not impose such restrictive assumptions, but they always involve other a priori hypotheses that cannot be validated.

Another problem is that econometrics fails to distinguish between correlations and causal relationships. Simulation models must portray the causal relationships in a system if they are to mimic its behavior, especially its behavior in new situations. But the statistical techniques used to estimate parameters in econometric models don't prove whether a relationship is causal. They only reveal the degree of past correlation between the variables, and these correlations may change or shift as the system evolves. The prominent economist Robert Lucas (1976) makes the same point in a different context.

Consider the Phillips curve as an example. Though economists often interpreted the Phillips curve as a causal relationship—a policy trade-off between inflation and unemployment—it never did represent the causal forces that determine inflation or wage increases. Rather, the Phillips curve was simply a way of restating the past behavior of the system. In the past, Phillips said, low unemployment had tended to occur at the same time inflation was high, and vice-versa. Then, sometime in the early 1970s, the Phillips curve stopped working; inflation rose while unemployment worsened. Among the explanations given by economists was that the structure of the system had changed. But a modeler's appeal to "structural change" usually means that the inadequate structure of the model has to be altered because it failed to anticipate the behavior of the real system!

What actually occurred in the 1970s was that, when inflation swept prices to levels unprecedented in the industrial era, people learned to expect continuing increases. As a result of the adaptive feedback process of learning, they learned to deal with high inflation through indexing, COLAs, inflation-adjusting accounting, and other adjustments. The structure, the causal relationships of the system, did not change. Instead, causal relationships that had been present all along (but were dormant in an era of low inflation) gradually became active determinants of behavior as inflation worsened. In particular, the ability of people to adapt to continuing inflation existed all along but wasn't tested until inflation became high enough and persistent enough. Then the behavior of the system changed, and the historical correlation between inflation and unemployment broke down.

The reliance of econometric estimation on numerical data is another of its weaknesses. The narrow focus on hard data blinds modelers to less tangible but no less important factors. They ignore both potentially observable quantities that haven't been measured yet and ones for which no numerical data exist. (Alternatively, they may express an unmeasured factor with a proxy variable for which data already exists, even though the relationship between the two is tenuous—as when educational expenditure per capita is used as a proxy for the literacy of a population.)

Among the factors excluded from econometric models because of the hard data focus are many important determinants of decision making, including desires, goals, and perceptions. Numerical data may measure the results of human decision making, but numbers don't explain how or why people made particular decisions. As a result, econometric models cannot be used to anticipate how people would react to a change in decision-making circumstances.

Similarly, econometric models are unable to provide a guide to performance under conditions that have not been experienced previously. Econometricians assume that the correlations indicated by the historical data will remain valid in the future. In reality, those data usually span a limited range and provide no guidance outside historical experience. As a result, econometric models are often less than robust: faced with new policies or conditions, the models break down and lead to inconsistent results.

An example is the model used by Data Resources, Inc. in 1979 to test policies aimed at eliminating oil imports. On the basis of historical numerical data, the model assumed that the response of oil demand to the price of oil was rather weak—a 10 percent increase in oil price caused a reduction of oil demand of only 2 percent, even in the long run. According to the model, for consumption to be reduced by 50 percent (enough to cut imports to zero at the time), oil would have to rise to $800 per barrel. Yet at that price, the annual oil bill for the remaining 50 percent would have exceeded the total GNP for that year, an impossibility (Sterman 1981). The model's reliance on historical data led to inconsistencies. (Today, with the benefit of hindsight, economists agree that oil demand is much more responsive to price than was earlier believed. Yet considering the robustness of the model under extreme conditions could have revealed the problem much earlier.)

Validation is another problem area in econometric modeling. The dominant criterion used by econometric modelers to determine the validity of an equation or a model is the degree to which it fits the data. Many econometrics texts (e.g., Pindyck and Rubinfeld 1976) teach that the statistical significance of the estimated parameters in an equation is an indicator of the correctness of the relationship. Such views are mistaken. Statistical significance indicates how well an equation fits the observed data; it does not indicate whether a relationship is a correct or true characterization of the way the world works. A statistically significant relationship between variables in an equation shows that they are highly correlated and that the apparent correlation is not likely to have been the result of mere chance. But it does not indicate that the relationship is causal at all.

Using statistical significance as the test of model validity can lead modelers to mistake historical correlations for causal relationships. It also can cause them to reject valid equations describing important relationships. They may, for example, exclude an equation as statistically insignificant simply because there are few data about the variables, or because the data don't contain enough information to allow the application of statistical procedures.

Ironically, a lack of statistical significance does not necessarily lead econometric modelers to the conclusion that the model or the equation is invalid. When an assumed relationship fails to be statistically significant, the modeler may try another specification for the equation, hoping to get a better statistical fit. Without recourse to descriptive, micro-level data, the resulting equations may be ad hoc and bear only slight resemblance to either economic theory or actual behavior. Alternatively, the modelers may attempt to explain the discrepancy between the model and the behavior of the real system by blaming it on faulty data collection, exogenous influences, or other factors.

The Phillips curve again provides an example. When it broke down, numerous revisions of the equations were made. These attempts to find a better statistical fit met with limited success. Some analysts took another tack, pointing to the oil price shock, Russian wheat deal, or other one-of-a-kind events as the explanation for the change. Still others argued that there had been structural changes that caused the Phillips curve to shift out to higher levels of unemployment for any given inflation rate.

These flaws in econometrics have generated serious criticism from within the economic profession. Phelps-Brown notes that because controlled experiments are generally impossible in economics "running regressions between time series is only likely to deceive" (Phelps-Brown 1972, p. 6). Lester Thurow notes that econometrics has failed as a method for testing theories and is now used primarily as a "showcase for exhibiting theories." Yet as a device for advocacy, econometrics imposes few constraints on the prejudices of the modeler. Thurow concludes:

> By simple random search, the analyst looks for the set of variables and functional forms that give the best equations. In this context the best equation is going to depend heavily upon the prior beliefs of the analyst. If the analyst believes that interest rates do not affect the velocity of money, he finds a 'best' equation that validates his particular prior belief. If the analyst believes that interest rates do affect the velocity of money, he finds a 'best' equation that validates this prior belief. (Thurow 1983, pp. 107-8)

But the harshest assessment of all comes from Nobel laureate Wassily Leontief:

> Year after year economic theorists continue to produce scores of mathematical models and to explore in great detail their formal properties; and the econometricians fit algebraic functions of all possible shapes to essentially the same sets of data without being able to advance, in any perceptible way, a systematic understanding of the structure and the operations of a real economic system. (Leontief 1982, p. 107; see also Leontief 1971.)

But surely such theoretical problems matter little if the econometric models provide accurate predictions. After all, the prime purpose of econometric models is short-term prediction of the exact future state of the economy, and most of the attributes of econometrics (including the use of regression techniques to pick the "best" parameters from the available numerical data and the extensive reliance on exogenous variables) have evolved in response to this predictive purpose.

Unfortunately, econometrics fails on this score also; in practice, econometric models do not predict very well. The predictive power of econometric models, even over the short-term (one to four years), is poor and virtually indistinguishable from that of other forecasting methods. There are several reasons for this failure to predict accurately.

As noted earlier, in order to forecast, the modeler must provide estimates of the future values of the exogenous variables, and an econometric model may have dozens of these variables. The source of the forecasts for these variables may be other models but usually is the intuition and judgment of the modeler. Forecasting the exogenous variables consistently, much less correctly, is difficult.

Not surprisingly, the forecasts produced by econometric models often don't square with the modeler's intuition. When they feel the model output is wrong, many modelers, including those at the "big three" econometric forecasting firms—Chase Econometrics, Wharton Econometric Forecasting Associates, and Data Resources – simply adjust their forecasts. This fudging, or add factoring as they call it, is routine and extensive. The late Otto Eckstein of Data Resources admitted that their forecasts were 60 percent model and 40 percent judgment ("Forecasters Overhaul Models of Economy in Wake of 1982 Errors," *Wall Street Journal, 17* February 1983). *Business Week* ("Where Big Econometric Models Go Wrong," 30 March 1981) quotes an economist who points out that there is no way of knowing where the Wharton model ends and the model's developer, Larry Klein, takes over. Of course, the adjustments made by add factoring are strongly colored by the personalities and political philosophies of the modelers. In the article cited above, the *Wall Street Journal* quotes Otto Eckstein as conceding that his forecasts sometimes reflect an optimistic view: "Data Resources is the most influential forecasting firm in the country...If it were in the hands of a doom-and-gloomer, it would be bad for the country."

In a revealing experiment, the Joint Economic Committee of Congress (through the politically neutral General Accounting Office) asked these three econometric forecasting firms (DRI, Chase, and Wharton) to make a series of simulations with their models, running the models under different assumptions about monetary policy. One set of forecasts was "managed" or add factored by the forecasters at each firm. The other set consisted of pure forecasts, made by the GAO using the untainted results of the models. As an illustration of the inconsistencies revealed by the experiment, consider the following: when the money supply was assumed to be fixed, the DRI model forecast that after ten years the interest rate would be 34 percent, a result totally contrary to both economic theory and historical experience. The forecast was then add factored down to a more reasonable 7 percent. The other models fared little better, revealing both the inability of the pure models to yield meaningful results and the extensive ad hoc adjustments made by the forecasters to render the results palatable (Joint Economic Committee 1982).

Add factoring has been criticized by other economists on the grounds that it is unscientific. They point out that, although the mental models used to add factor are the mental models of seasoned experts, these experts are subject to the same cognitive limitations other people face. And whether good or bad, the assumptions behind add factoring are always unexaminable.

The failure of econometric models have not gone unnoticed. A representative sampling of articles in the business press on the topic of econometric forecasting include the following headlines:

> "1980: The Year The Forecasters Really Blew It." (Business Week, 14 July 1980).
>
> "Where The Big Econometric Models Go Wrong." (Business Week, 30 March 1981).
>
> "Forecasters Overhaul Models of Economy in Wake of 1982 Errors." (Wall Street Journal, 17 February, 1983).
>
> "Business Forecasters Find Demand Is Weak in Their Own Business: Bad Predictions Are Factor." (Wall Street Journal, 7 September 1984).
>
> "Economists Missing The Mark: More Tools, Bigger Errors." (New York Times, 12 December 1984).

The result of these failures has been an erosion of credibility regarding all computer models no matter what their purpose, not just econometric models designed for prediction. This is unfortunate. Econometric models are poor *forecasting* tools, but well-designed simulation models can be valuable tools for *foresight* and *policy design.* Foresight is the ability to anticipate how the system will behave if and when certain changes occur. It is not forecasting, and it does not depend on the ability to predict. In fact, there is substantial agreement among modelers of global problems that exact, point prediction of the future is neither possible nor necessary:

> ...at present we are far from being able to predict social-system behavior except perhaps for carefully selected systems in the very short term. Effort spent on attempts at precise prediction is almost surely wasted, and results that purport to be such predictions are certainly misleading. On the other hand, much can be learned from models in the form of broad, qualitative, conditional understanding—and this kind of understanding is useful (and typically the only basis) for policy formulation. If your doctor tells you that you will have a heart attack if you do not stop smoking, this advice is helpful, even if it does not tell you exactly when a heart attack will occur or how bad it will be. (Meadows, Richardson, and Bruckmann 1982, p. 279)

Of course, policy evaluation and foresight depend on an accurate knowledge of the history and current state of the world, and econometrics has been a valuable stimulus to the development of much-needed data gathering and measurement by governments and private companies. But econometric models do not seem well-suited to the types of problems of concern in policy analysis and foresight. Though these models purport to simulate human behavior, they in fact rely on unrealistic assumptions about the motivations of real people and the information available to them. Though the models must represent the physical world, they commonly ignore dynamic processes, disequilibrium, and the physical basis for delays between actions and results. Though they may incorporate hundreds of variables, they often ignore soft variables and unmeasured quantities. In real systems the feedback relationships between environmental, demographic, and social factors are usually as important as economic influences, but econometric models often omit these because numerical data are not available. Furthermore, econometrics usually deals with the short term, while foresight takes a longer view. Over the time span that is of concern in foresight, real systems are likely to deviate from their past recorded behaviors, making unreliable the historical correlations on which econometric models are based.

Checklist for the Model Consumer

The preceding discussion has focused on the limitations of various modeling approaches in order to provide potential model consumers with a sense of what to look out for when choosing a model. Despite the limitations of modeling, there is no doubt that computer models can be and have been extremely useful foresight tools. Well-built models offer significant advantages over the often faulty mental models currently in use.

The following checklist provides further assistance to decision makers who are potential model users. It outlines some of the key questions that should be asked to evaluate the validity of a model and its appropriateness as a tool for solving a specific problem.

> What is the problem at hand? What is the problem addressed by the model?
>
> What is the boundary of the model? What factors are endogenous? Exogenous? Excluded? Are soft variables included? Are feedback effects properly taken into account? Does the model capture possible side effects, both harmful and beneficial?
>
> What is the time horizon relevant to the problem? Does the model include as endogenous components those factors that may change significantly over the time horizon?
>
> Are people assumed to act rationally and to optimize their performance? Does the model take non-economic behavior (organizational realities, non-economic motives, political factors, cognitive limitations) into account?
>
> Does the model assume people have perfect information about the future and about the way the system works, or does it take into account the limitations, delays, and errors in acquiring information that plague decision makers in the real world?

Are appropriate time delays, constraints, and possible bottlenecks taken into account?

Is the model robust in the face of extreme variations in input assumptions?

Are the policy recommendations derived from the model sensitive to plausible variations in its assumptions?

Are the results of the model reproducible? Or are they adjusted (add factored) by the model builder?

Is the model currently operated by the team that built it? How long does it take for the model team to evaluate a new situation, modify the model, and incorporate new data?

Is the model documented? Is the documentation publicly available? Can third parties use the model and run their own analyses with it?

Conclusions

The inherent strengths and weaknesses of computer models have crucial implications for their application in foresight and policy analysis. Intelligent decision-making requires the appropriate use of many different models designed for specific purposes—not reliance on a single, comprehensive model of the world. To repeat a dictum offered above, "Beware the analyst who proposes to model an entire social or economic system rather than a problem." It is simply not possible to build a single, integrated model of the world, into which mathematical inputs can be inserted and out of which will flow a coherent and useful understanding of world trends.

To be used responsibly, models must be subjected to debate. A cross-disciplinary approach is needed; models designed by experts in different fields and for different purposes must be compared, contrasted, and criticized. The foresight process should foster such review.

The history of global modeling provides a good example. The initial global modeling efforts, published in *World Dynamics* (Forrester 1971) and *The Limits To Growth* (Meadows et al. 1972) provoked a storm of controversy. A number of critiques appeared, and other global models were soon developed. Over a period of ten years, the International Institute for Applied Systems Analysis (IIASA) conducted a program of analysis and critical review in which the designers of global models were brought together. Six major symposia were held, and eight important global models were examined and discussed. These models had different purposes, used a range of modeling techniques, and were built by persons with widely varying backgrounds. Even after the IIASA conferences, there remain large areas of methodological and substantive disagreement among the modelers. Yet despite these differences, consensus did emerge on a number of crucial issues (Meadows, Richardson, and Bruckmann 1982), including the following:

Physical and technical resources exist to satisfy the basic needs of all the world's people into the foreseeable future.

Population and material growth cannot continue forever on a finite planet.

Continuing "business as usual" policies in the next decades will not result in a desirable future nor even in the satisfaction of basic human needs.

Technical solutions alone are not sufficient to satisfy basic needs or create a desirable future.

The IIASA program on global modeling represents the most comprehensive effort to date to use computer models as a way to improve human understanding of social issues. The debate about the models created agreement on crucial issues where none had existed. The program helped to guide further research and provided a standard for the effective conduct of foresight in both the public and private sectors.

At the moment, model-based analyses usually take the form of studies commissioned by policy makers. The clients sit and wait for the final reports, largely ignorant of the methods, assumptions, and biases that the modelers put into the models. The policy makers are thus placed in the role of supplicants awaiting the prophecies of an oracle. When the report finally arrives, they may, like King Croesus before the Oracle at Delphi, interpret the results in accordance with their own preconceptions. If the results are unfavorable, they may simply ignore them. Policy makers who use models as black boxes, who accept them without scrutinizing their assumptions, who do not examine the sensitivity of the conclusions to variations in premises, who do not engage the model builders in dialogue, are little different from the Delphic supplicants or the patrons of astrologers. And these policy makers justly alarm critics, who worry that black box modeling abdicates to the modelers and the computer a fundamental human responsibility (Weizenbaum 1976).

No one can (or should) make decisions on the basis of computer model results that are simply presented, "take 'em or leave 'em." In fact, the primary function of model building should be educational rather than predictive. Models should not be used as a substitute for critical thought, but as a tool for improving judgment and intuition. Promising efforts in corporations, universities, and public education are described in Senge 1989; Graham, Senge, Sterman, and Morecroft 1989; Kim 1989; and Richmond 1987.

Towards that end, the role of computer models in policy making needs to be redefined. What is the point of computer modeling? It should be remembered that we all use models of some sort to make decisions and to solve problems. Most of the pressing issues with which public policy is concerned are currently being handled solely with mental models, and those mental models are failing to resolve the problems. The alternative to continued reliance on mental models is computer modeling. But why turn to computer models if they too are far from perfect?

The value in computer models derives from the differences between them and mental models. When the conflicting results of a mental and a computer model are analyzed, when the underlying causes of the differences are identified, both of the models can be improved.

Computer modeling is thus an essential part of the educational process rather than a technology for producing answers. The success of this dialectic depends on our ability to create and learn from shared understandings of our models, both mental and computer. Properly used, computer models can improve the mental models upon which decisions are actually based and contribute to the solution of the pressing problems we face.

References

Barney, Gerald O., ed. 1980. *The Global 2000 Report to the President.* 3 vols. Washington, DC: US Government Printing Office.

Business Forecasters Find Demand Is Weak in Their Own Business: Bad Predictions Are Factor. *Wall Street Journal, 7* September 1984.

Cyert, R., and March, J. 1963. A *Behavioral Theory of the Firm.* Englewood Cliffs, NJ: Prentice Hall.

Department of Energy. 1979. *National Energy Plan II.* DOE/TIC-10203. Washington, DC: Department of Energy.

______. 1983. *Energy Projections to the Year 2000.* Washington, DC: Department of Energy, Office of Policy, Planning, and Analysis.

Eckstein, O. *1983. The DRl Model of the US Economy.* New York, McGraw Hill.

Economists Missing the Mark: More Tools, Bigger Errors. *New York Times*, 12 December 1984.

Federal Energy Administration. 1974. *Project Independence Report.* Washington, DC: Federal Energy Administration.

Forecasters Overhaul Models of Economy in Wake of 1982 Errors. *Wall Street Journal, 17* February, 1983.

Forrester, Jay W. 1969. *Urban Dynamics.* Cambridge, Mass.: MIT Press.

______. 1971. *World Dynamics.* Cambridge, Mass.: MIT Press.

______. 1980. Information Sources for Modeling the National Economy. *Journal of the American Statistical Association* 75(371):555-574.

Graham, Alan K.; Senge, Peter M.; Sterman, John D.; and Morecroft, John D. W. 1989. Computer Based Case Studies in Management Education and Research. In *Computer-Based Management of Complex Systems,* eds. P. Milling and E. Zahn, pp. 317-326. Berlin: Springer Verlag.

Grant, Lindsey, ed. 1988. *Foresight and National Decisions: The Horseman and the Bureaucrat.* Lanham, Md.: University Press of America.

Greenberger, M., Crenson, M. A., and Crissey, B. L. 1976. *Models in the Policy Process.* New York: Russell Sage Foundation.

Hogarth, R. M. 1980. *Judgment and Choice.* New York: Wiley.

Joint Economic Committee. 1982. *Three Large Scale Model Simulations of Four Money Growth Scenarios.* Prepared for subcommittee on Monetary and Fiscal Policy, 97th Congress 2nd Session, Washington, DC

Jorgenson, D. W. 1963. Capital Theory and Investment Behavior. *American Economic Review* 53:247-259.

Kahneman, D., Slovic, P., and Tversky, A. 1982. *Judgment Under Uncertainty: Heuristics and Biases.* Cambridge: Cambridge University Press.

Kaldor, Nicholas. 1972. The Irrelevance of Equilibrium Economics. *The Economic Journal* 82: 1237-55.

Kim, D. 1989. Learning Laboratories: Designing a Reflective Learning Environment. In *Computer-Based Management of Complex Systems, P.* Milling and E. Zahn, eds., pp. 327-334. Berlin: Springer Verlag

Leontief, Wassily. 1971. Theoretical Assumptions and Nonobserved Facts. *American Economic Review* 61(1):1-7.

______. 1982. Academic Economics. *Science*, 217: 104-107.

Lucas, R. 1976. Econometric Policy Evaluation: A Critique. In *The Phillips Curve and Labor Markets,* K. Brunner and A. Meltzer, eds. Amsterdam: North-Holland.

Meadows, Donella H.; Meadows, Dennis L.; Randers, Jorgen.; and Behrens, William W. 1972. *The Limits to Growth.* New York: Universe Books.

Meadows, Donella H. 1982. Whole Earth Models and Systems. *CoEvolution Quarterly,* Summer 1982, pp. 98-108.

Meadows, Donella H.; Richardson, John; and Bruckmann, Gerhart 1982. *Groping in The Dark.* Somerset, NJ: Wiley.

Morecroft, John D. W. 1983. System Dynamics: Portraying Bounded Rationality. *Omega II:* 131-142.

______. 1988. System Dynamics and Microworlds for Policy Makers. *European Journal of Operational Research* 35(5):301-320.

______. 1980: The Year The Forecasters Really Blew It *Business Week,* 14 July 1980.

Phelps-Brown, E. H. 1972. The Underdevelopment of Economics. *The Economic Journal* 82:1-10.

Pindyck, R., and Rubinfeld, D. 1976. *Econometric Models and Economic Forecasts.* New York: McGraw Hill.

Richmond, B. 1987. *The Strategic Forum.* High Performance Systems, Inc., 45 Lyme Road, Hanover, NH 03755, USA.

Senge, Peter M. 1980. A System Dynamics Approach to Investment Function Formulation and Testing. *Socioeconomic Planning Sciences* 14:269-280.

______. 1989. Catalyzing Systems Thinking Within Organizations. In *Advances in Organization Development,* F. Masaryk, ed., forthcoming.

Simon, Herbert. 1947. *Administrative Behavior.* New York: MacMillan.

______. 1957. *Models of Man.* New York: Wiley.

______. 1979. Rational Decisionmaking in Business Organizations. *American Economic Review* 69:493-513.

Sterman, John D. 1981. The Energy Transition and the Economy: A System Dynamics Approach. Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge.

______. 1985. A Behavioral Model of the Economic Long Wave. *Journal of Economic Behavior and Organization* 6(1):17-53.

______. 1989. Modeling Managerial Behavior: Misperceptions of Feedback in a Dynamic Decision Making Experiment. *Management Science* 35(3):321-339.

Stobaugh, Robert and Yergin, Daniel. 1979. *Energy Future.* New York: Random House.

Thurow, Lester. 1983. *Dangerous Currents.* New York: Random House.

Weizenbaum, J. 1976. *Computer Power and Human Reason: from Judgment to Calculation.* San Francisco: *W.* H. Freeman.

Where The *Big* Econometric Models Go Wrong. *Business Week,* 30 March 1981.

The author wishes to acknowledge that many of the ideas expressed here emerged from discussions with or were first formulated by, among others, Jay Forrester, George Richardson, Peter Senge, and especially Donella Meadows, whose book Groping in the Dark (Meadows, Richardson, and Bruckmann 1982), was particularly helpful.