ENERGY MODEL COMPARISON: AN OVERVIEW

J.L. Sweeney

Large-Scale Energy Models -
Prospects and Potential

EMF PP6.4
1983

Energy Modeling Forum
Stanford University
Stanford, California 94305

ACKNOWLEDGMENTS

The activities of the Energy Modeling Forum are funded by gifts from corporate affiliates and by grants from or contracts with the U.S. Department of Energy and the Electric Power Research Institute. Corporate affiliates are:

Aluminum Company of America	Pacific Gas and Electric Company
Chevron U.S.A.	Pembina Resources Limited
Conoco, Inc.	Rocky Mountain Energy Company
Exxon Company, U.S.A.	Royal Bank of Canada
Gas Research Institute	Standard Oil Company of Ohio
General Motors Corporation	Union Carbide Corporation
Natomas North America, Inc.	Union Oil Company of California

Reprinted from:

Large-Scale Energy Models

Prospects and Potential

Edited by Robert M. Thrall, Russell G. Thompson, and Milton L. Holloway

AAAS Selected Symposium **73**

Published by Westview Press, Boulder, Colorado 80301

James L. Sweeney

12. Energy Model Comparison: An Overview

Abstract

Organization and manipulation of complex data bases always requires models of some sort, and all models, being of necessity simplifications, are imperfect. Systematic model comparisons can produce benefits through identification of errors, clarification of disagreements, and guidance for model selection. Model comparison categories include methods and equations, forecast, aggregate behavior, and model regeneration. Modeling-the-model helps to structure and to communicate these comparison results.

Introduction

The policymaker or planner confronting any system as complex and controversial as the energy system must assimilate and analyze a vast array of information about technology, existing policies, regulations, economics, and political forces, and anticipate the reactions of constituents and other interested parties to any action. The planning and using of the requisite data base always requires models of some form--explicit or implicit, mathematical or intuitive--since a model is simply an organized set of cause-and-effect relationships and basic data. Although formal mathematical models may not be utilized, information about possible futures or about the effects of alternative actions on these futures must at least implicitly be shaped through modeling, unless the policymaker or planner depends primarily upon soothsayers for advice.

In organizing and manipulating information, judgments must be made about the relative importance of and relationships among diverse pieces of information. The basic cause-and-effect relationships accepted by the individual are shaped by previous experiences, judgments, evidence, and professional intuition. Therefore, the model developed by one person to describe a complex system will normally differ, often in important ways, from that developed by another person. Such model differences may lead naturally to conflicting conclusions as to what policy choices should be accepted, even when there is a common goal.

Whenever differences in policy or planning recommendations stem from discrepancies in the information base, an important aspect of debate resolution involves a comparison of the implicit and explicit assumptions made by participants in the debate. A system of examining assumptions is valuable for understanding the fundamental points of agreement and disagreement among the parties and for striving toward consensus.

While comparison of assumptions is crucial for achieving mutual understanding and consensus, the examination of information sources is often hampered by obstacles common to all applications of modeling, particularly formal modeling (7).

(1) Communication between developers and users of formal models is extremely limited.

(2) Formal and mental models often remain as black boxes, opaque to potential users, and translucent only to their developers.

(3) Although the various models may exhibit widely differing behavior, and therefore provide different answers to any question, they are seldom used in a comparative mode to address policy or planning questions.

A formal model comparison can alleviate many of the difficulties that are faced by modelers and model users. Development of a formal methodology of model comparison and assessment will complement the usual procedure of model assessment that naturally occurs with the development of a model. Formal assessments of a model focus on the evaluation of the quality of a model, the accuracy of its projections, or similarly normative concepts. Assessments imply a comparison between a structural model and a mental model or between two formal models. Although assessment implies comparison, the converse is not necessarily true: comparisons can be made without necessarily reaching conclusions about the relative quality of models.

I will next discuss the process of model comparison, without addressing the more difficult process of model assessment once the comparisons are accomplished. Benefits of formal, extensive comparison efforts are discussed, followed by an outline of some comparison methods. Finally I present a purely subjective view of the relations between the methods utilized and the benefits obtained by model comparisons. While formal model comparisons and assessments are being conducted intensively through groups such as the MIT Model Assessment Laboratory, the Utility Modeling Forum, and the Texas Energy Advisory Council, most examples will be drawn from the author's experience with the Energy Modeling Forum (12).

Reasons for Model Comparison

The essence of modeling is simplification. From the wealth of information potentially available, relatively few crucial cause-and-effect relationships are selected. These relationships form the basis of the model. Relationships and data that are judged less important are given little or no role in the modeling process. This simplification permits an emphasis on important issues and allows easy manipulation of a complex system. Thus the simplification process relies on individual human judgment. As a result, there are at least as many models as "modelers." Even if all modelers were to have the same basic perceptions of the systems being examined, they would still invariably develop different models, based upon their different time constraints, goals, styles, research budgets, organizational talents, motivations, and judgment.

In addition, basic perceptions and assumptions vary: probably no two modelers possess the same information or judgments about the relative importance of various relationships in the system. Since a model is a product of human understanding, limits to that understanding are translated into limits in our ability to represent the inherent complexities of the world accurately. What we understand imperfectly, we model imperfectly.

Because of this dilemma, modelers should examine the implications of each simplification. This examination inherently involves a comparison of models. At the earliest stages of research, the comparisons are normally made among several alternative conceptual structures formulated only into mental models. At later stages, comparisons are made between the evolving formal model and the results of mental models. Such comparisons are considered by most modelers as a necessary part of the model development process. At more mature stages of model development, systematic comparisons among the various formal models and mental models can further enhance understanding of the specific modeling simplifications.

Thus, the issue is not whether model comparison and assessment is important. Model comparisons of some form are conducted whenever formal models are developed or utilized and whenever people attempt to resolve differences of opinion. The resolution of those differences or the choice of one answer over another at least implicitly includes a comparative assessment of the several sources of information. The dilemma faced by modelers is the question of the depth of the model comparison. Model comparison is costly and time consuming. Resources devoted to comparison exercises may imply fewer resources available for model development or application. The appropriate mix of activities is not obvious.

Benefits of a formal model comparison are the identification of errors, clarification of disagreements, and guidance for model selection. These three areas of model improvement will be explored in turn to delineate the benefits that are possible with a formal model comparison.

Error Identification

Every modeler makes errors in the process of developing a model. The goal is to identify and correct those errors as early as possible. Some are identified when the formal model is first compared with a mental model. Some errors are not apparent by comparison with mental models but may persist until more quantitative comparisons with other formal models are conducted.

The simplest type of mistake, computer coding error, often can be identified by comparison between the evolving formal model and a mental model. However, some more subtle coding errors may persist until formal model comparisons have been conducted.

Conceptual errors may be more difficult to identify without formal comparison exercises. A common conceptual error is to calibrate a model based upon data which bear the right label but which measure the wrong concept. For example, oil resources can loosely be identified as recoverable resources or as oil-in-place. The three-to-one difference can raise havoc unless one knows which concept is being employed. Or, more subtle differences can occur between data on production of crude oil versus all petroleum liquids. Errors of this type have been found in completed models through the Energy Modeling Forum (EMF) studies.

Another example of a conceptual error is provided by the experience of EMF 2, "Coal in Transition" (2). This study analyzed several models of coal production and distribution that used optimization algorithms to simulate a competitive equilibrium. In one model, the optimization was based implicitly on the assumption that producers could charge different prices to different customers for the same commodity. Comparison of results from that model with results of other models made the error apparent--the model was implicitly simulating a monopoly solution when the competitive solution was desired. That observation led to restructuring the solution algorithm so as to modify the implicit behavioral assumption of the model to that of a competitive equilibrium system.

Clarification of Disagreements

The second major benefit of formal model comparisons is the clarification of points of contention and disagreement. This advantage itself serves as a major stimulus for the growing professional interest in model comparison and assessment. When inputs of the various models are standardized, the results and projections may still vary significantly. These remaining variations may motivate a search for differences in modeling assumptions, data, or parameters--differences which may be fundamental to policy conclusions. Underlying differences--"contention points" (13, 7)--can be examined, debated, and researched. In fortunate situations this process may lead to resolution of the differences. In any event, the isolation of the basic disagreement tends to focus and sharpen the debate and may lead to enhanced understanding of the fundamental issues.

For example, differing assumptions about how OPEC nations set production quantities can lead to differences in the forecasts of world oil price. Some models include assumptions that OPEC suppliers maximize a discounted stream of revenues net of costs (monopoly or cartel pricing). Others, such as the Department of Energy/Oil Market Simulation (DOE/OMS) model, include an implicit assumption that OPEC quantity trajectories are independent of oil prices or consumer response. When the two classes of models are exercised to project the impacts on world oil prices of reductions in U.S. oil imports, different answers are produced by the two: the DOE/OMS model typically projects a larger price impact than do the monopoly models. If the two classes of models were used to estimate the economic value to the U.S. of reducing oil imports, then two different answers would emerge.

The differences in price impacts seem to be the result of these behavioral assumptions. Monopoly pricing requires a production rate such that the demand elasticity facing the monopolist exceeds unity. In contrast, many analyses, including many conducted with the DOE/OMS model, embody an assumption that the demand elasticities facing OPEC are significantly smaller than unity. The price impact of a given quantity change is inversely related to the elasticity of demand for oil. Therefore, the structure of monopoly implies that these projected oil price impacts of reduced imports will tend to be smaller than the projected impacts based upon models such as DOE's Oil Market Simulation (8).

The above example does not tell us which results are more nearly correct, since the extent to which OPEC currently is behaving as a monopolist is subject to intense debate. However, through understanding the implicit contention point the debate over the benefits of reducing imports can be clarified.

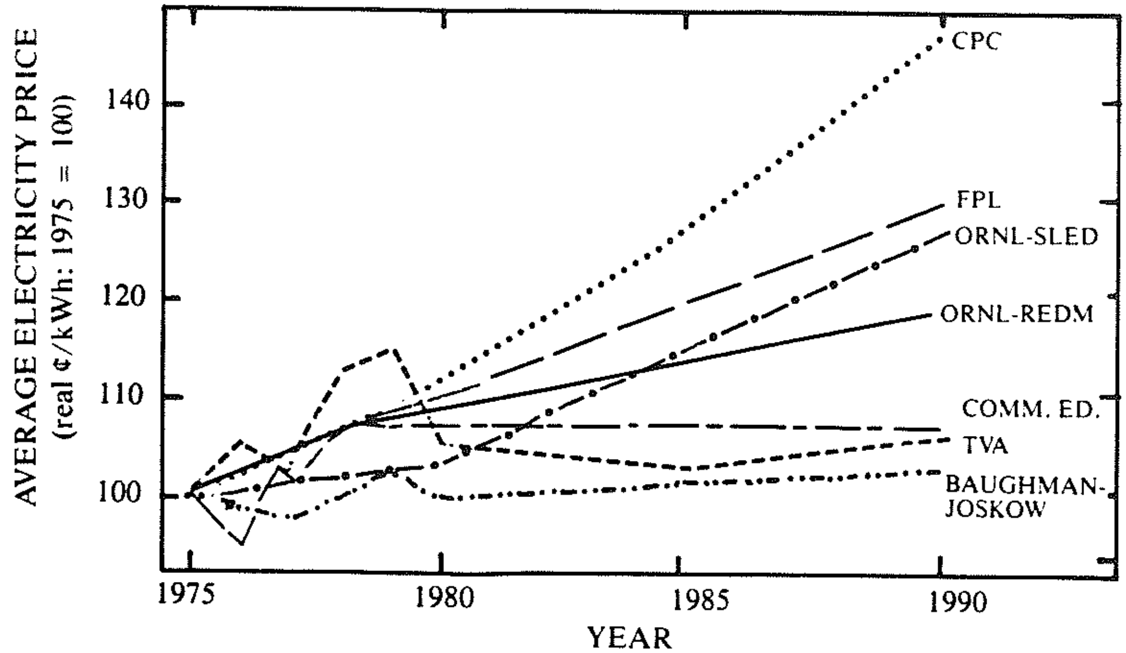

Figure 1. Reference Case Electricity Price Assumptions. Source: Electric Load Forecasting: Probing the Issues with Models.

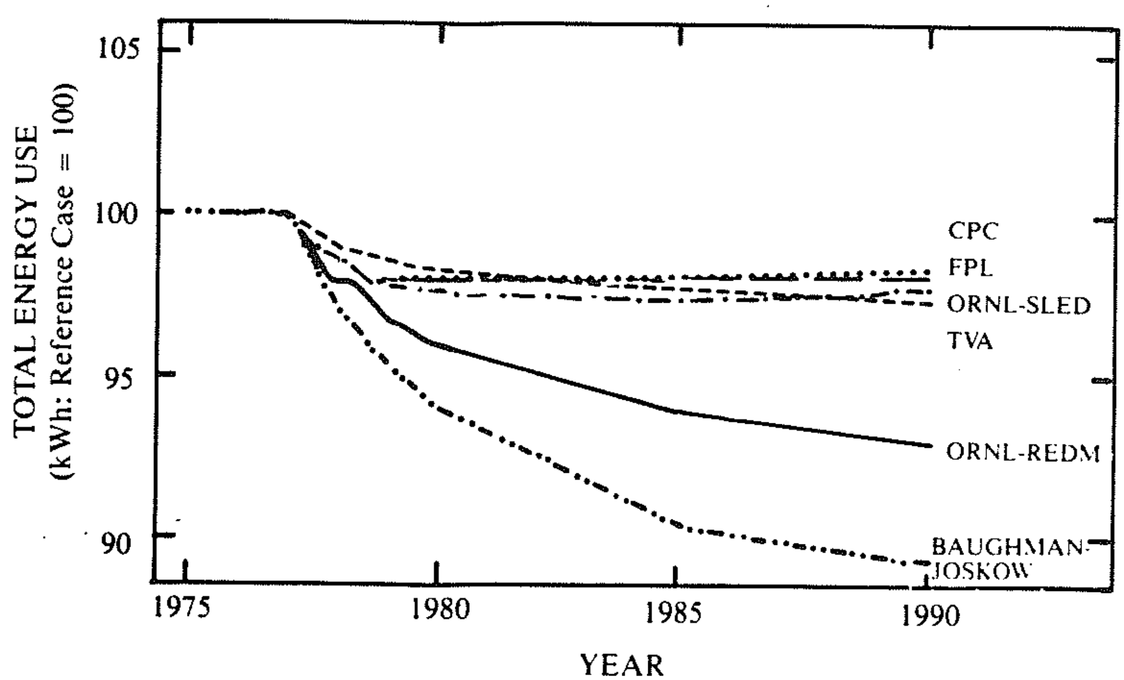

Figure 2. Change in Electricity Consumption with 10 Percent Price Increase Relative to Reference Case. Source: Electric Load Forecasting: Probing the Issues with Models.

Discussion of contention points may hasten diffusion of new ideas into many models, since this discussion may enhance basic understanding of the modeling issues. This flow of information into the body of analytical tools may be in the collective interest. However, the originator of a novel approach may be ambivalent about an instant diffusion of ideas into the competitor's models, especially if the only acknowledgment appears in a footnote of the mathematical addendum to Appendix 7.

A second example of model comparisons for clarifying contention points is EMF 3: "Electric Load Forecasting, Probing the Issues with Models" (3). In this study, the participants examined the responsiveness to electricity and natural gas prices of several electricity demand models. It was generally believed by the participants in the study group that electricity prices would continue to rise significantly (Figure 1). Thus, differences in the price elasticity of demand for electricity are fundamental to differences in forecasts of needed capital expansion for electric utilities. Figure 2 shows the results of model comparisons. One group of models had electricity price elasticities of demand equal to about -0.2, while two models--the Baughman-Joskow model and the Oak Ridge SLED model--showed elasticities of -0.7 and -1.0. The working group as a whole could not agree as to which behavior was closer to reality, although most working group members developed personal convictions. Vociferous disagreement arose on the correct resolution of this point. However, group members agreed on the value of actively debating and communicating those differences. Many hoped that publication of the study would facilitate and motivate a scientific debate, particularly within the electric utility industry, as to how significantly electricity price increases could be expected to reduce electricity demand growth.

The general category, clarification of disagreements, can be delineated into a number of specific areas as follows:

- Identification and communication of model limitations
- Identification of forecast uncertainty
- Enhancement of system understanding
- Guidance for future research

An examination of these specific benefits follows.

Identification and Communication of Model Limitation

Every mental and formal model has limitations. Limits of application should not be embarrassments. However, the failure to identify and communicate these limitations increases the probability that the model will be misused. Identification of the limitations of a model is the first necessary step for avoiding such misuse. This can be accomplished through discussion of equations, through examination of sensitivity tests, and so on. Even when the limits are apparent to the developer of a model, they may not be discernible to those who have not been closely associated with the specific modeling project.

For example, reduced form econometric models of gasoline demand were developed using data that predate the corporate average fuel efficiency standards on new cars. Thus, the structural change associated with the recent implementation of these standards implies that those reduced form models will tend to both over-estimate gasoline demand and over-estimate the role of higher gasoline prices in reducing demand (4, 9). What had not been a limitation to these models now has become a subtle, but important and poorly understood limitation. Comparison of the current generation of structural gasoline demand models with the reduced form econometric models can crystallize understanding of this type of limitation.

Model benefits depend upon communicating the limitations to other modelers and model users. The essence of truth-in-modeling is disclosure: disclosure of methods, assumptions, data, approximations, sensitivity results, insights, advancements, and embarrassments. The converse "untruth in modeling" is normally associated with a failure to allocate sufficient resources to the examination of models and communication of those ideas. Full disclosure is a most difficult task faced by modelers. Lack of resources is probably the single most critical barrier to full communication. In addition, political issues, competition, gamesmanship, and contractual obligations encourage the modeler to bypass the discussion of weak points and to concentrate on strengths and insights.

Serious effort on the part of modelers and model users is needed to overcome the barriers. For example, model assessment by third parties has certain advantages; see for example (11). Third-party assessment often provides an opportunity for the modeler to step back and to ask broad questions about key limitations. The assessor is often not initially aware of the complete structure of the model and thus has the opportunity to provide a new perspective. Since professional recognition for assessment activities is usually associated with the identification and communication of the limits of a model, third-party assessment activities may be quite effective.

Group exercises concentrating on the comparison of models' aggregate behavior are also useful. Such exercises can identify and solidify understanding of model capabilities and limitations and can facilitate communications to a broad audience of potential model users. Since all models have limitations and since both strengths and limitations of models being compared can be discussed, the modelers collectively benefit from the truth-in-modeling encouraged by such an exercise.

Identification of Forecast Uncertainty

Model comparisons make us more humble about our ability to forecast anything, using either mental or formal models, with precision. More positively, model comparisons reveal the full range of uncertainty of any forecasts. In modeling, three types of uncertainty exist: in inputs, in parameters, and in structure. These sources of uncertainty can each be addressed through model comparisons.

In order to examine the uncertainties of a forecast, modelers will often vary a set of critical input assumptions to the models over a range of values and observe changes in outputs. For example, this is a procedure used in the 1977 Annual Report to Congress (15) by the Energy Information Administration. Such sensitivity tests can provide a mapping from uncertainties in critical inputs to uncertainties in the output projections. These sensitivity tests allow researchers to identify the pivotal uncertain data elements to be input to the model. A high degree of uncertainty about a pivotal data element will severely limit the precision of any forecast and should motivate an effort to better estimate this element. Thorough scrutiny of these uncertainties through model comparison can provide more precise estimates of uncertainty.

A second element of uncertainty that is less frequently examined is the relationship between the parametric estimates and the real system being modeled. Typically, parameters are estimated using either statistical or engineering data methods along with professional judgment. Uncertainty in outputs associated with uncertainty of these parameters can be examined using a Monte Carlo simulation, randomly choosing alternative possible values of those parameters. Such simulations conceptually allow a mapping from uncertainties in critical parameters to uncertainties in the outputs.

Both types of uncertainties could be examined simultaneously, through use of Monte Carlo simulation techniques. As long as parameter uncertainties and input uncertainties are not correlated, the joint output uncertainty will be greater than the measurement obtained by allowing only one of these to vary. Therefore, examination of only input uncertainty will lead to underestimates of the real uncertainty about outputs.

The third source of uncertainty can be thought of as structural uncertainty. Inevitably forecast errors exist because the modeling process always results in a model structure that differs from reality, often in important ways. And often there is great uncertainty as to which of several structures best approximate reality. For example, there may be alternative paradigms utilized to forecast the impacts of policy choices. The debate in economics between the Macroeconomists influenced by Keynesian concepts and the Monetarists provides such an example. Even more dramatic are the paradigm differences among the so-called radical economists and the neoclassical economists. At a less basic level are the differences cited earlier among the alternative conceptual structures underlying models of world oil pricing. These differences imply a deep uncertainty that goes beyond differences in inputs or parameters.

Forecast uncertainty stemming from structural uncertainty is poorly understood and difficult to analyze. It is even unclear how one could conceptually define a measure of structural uncertainty since the set of feasible alternative structures cannot be observed and can only be vaguely imagined.

Some understanding of the degree of structural uncertainty can be obtained by comparison of several models that incorporate different representations of the same system. At least such comparisons help one to develop a more profound humility about the quality of our knowledge of the world than would be developed if forecasts from only one model were used.

Model comparisons do not fully answer the questions posed by structural uncertainty. However, they do provide a starting point for future analysis of uncertainty, since they identify the areas of uncertainty, explore the quantitative significance, and motivate further analysis.

Enhancement of System Understanding

Another important benefit resulting from a formal comparison is increased understanding of the system being modeled. There are a number of examples in the energy area. Models by Edward Hudson and Dale Jorgenson (10) and Ernie Berndt and David Wood (17) have forced a reevaluation of the relationships between energy prices and the rate of capital formation in the U.S. economy. The debate between Berndt/Wood and James Griffin (14, 18) concerning their models of capital formation has led to an enhanced understanding of capital-energy interrelationships.

Work produced from the first Energy Modeling Forum working group (1), under the leadership of Bill Hogan, reexamined the popular belief that reductions in energy would inherently lead to almost proportionate reductions in economic growth. Attempts to compare and explain differences between economic impacts projected using various energy-economy models provided insight for Bill Hogan's and Alan Manne's paper, "Energy-Economy Interactions: the Fable of the Elephant and the Rabbit?"(6) This paper provided important insights about the relationship between the elasticity of demand for energy and the economic impact of reduced energy availability.

Guidance for Model Selection

The last category of model comparison benefits is guidance in selection of appropriate models for specific applications. For most analysis, it is important to choose a model that matches the desired function because the choice of model may greatly influence the results obtained.

Analysis of the impacts of an oil import quota demonstrates this idea. The PILOT model (19), which uses a very small elasticity of demand, would show a very large impact of a quota on GNP, while the Hudson-Jorgenson model (10), which has a much higher demand elasticity, would show a much smaller impact. The projected impacts on world prices obtained by utilizing a formal or mental model, such as the Salant model (16), which envisions OPEC as a profit-maximizing monopoly or cartel, will be very different from those obtained by a model such as the DOE/OMS, which implicitly views the world oil market as competitive. The latter class of model would imply that the quota, by reducing the demand for world oil, would reduce oii prices. The former class--monopoly models--might indicate that a broadly applied quota which reduces the elasticity of demand facing OPEC will motivate price <u>increases</u>.

The extent to which the model comparison exercise should be carried is a function of the importance of the decision or policy and the potential influence of analyses on the ultimate decision. Clearly, when the issue is as important as the imposition of oil import quotas by the U.S. and its allies, it is inexcusable to fail to use a number of formal models, comparing their results with those obtained using purely mental models. Failure to do so may result in decisions based on an inadequate information base.

For some decisions, however, it is not economical or feasible to conduct an adequate model comparison exercise. The urgency of a decision may require expedience or the sensitivity of an issue may demand a minimal number of people involved. In these cases it is valuable to have an information base about the relative behavior of several models. This information base can allow a shortcut to understanding the assumptions that are implicitly made by choice of a specific model. The informed user can adapt the results from those models that most closely correspond to his or her interpretation of the evidence. If the model has already been selected (which is often the case with "inhouse" model development), the model user can adjust the results according to his or her judgment and knowledge of other models.

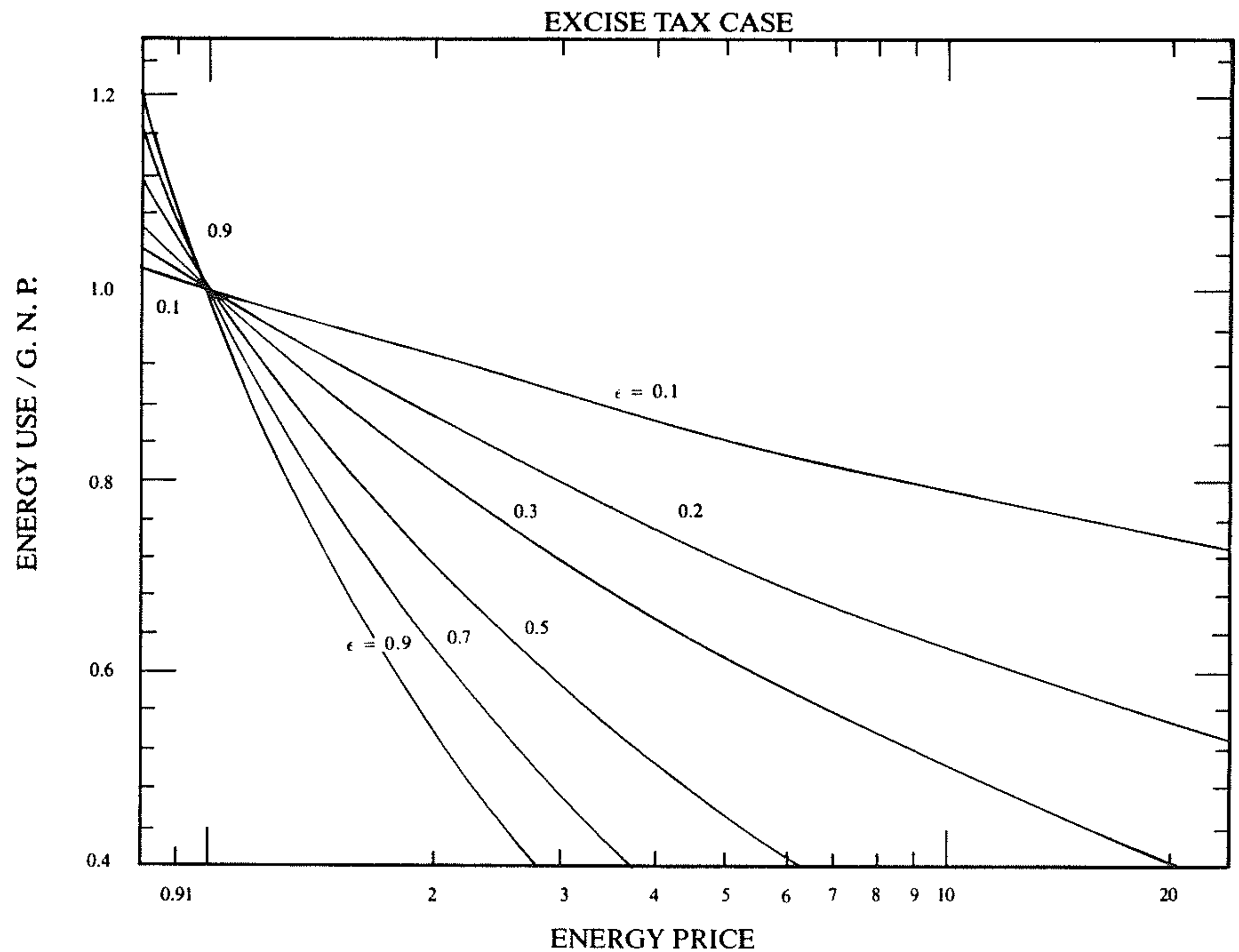

Figure 3. Projected ratio of energy consumption to GNP for varying Demand Elasticities (ε). Source: Aggregate Elasticity of Energy Demand.

For example, if it were necessary to project the use of energy to the year 2000, model selection could be influenced by knowledge of the responsiveness of different models to higher energy prices. The degree of responsiveness is quantified by the aggregate elasticity of demand for energy implicit in the model. This elasticity is commonly measured at one of three points in the supply chain: primary energy (at the wellhead or minemouth), secondary energy (at the refinery gate or electric utility busbar), or delivered energy (at the gas pump or meter). Figure 3 shows how the projected ratio of energy consumption to GNP varies with aggregate energy price for various demand elasticities measured at the secondary level. For a given future price of energy, the projected energy/GNP ratio depends upon the aggregate elasticity of energy demand implied in the model. The estimate of aggregate elasticity of energy demand will fundamentally shape projections. The importance of this measurement of elasticity for forecasting energy consumption encouraged the Energy Modeling Forum to compare the elasticities of sixteen energy demand models used in their EMF 4 study (4).

In Table 1, the models have been classified on the basis of the parameter estimation procedure used--statistical, engineering, or judgmental. Aggregate elasticities of demand at the secondary level are shown for the separate consuming sectors of the various models. Such a tabulation can be useful for guiding users in the selection of a model for specific analyses. Presumably, that choice will be based upon the evidence, as interpreted by the user, of the responsiveness of energy demand to prices. On a more cynical note, such a table could be used to select that model which gives the answer most convenient for the user. Even for such cynical purposes, the existence of the data-base gives both the primary user of the model and other users--adversaries, allies, and impartial observers--a firm understanding in order to debate specific forecasts, analyses, and conclusions.

Methods of Model Comparison

Implicit in the preceding discussion of model comparison benefits are some notions as to how those comparisons might be conducted. This section makes explicit those implicit notions. Four broad categories of model comparison are discussed: methods and equations, forecast, aggregate behavior, and model regeneration. In addition, "modeling-the-model" helps in structuring and communicating these comparisons.

Methods and Equations Comparisons

The most frequently developed model comparisons typically describe the methodologies incorporated in the various models and compare and contrast several equations from the models. Such

Table 1

Aggregate Elasticity of Energy Demand

SECTOR	STATISTICAL		ENGINEERING		JUDGMENTAL	
RESIDENTIAL	GRIFFIN OECD	.9	BECOM	.6		
			HIRST RESIDENTIAL	.4		
	MEFS	.5				
	PINDYCK	1.0				
RESIDENTIAL/ COMMERCIAL	BAUGHMAN-JOSKOW	.8	BECOM	.5	EPM	.5
	BESOM/H-J	.7				
	MEFS	.5				
COMMERCIAL	MEFS	.5	BECOM	.3	JACKSON	.4
			JACKSON COMMERCIAL	.4		
COMMERCIAL/ INDUSTRIAL	GRIFFIN OECD	.3				
	PINDYCK	.7				
INDUSTRIAL	BAUGHMAN-JOSKOW	.4	ISTUM	.2	EPM	.7
	BESOM/H-J	.5				
	MEFS	.2				
TRANSPORTATION *	FEA-FAUCETT	.1			EPM	.4
	GRIFFIN OECD	.5				
	BESOM/H-J	.2				
	SWEENEY AUTO	.5				
	WHARTON MOVE	.2				
	PINDYCK	.5				
	MEFS	.3				
ALL SECTORS	BAUGHMAN-JOSKOW **	.6			ETA-MACRO	.2
	GRIFFIN OECD	.5			FOSSIL 1	.1
	MEFS	.3			FOSSIL 1/ CONSERV.	.2
	PINDYCK	.7			PARIKH	
	BESOM/H-J	.4			WEM	.1
					EPM LL)	.6

* The FEA-Faucett, Wharton, and Sweeney results are for automobile gasoline only. These are 15 year elasticities.

** Excludes the transportation sector

comparisons are often based only upon the existing model documentation, possibly enhanced by discussions with the model developers. Often, they can be recognized by a title: "A State-of-the-Art Review of . . . ." Model comparisons of this type are also familiar to readers of academic literature, since a summary of previous work is a normal component of academic journal articles. In the MIT Energy Laboratory this type of model assessment is incorporated in the "Overview Assessment" phase of analysis (11). This type of comparison is the "descriptive" phase where the predominant methodology is common sense and professional judgment.

There are three basic elements of Methods and Equations comparison. The first is a discussion of the basic structure and underlying theory of the models. The relationship between the theory and its implementation into the model's structure is a crucial element of this component. The goal is to pinpoint important differences between theories and between various interpretations of a theory.

The second component evaluates the techniques of measurement and estimation used in the models. This allows the analyst to ponder possible biases in model behavior. More importantly, it allows judgment of the dependability of parameter estimation.

A key aspect of methods and equations comparison that improves communication is the reproduction of key equations and parameter values. For models having thousands of equations, this step may be impossible for more than a few of those equations. However, even in the most complicated models, often only a limited set of equations need to be examined carefully. Most of the equations are necessary only for details.

A part of a methods and equations comparison is a delineation between the new model and the "state-of-the-art" knowledge. Such comparisons provide a basic perspective for the subsequent comparisons. For example, at the MIT Energy Laboratory a review of the literature occurs early in the assessment procedure. This helps to place a specific model within a context of alternative models or approaches to an issue.

Forecast Comparisons

The second category of model comparison involves contrasting the predictions of the models. Forecast comparisons can be conducted on four levels. The first level is essentially a descriptive comparison of the array of forecasts. Since it is a cursory analysis, it alone seldom gives insight into the source of differences.

The second step attempts to adjust the various forecasts to a common basis. Forecasters generally will have used widely differing assumptions about the future values of the driving variables. Knowledge of those differing input values and knowledge of the mapping from input changes into output changes together allow adjustments of the forecasts to common bases.

The third and more complete level requires operation of at least one model. It then is possible to choose input values identical to those of a published forecast so as to allow a baseline comparison. This procedure was used, for example, in the early development of the Project Independence Evaluation System (PIES) at the Federal Energy Administration. Projections were made for prices remaining at pre-embargo levels, and these were compared against the Dupree-West forecasts published by the U.S. Department of the Interior (20) before the steep rise in energy prices. It was reassuring that the PIES projections fairly closely matched the Dupree-West forecasts for a low price situation. More importantly, errors in the model were discovered early through these comparisons.

The fourth phase of forecast comparison involves examination of all the models through operation with standard data. This assures that forecast variations stem only from model differences and not from the specific data inputs. The Energy Modeling Forum typically uses this procedure.

Another type of comparison that is closely related to the forecast procedure is called a "backcasting" comparison. The procedure requires the operation of each model with historical standardized data. Then the forecast "predicts" an event that has already occurred.

The Conference on Econometrics and Mathematical Economics (CEME), sponsored by the National Bureau of Economic Research and the National Science Foundation (21), provided a forum for the comparison of macroeconomic models. Inputs have been set at historical values in order to backcast the macroeconomic behavior that actually occurred. Summary statistics have been compared, describing how well the various models performed in these structured tests. Such backcasts, when feasible, provide valuable information.

Forecast comparisons, by themselves, do not necessarily expose the precise reasons for differences in the predictions. The information derived from the procedure may be of limited value when considered by itself. As a component of model comparison, the forecast exercise can prove to be crucial.

Aggregate Behavior Comparison

Most models contain many complex relationships and provide detailed projections, often disaggregated by sector, product, location, and so on. These detailed relationships are often important for precise delineation among policy options or for projections of differential impacts among locations, sectors, or products. But these detailed relationships may not be important for understanding the more aggregate outputs from the models, in fact, may help hide the more basic forces. Therefore, it is often valuable to develop comparisons among the broad aggregate input-output relations of the various models. This is the process of aggregate behavior comparison.

For example, some oil supply models project production of crude oil by region, by specific gravity of the oil, and by sulfur content. Inputs required are current and future world oil prices, tax structure, detailed price control rules, and so on. Aggregate behavior comparisons may ignore this detail and only examine the total lower-48-state oil production rate and the influence of a proportionate increase in all oil prices on this total.

Several examples of aggregate behavior comparisons can be cited. The CEME has developed aggregate behavior comparisons of macroeconomic models, focusing most attention on the various multipliers--government expenditure, tax, investment. The MIT Model Assessment Laboratory also incorporates this type of exercise in their "In-depth" phase of model assessment. The Energy Modeling Forum studies include extensive efforts to characterize the aggregate behavior of models, based upon common scenarios run by each modeler. For instance, in EMF 4: Aggregate Elasticity of Energy Demand (4), energy prices and economic growth assumptions were standardized for each of nine scenarios, predicting different rates of change of energy prices. The aggregate elasticity of energy demand was characterized for each model, based upon the outputs. Results appeared in Table 1. In EMF 5: U.S. Oil and Gas Supply (5), the responsiveness of total oil supply and total gas supply to assumptions about geological potential, oil price trajectories, or policy changes was characterized. Other examples include characterization of the influence of appliance efficiency standards, of average price and price structure, or of natural gas prices on electricity demand (EMF 3; see Figure 2); or the reductions in economic growth associated with restrictions on the availability of energy (EMF 1) (1).

Aggregate behavior comparisons normally include three key elements:

- Standardization of inputs for the models
- A set of structural sensitivity tests

- Summary characterization and quantification of key aggregate relationships

In contrast to the methods and equation comparison, the aggregate behavior of the models can be compared, even though the models may have different structures. For example, we can compare the response of oil supply to price in an intertemporal optimization, linear programming model (LORENDAS), a mixed econometric-engineering model (AGA-TERA), and a purely econometric model (Rice-Smith), even though these models have vastly different structures (5). However, it seems impossible to compare these models on an equation-by-equation basis.

Model Regeneration

The final mode of comparison may be called model regeneration. Critical relationships are examined by systematically replacing one component of the model (parameter set, equation, or system of equations) with an alternative version of the same component. By this procedure a family of models is generated based on one primary model. Changes in the model behavior then can be examined. An example of model regeneration is the MIT Energy Laboratory "Counter-Analysis" activities used in the in-depth evaluation of the Baughman/Joskow Regionalized Electricity Model (11). While the most costly mode of comparison, model regeneration exercises offer perhaps the greatest potential for model assessment of any method discussed above.

Modeling-the-Model

The final exercise of model comparison, "modeling-the-model," provides a means to consolidate the results from the other comparisons and a forum for interpretation of the results. The modeling-the-model idea requires the development of a simple model characterized by a very limited number of key parameters. This method can roughly mimic the actual model in order to illustrate, analyze, and communicate the structure of complex models.

For example, for EMF 1, Bill Hogan and Alan Manne postulated a simple constant-elasticity-of-substitution production function between energy and all other inputs to the economy. Two parameters were free, the value share of energy in the economy and the elasticity of substitution between energy and other productive factors. Differences in the aggregate elasticity of substitution implicit in the model led to fundamental differences in the projected economic growth reductions associated with restrictions on energy availability. It was shown that the aggregate elasticity of substitution could be closely approximated by the aggregate elasticity of energy demand discussed in previous sections of this paper. The relationship between energy availability and economic growth for various aggregate demand elasticities (ranging from 0.1 to 0.9) is illustrated in Figure 4. While this very simple structure could not capture many elements of the more complicated models, it captured perhaps the most important three issues. This simple model-of-models allows individuals to understand the role of elasticity of substitution in governing model-based projections of energy-economy interactions.

The modeling-the-model technique has been used within EMF 5 to emphasize the importance of the shape of the finding curve, the mathematical relationship between new discoveries of oil per additional unit of exploration and cumulative exploration for oil. The assumed shape of the finding curve can greatly influence the model's projections of the ultimate quantities of oil discovered and the price elasticity of the cumulative discoveries.

Within many models is a maintained assumption that the discoveries per unit of additional exploration (feet drilled, wells completed, etc.) is an exponentially declining function of cumulative exploration. This assumption of an exponential finding curve implies that the discoveries per additional well are always proportional to undiscovered reserves. A simple alternative "model-of-the-models" embodies an assumption that the discoveries per additional exploration are always proportional to undiscovered reserves, raised to the power B. If B=1, then the finding curve is exponential, corresponding to most of the complex models. Low values of B imply that the productivity of drilling is almost independent of remaining reserves, while a large value of B implies that the productivity of drilling declines sharply as additional reserves are found. For all values of B the postulated finding curve can be calibrated to the same estimates of initially undiscovered resources and initial rate of oil discoveries per well.

Figure 5 shows results from the simple model, plotting the fraction of resources which will be cumulatively discovered under a profit-maximizing regime as a function of the "relative price." This relative price is the value of a discovery relative to the initial exploration cost per discovery. Relative price is roughly proportional to the price of oil. The initial exploration cost per discovery is the cost per well divided by the the initial discoveries per well. As B ranges from 0.6 to 3.0, the fraction of resources cumulatively discovered by a profit maximizing competitive industry drops sharply, varying by a factor of 2.0 or more for low relative price. Figure 6 plots the elasticity of cumulative discoveries with respect to the value of a discovery for various B's. This supply elasticity drops as B decreases, varying by a factor of 3.0 or more as B varies from 0.6 to 3.0.

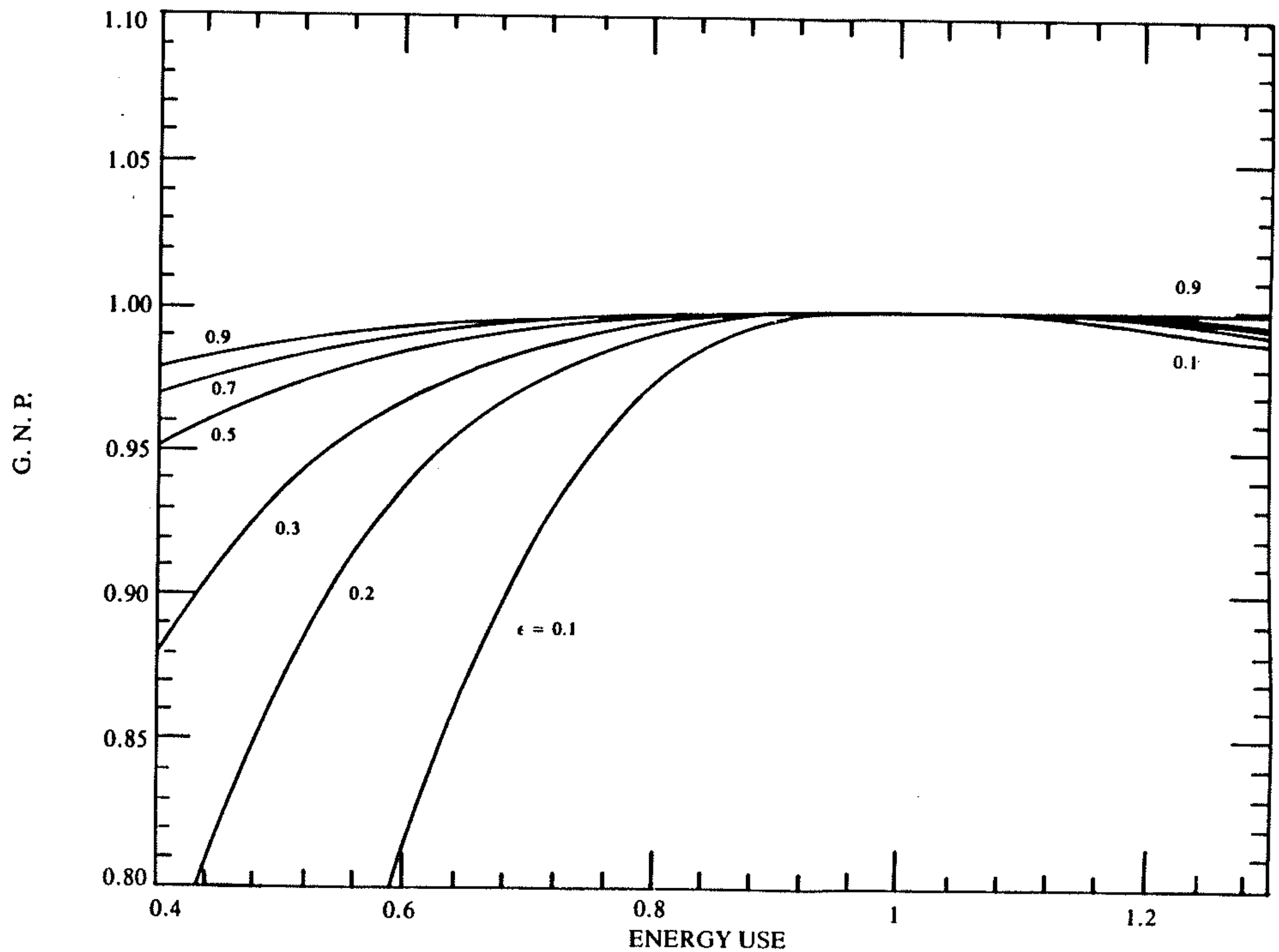

Figure 4. Impact of reduced energy use on GNP: Excise Tax Case. Source: Derived from Energy Modeling Forum Working Paper 4.4., "Energy Modeling Forum 4: Aggregate Elasticity of Energy Demand," October 1979.

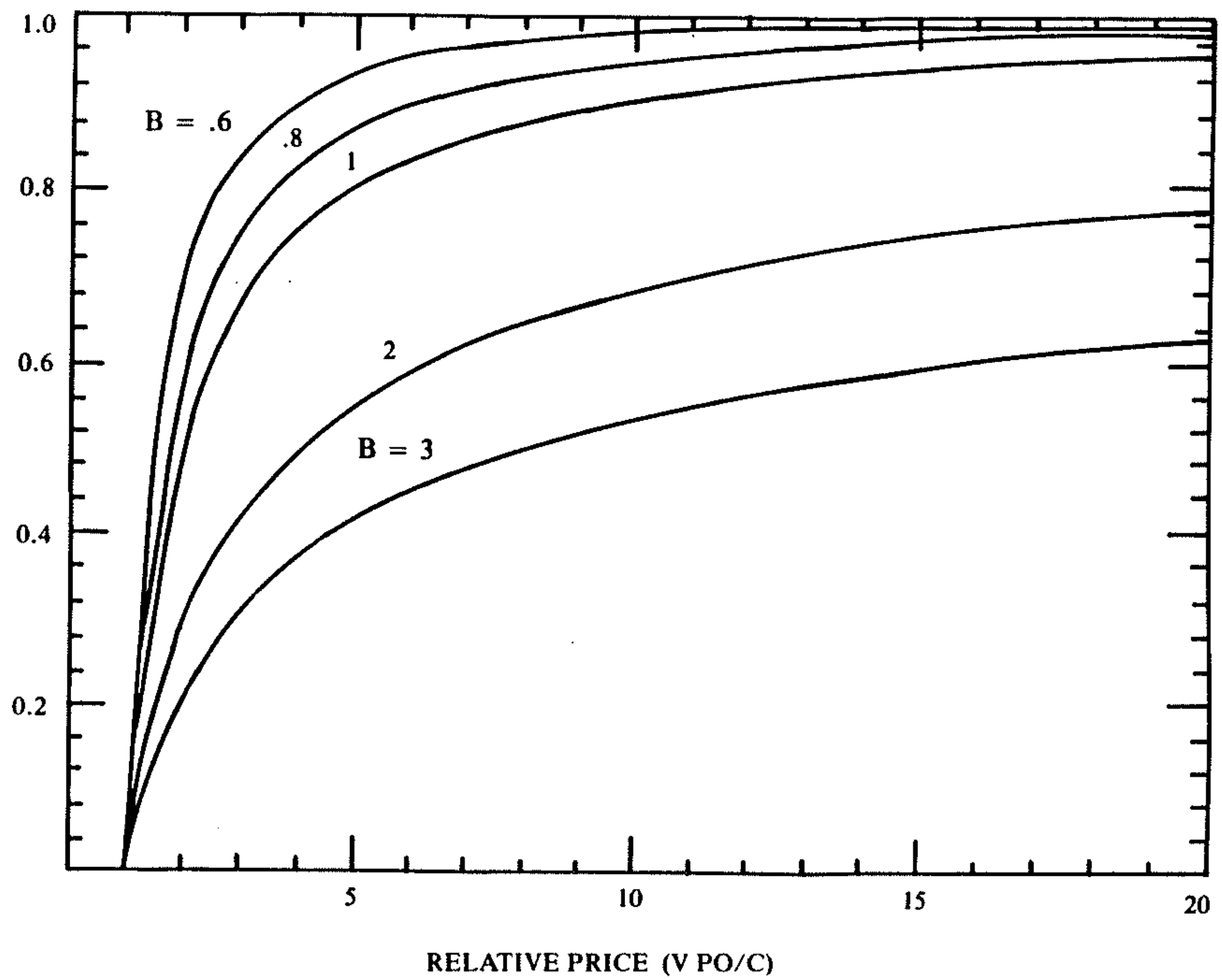

Figure 5. Projected fraction of resources ultimately discovered for various values of B.

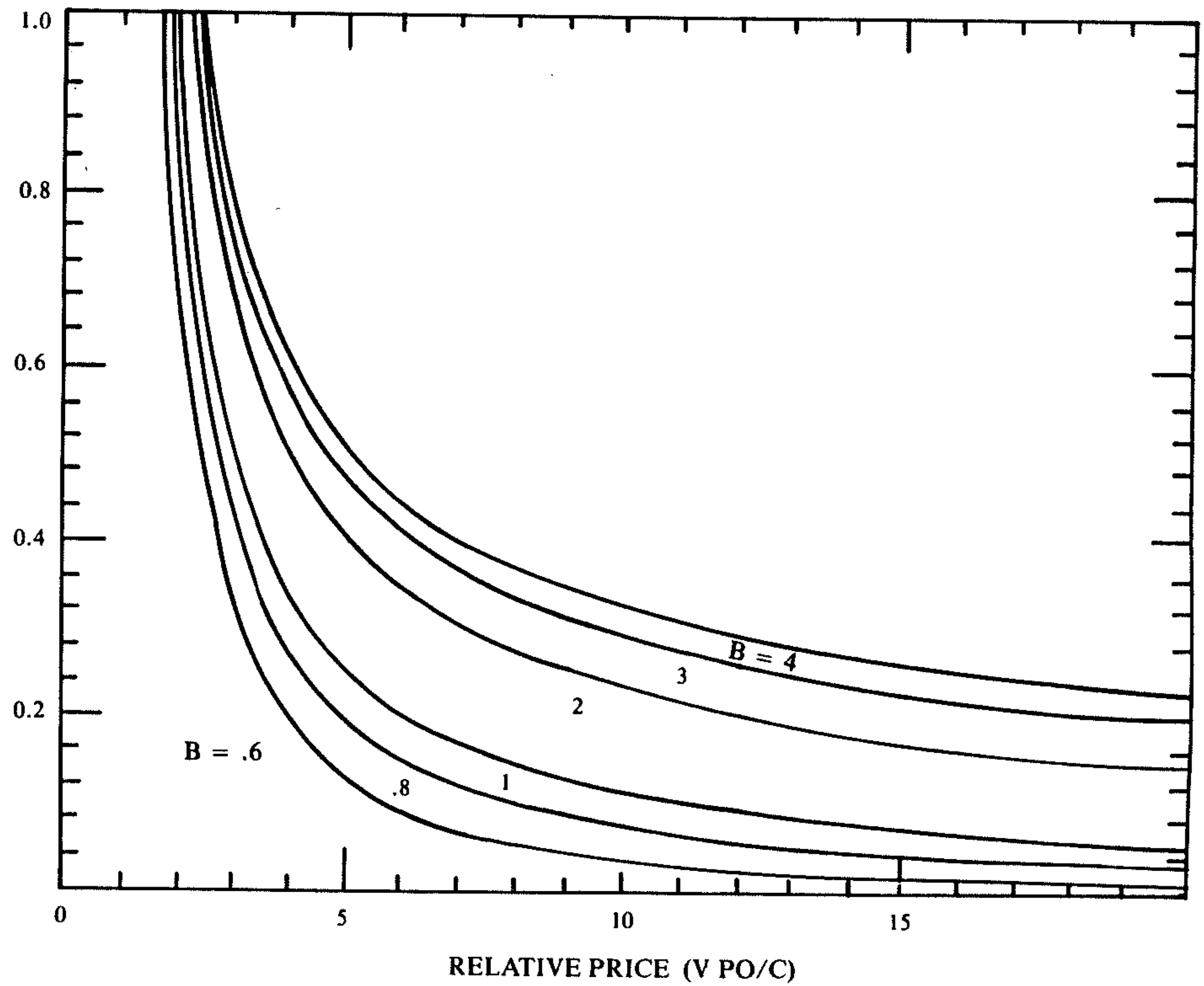

Figure 6. Elasticity of cumulative discoveries for various values of B.

Both cumulative supply and the price elasticity of supply will depend critically upon B. How well is the value of B known? Current information does not allow one to limit B between 0.6 and 3.0. Current research results certainly do not strongly justify a choice of B equal to unity. Thus, in this case, the simple model (1) allows vivid interpretation of implicit assumptions embedded within the models; (2) highlights the finding curve shape as a critical determinant of long-run oil and gas supply; and (3) suggests the potential value of research into finding curve shapes.

Conclusions

Formal model comparison activities can provide important benefits, ranging from simple identification of coding errors to greater fundamental understanding of the system being examined. The benefits obtained may vary systematically with the comparison methods used.

Methods and equations comparisons can assist in understanding the underlying theories employed and can enhance understanding of the system being studied. Yet this approach, taken alone, can be deceptive by overemphasizing methodological differences that may be insignificant quantitatively or by underemphasizing numerical or structural differences that may lead to fundamental differences in aggregate behavior.

The second and third modes of analysis, forecast and aggregate behavior comparison, are particularly useful in developing understanding of which issues are quantitatively significant. However, when taken alone, these methods can produce a cursory analysis. Numbers without analysis of the source of these numbers may provide little insight into the dependability of a model or the priorities for its improvement.

Model regeneration exercises can provide the most detailed understanding of the relationships between the parts of a model and its whole and can provide a means of testing the implications of alternative theories. However, regeneration exercises, taken alone, can waste vast amounts of computer and human time without producing basic understanding.

Modeling-the-model can focus discussion on the few most fundamental issues and can provide a method of disseminating results. However, unless the simple results are related to the aggregate behavior of the models, this approach may be little more than a mathematical exercise.

In short, the various methods each have severe limitations when taken alone. When taken together they can be highly complementary. Table 2 presents rough summary judgments about

Table 2

Model Comparison Methods and Their Benefits

	METHODS AND EQUATION COMP.	FORECAST COMP.	AGGREGATE BEHAVIOR COMP.	MODEL GENERATION	"MODEL-THE-MODEL" COMPARISON
Identification of Errors	G	G	G	E	P
Clarification of Disagreements	G	P	G	G	E
Identification & Commun. of Limits	G	G	E	G	P
Identification of Uncertainty:					
Data Inputs	P	G	G	G	G
Parameter Estimates	G	G	E	E	E
Structural	P	P	G	G	P
Guidance for Research	G	P	E	G	G
Enhancement of System Understanding	E	P	E	G	E
Guidance in Model Selection	G	G	E	G	G

P = Poor G = Good E = Excellent

the relations between comparison methods and benefits. This tentative summary is open to severe debate, criticism, and revision. The basic point is that one can expect a systematic relationship between the modes employed and the benefits obtained. For example, modeling-the-model may be useless for the identification of coding errors while model regeneration could be invaluable. Conversely, clarification and communication of disagreements can be organized through modeling-the-model whereas forecast comparisons alone will give little insight.

Three problems identified in the Introduction provided much of the motivation for this paper: (1) communication among model developers is surprisingly limited; (2) models remain as black boxes, especially to potential users; and (3) models are seldom used in a comparative mode to address policy or planning questions. These problems are a result of lack of understanding and communication. Comparative model studies can alleviate some of the difficulties associated with these problems. The current trend toward increased model comparison could lead to more effective model application of models to policy issues. This type of endeavor is worthy of continued encouragement.

Acknowledgments

I would like to gratefully acknowledge, without implicating, helpful comments and critiques from Gary Ardell, Patrick Coene, Douglas Finlay, David Froker, Douglas Logan, and John Weyant, and editorial assistance from Dorothy Sheffield and Pamela Rosas. I would especially like to thank Tricia Rigot for her extensive research and writing assistance.

References

1. Energy Modeling Forum, Energy and the Economy, EMF 1, Stanford University, Stanford: September 1977.

2. Energy Modeling Forum, Coal in Transition: 1980-2000, EMF 2, Stanford University, Stanford: July 1978.

3. Energy Modeling Forum, Electric Load Forecasting: Probing the Issues with Models, EMF 3, Stanford University, Stanford: April 1979.

4. Energy Modeling Forum, Aggregate Elasticity of Energy Demand, EMF 4, Stanford University, Stanford: August 1980.

5. Energy Modeling Forum, "U.S. Oil and Gas Supply Summary Report," Working Paper, EMF 5, Stanford University, Stanford: November 1979.

6. Hogan, W.W. and A.S. Manne, "Energy Economy Interactions: the Fable of the Elephant and the Rabbit?," in Hitch, C.J. (ed.), Modeling Energy-Economy Interactions: Five Approaches, Resources for the Future, Washington, D.C.: 1977.

7. Greenberger, M., M.A. Crenson, and B.L. Crissey. 1976. Models in the Policy Process: Public Decision Making in the Computer Era. New York, Russell Sage Foundation.

8. William Hogan. Verbal Communication.

9. Sweeney, J.W., "Energy Policies and Automobile Use of Gasoline," in Hans Landsberg, ed., Selected Studies on Energy. Background Papers for Energy: The Next Twenty Years, Ballanger Publishing Company, Cambridge, Mass., 1980.

10. Hudson, E.A. and D.W. Jorgenson, "U.S. Energy Policy and Economic Growth, 1975-2000," Bell Journal of Economics and Management Science, Vol. 5, No. 2, Autumn 1974, pp. 461-514.

11. M.I.T. Model Assessment Group, "Independent Assessment of Energy Policy Models: Two Case Studies," Report no. 78-011, Cambridge, Massachusetts: M.I.T. Energy Laboratory and Center for Computational Research in Economics and Management Science, 1978.

12. Sweeney, J.L. and J.P. Weyant, "The Energy Modeling Forum: Past, Present, and Future," Energy Policy, 1979. Also, Hogan, W.W., "The Energy Modeling Forum: A Communication Bridge," Paper presented at the Eighth Triennial IFORS Conference, Toronto, June 20, 1978. Also, Hogan, W.W., "Energy Models: Building Understanding for Better Use," Paper presented at the Second Lawrence Symposium on Systems and Decision Sciences, Berkeley, October 3-4, 1978.

13. Crissey, B.L, "A Rational Framework for the Use of Computer Simulation Models in a Policy Context," Ph.D. dissertation, Baltimore: The Johns Hopkins University, November 1975.

14. Berndt, E.R. and D.O. Wood, "Engineering and Economic Interpretation of Energy-Capital Complementarity," American Economic Review, June 1979, Vol. 69, No. 3, pp. 242-254.

15. Energy Information Administration. Annual Report to Congress, 1977.

16. Salant, S., "Imperfect Competition in the World Oil Market," Lexington Books, October 1981.

17. Berndt, E.R. and D.O. Wood, "Technology, Prices, and the Derived Demand for Energy," Review of Economics and Statistics, Vol. 58, No. 1, (February 1976) pp. 1-10.

18. Griffin, J.M. and P.R. Gregory, "An Intercountry translog Model of Energy Substitution Responses," American Economic Review, December 1976, Vol. 66, No. 5, pp 845-857.

19. Dantzig, G.B., T.J. Connolly, and S.C. Parikh, "Stanford Pilot Energy/Economic Model," Electric Power Research Institute, EPRI EA-626, May 1978.

20. Dupree, W.E., and J.A. West, "United States Energy through the Year 2000," U.S. Department of the Interior, December 1972.

21. Klein, L.R. and E. Burmeister, Econometric Model Performance, University of Pennsylvania Press, 1976.