WHY MATHEMATICAL MODELS OF PUBLIC POLICY ISSUES OFTEN DON'T WORK:

The Case of School Finance Reform in New York State

David F. Andersen
Institute for Government and Policy Studies
The Rockefeller College of Public Affairs and Policy
State University of New York at Albany

The arguments and conclusions of this study are based upon a case study involving the comparative analysis of three mathematical models used to analyze school finance reform policies in the State of New York. School finance reform is an interesting case study because proposed reforms involve the redistribution of a pool of resources of an essentially fixed size. Distribution questions in New York State (as elsewhere) are highly politicized and hence this study reduces to a case study of the use of complex models applied to highly politicized policy questions.

The central premise of this study is that complex models of social processes often fail to provide direct and useful evidence for policy makers because, of necessity, complex models are based upon five distinct classes of assumptions. At least two of these five classes of assumptions are based upon a priori or theoretical arguments rather than strict empirical arguments. Because of their inherent speculative nature (at least in part), complex models produce forecasts that are not admissable as evidence in an essentially political debate.

Hence, quantitative analysts face a fundamental dilemma. Simple models containing relatively little speculation (i.e., models that "stick to the facts") can produce evidence that is directly admissable into a political debate. However, such models may fail to take into account a full range of known (or suspected) interactions and can produce forecasts that are incorrect

when a policy situation is complex (involving interactions of several variables) and dynamic. On the other hand, models containing speculative arguments may be the best available technical forecasts of complex, dynamic situations, but evidence gleaned from such models is not directly admissable into political debate.

In the sections that follow, this line of reasoning will be developed as follows. First the broad roots of this dilemma will be discussed briefly. Second, some background to the school finance reform case and the three models being studied will be presented. Third, the three models will each be discussed in relation to the five classes of assumptions involved in a modeling study. Finally, several suggestions for improving the utility of complex models are presented.

A Dilemma Facing Public Policy Modelers

Large scale social policy reforms, whether initiated by legislation or mandated by the courts, are designed to change the behavior of individuals or institutions. Sometimes these reforms involve a carrot of fiscal incentives as in the case of the Headstart program or the New Jersey Income Maintenance Program and sometimes the reforms involve a stick in the form of legal constraints as in the case of OHSA regulation of the industrial workplace. Often reforms combine the carrot and stick by both imposing legal constraints and providing fiscal incentives as in the case of Education for the Handicapped in PL 94-142.

What these reforms have in common is that they are designed to break established patterns of behavior and to create new modes of social exchange. From an analytic point of view, such programs are difficult to deal with because, by definition, they are designed to violate most ceteris paribus assumptions, the mainstay of the analyst's trade. Hence, policy analysts charged with program design (as opposed to those charged with program evaluation) must deal with a host of puzzling technical and non-technical dilemmas. Non-technical puzzles center around the essentially political nature of the policy design process and the limitations of established organizations for implementing policy changes. Although perhaps ultimately the most important set of forces in the policy design process, these non-technical puzzles are not the locus of concern here. Instead we are interested in the types of technical dilemmas that must be solved by policy analysts working in the public sector.

The central technical dilemma centers around how to project the impact of social reforms (in fiscal or other terms) when ceteris paribus conditions, by definition, are not met. As a matter of logical necessity, analysts are forced to rely on prior assumptions of one sort or another. Analysts must assume that a prior body of theory holds (such as standard micro-economic assumptions), that individuals or institutions will react with predictable behaviors, or that some form of an equation in some sense "fits" the underlying processes at work (such as assuming that log-linear regression equations model production processes).

Clearly, an ability to postulate, test, and explain one's prior assumptions is key to producing insightful and believable policy design models. The observation that policy design results emerge solely from prior assumptions coupled with mathematics is almost true. Since one's mathematics is most often not "false," the quality of analysis appears to rest squarely on the quality of the underlying assumptions.

This necessary reliance on assumptions creates a critical tension for the analyst. One school of thought argues that analysts should "stick to the facts." Unfortunately, for policy design questions, there are few relevant "facts" concerning what would happen if a reform were to be implemented and hence attempts to stay close to one's data produce analyses that are biased toward the status quo and are often just point blank wrong (as seen with the advantage of hindsight).[1]

Another school of thought says that if someone (usually an agency head or politician) can articulate a policy conclusion, then a reasonably well-trained analyst should be able to come up with the assumptions necessary to justify that conclusion. Although undoubtedly true in part, this position can create ethical problems for conscientious analysts and reduces what should be a serious inquiry to modern day sophism.

Finally, truly conscientious analysts may find it convenient (or necessary) to rely upon rather sophisticated statistical or management scientific models, such models often being underpinned by subtle assumptions requiring years of analytic study to comprehend fully. These modeling forms with their prior and often partially camouflaged assumptions may either baffle or render suspicious non-technical policy makers.

This dilemma is exacerbated when the policy under study is one that is highly politicized. In the case chosen for study here, the allocation of state aid to localities is highly politicized because it involves the distribution of a fixed pool of funds. In such a zero sum situation, additional aid to some communities must result in relatively less state support for other communities or activities underwritten by the state. If the fixed pool of funds is expanded, legislators will have to impose an additional tax burden on their constituents.

A Case Study: School Finance Reform in New York State

As with most states in the U. S., the largest single item in the New York State budget is state aid to localities for education.[2] A multi-billion dollar item, annual negotiations around state aid to education are supported by an array of data base management and simulation systems in the State Education Department, the Division of the Budget, and both houses of the Legislature.[3] In 1978, the method that the State used to support local schools was declared unconstitutional by a lower court ruling.[4] Pending appeal of that decision, a multitude of task forces, commissions, and special interest groups have arisen to study the school finance reform question. Most of these study groups are supported by sophisticated, policy-oriented, mathematical modeling capabilities designed to project how policy changes might impact a broad range of equity and distribution questions, these questions being the basis of the original court opinion.

One of several options proposed to remedy inequities in the State's allocation of aid to localities is a cost of education index (COEI). A COEI is designed to reimburse communities for price differentials that they have to pay to purchase a standard unit of input to the educational system.[5] For example, differences in starting salaries for teachers with similar background and training might reflect in part different "prices" for services faced by districts throughout the State.

Since the early 1970's, COEIs have been widely studied within the educational finance community, but have not received wide acceptance in practice. Most researchers have proposed that COEIs be constructed using a statistical model that sorts out supply effects from demand effects. For example, differentials in teachers' starting salaries may be due to a community's desire to hire only the best teachers (a demand effect) or due to

the fact that teachers refuse to work in some areas (such as crime-ridden central cities) without a pay bonus (a supply effect).

The cost of education index question was chosen for study here because in the New York State case, three different computer-based models have been used to study the COEI question. Each of these models, is briefly described in turn.

The Tactical Simulation Model. In a recent national survey of models used to support school finance decisions, Keene identified the most commonly occurring type of model as "tactical simulation models."[6] Actually, the term "simulation" is a bit of a misnomer since these models are in fact elaborated data management systems. Large data files are maintained by the State Education Department for each local school district within the State. Drawing upon this data base, the tactical simulation model can compute aid levels on a locality-by-locality basis one year in advance. The key to developing and maintaining a tactical modeling capacity is keeping the data base current and without errors. Essentially similar versions of tactical simulation models exist within New York State in the State Education Department, the Division of Budget, and both houses of the Legislature. These tactical simulation models have long played a key part in legislative negotiations by producing accurate estimates of what will be the impacts of proposed changes on a fine-grained geographic basis. The role of models such as these has been accepted in the school finance process for quite some time.

The Econometrics Model. The econometric model chosen for study was completed by the Educational Commission of the States in Denver, Colorado under contract to the Governor's Task Force for Equity and Excellence in Education.[7] This task force was assembled to help plan the State's reaction to the court order declaring the present funding formula unconstitutional.

The specific model chosen derives COEIs for each local school district in the State by estimating an equation that sorts out supply versus demand effects in determining school teachers' salary differentials. It was one of several econometric models completed by the Educational Commission of the States and was chosen for study because it shows interesting parallels to the system dynamics model under study and because results from this model have been published and are hence in the public domain.[8]

The Systems Dynamics Model. The system dynamics model chosen for study was constructed by a team of researchers from the State University of New York at Albany. The EDFIN models were developed in consultation with staff from the New York State Division of the Budget and State Education Department. The EDFIN models were supported by a small institutional grant to SUNY-Albany as well as by a potpourri of financing involving small grants from state, and private interests. The model reported here examines the feedback implications of implementing a COEI. Results from the dynamic EDFIN models have also been made available through previous publication.[9]

Five Key Issues Involved in Managing One's Assumptions.

Based upon an analysis of three different classes of models all being used to analyze school finance reform policies in New York, five issues that span all three types of models have been identified as key in being able to manage effectively how assumptions are used and misused in formal models. These five issues are discussed in turn.

Defining the "Context" of Analysis. "Defining context" means defining a model's purpose, audience and boundary and arriving at a preliminary definition of proposed policy levers. The three models chosen for analysis in

this case all vary significantly along these critical contextual dimensions. For example, the major purpose of the tactical simulation model is to predict on a locality-by-locality basis what the implications of changes in the current formula will be one, or at most two years into the future. The system dynamics model is designed to explore the longer term (up to ten years) implications of policy changes as such changes touch off behavioral reactions in local districts. However, the system dynamics model treats aggregates of community types, not individual communities (there are approximately seven hundred local school districts in New York State). Finally, the econometric model is designed to identify and "hold constant" the effects of variables reflecting community preferences for more or less expensive educational services. Only the effects of "supply" variables beyond the control of local decision makers should be used in computing a COEI.

In a similar fashion, the three models differ in terms of their audience, and boundary of analysis. Interestingly enough for purposes of comparison however, all three models are addressing essentially the same proposed policy innovation.

Selection of Method. The econometric model under study assumes that relevant portions of the school finance reform reality may be captured by a method that assumes the following prior functional form:

$$\ln(Y) = a_o + \sum_{i=1}^{n} a_i \ln(X_i) + \sum_{j=1}^{m} a_j X_j + e \qquad (1)$$

where:

Y is the dependent variable of interest

X_i is one of n independent variables

e is a Gaussian distributed residual error term, and

a_i is an elasticity coefficient to be estimated by the following

$$\underset{\text{all } a_i}{\text{Min:}} \quad (Y-\hat{Y})^2 \qquad (2)$$

where:

$\hat{Y}$ is the predicted value of Y derived from equation (1) above.

The system dynamics model, on the other hand, made the prior assumption that the important aspects of school finance reform could be captured by a model that conformed to the functional form:

$$\underline{R} = \dot{\underline{X}} = \underline{F}(\underline{X},\underline{P}) \qquad (3)$$

where:

$\underline{R}$ is a vector of net rates of change

$\underline{X}$ is a vector of n system states

$\dot{\underline{X}}$ is the first derivative of $\underline{X}$

$\underline{P}$ is a generalized vector of parameters that may include table functions, and

$\underline{F}$ is a generalized non-linear function.

Finally, the tactical simulation model took on a rather simple mathematical form:

$$A_{i,t} = F(\underline{D}_{i,t}) \qquad (4)$$

where $A_{i,t}$ represents the aid being allocated to local school district i at time t and $\underline{D}_{i,t}$ represents a vector of data items collected on school district i for the aid period t with F being some non-linear function representing the state aid formula either in use or being proposed. The key to this tactical simulation rests in keeping the vector of data input as up-to-date as possible.

Each of these methods "shackles" the analyst in different ways. For example, the tactical model has no dynamics built into it and can not predict beyond the time period for which data (or data estimates) are available. However, within these limitations, the model is quite accurate. On the other hand, the system dynamics model has explicit dynamic capabilities, but must rely upon a series of specification and estimation assumptions concerning how local communities will react at future points in time. Verifying these assumptions is difficult.

Specification of Functional Form. Implicit in all of the three studies examined are assumptions concerning what variables to include and exclude from the study and how to specify exact functional forms. For example, the regression-centered model must decide which variables to use as measurements for supply and demand effects and whether variables should be entered solely upon the criteria of statistical significance (they were not in the case under study) and the system dynamics model had to specify detailed feedback relationships between variables chosen for study.

The details of how these specification assumptions were made, tested, and justified to policy makers is an important determinant of the reliability of model results.

Estimation of Model Parameters. Once functional forms were specified, all three models needed to obtain estimates of parameters to be used in the study. For the tactical simulation model, this process reduces to making sure that the data base is up-to-date and accurate. For the econometric model, the estimation procedure involved least squares estimation with a series of diagnostic tests being applied to test for violations of assumptions used in the estimation as well as computing summary statistics that indicate the goodness of the estimates for individual parameters and groups of parameters. Many of these statistical testing procedures are quite tricky, often evoking controversy among trained professionals, and more often baffling untrained but interested laypersons. Hence, management and reporting of how parameters were estimated can be an important determinant of perceived model utility.

Even though the system dynamics model uses less formal estimation techniques, the same set of concerns remains concerning how model parameters are derived.

Interpretation of Results. The system dynamics model relies upon an "intuitive eyeball" analysis of trajectories printed out from the model. The tactical simulation model merely sums various statistics concerning types of aid by geographic locality and presents these totals in tabular form. Laymen are often confident that they can follow and critique these interpretive processes.

However, the econometric model's parameters are not readily interpreted by the layman and nontrivial mathematical computations must be performed before the raw elasticities estimated by the regression can be turned into estimates that are meaningful to lay intuition. Several technical and conceptual difficulties exist in making these interpretations.

Quite obviously, these five issues are not distinct, but tightly interwoven. Selection of a method creates a series of methodological constraints that must be met as functional forms are specified (i.e., one can

not easily specify severe non-linearies in the independent variables of a regression equation if one wishes to use standard regression packages), and similarly how functions are specified influences whether or not data will be available to estimate parameters and how interpretable final results will be.

In the sections below, each of the three models is discussed in more detail with special attention paid to how the models differ along the five dimensions sketched above.

Discussion of the Tactical Simulation Model

Over the past five to ten years, tactical simulation models have gained wide acceptance as tools for forecasting distributional impacts of proposed policy changes in the school finance area. They are extensively used both within the executive and legislative branches of government. In fact, each of the major parties to the school finance policy debate maintains a tactical simulation capability of some sort. The Governor's annual school finance proposal is supported by such a model within the Division of Budget as is the Board of Regents' proposal supported by a model within the State Education Department. Both houses of the Legislature have access to tactical simulation capabilities during the period when the school aid budget is being negotiated.

The wide spread acceptance of tactical simulation models stems to a large degree from their technical properties. The methods being employed are easy to understand and do not require the inference of patterns using statistical models. Results from two or more such simulations can be reconciled since literally no assumptions are evoked by these models. An officially recognized data base is made available on a routine basis to all parties by the State Education Department.[10]

Most important, since the results of the tactical models are based upon a data base that is frequently updated through official forms filled out by localities for the State, the results emanating from the tactical models are subject to audit by the State. That is, the models are strictly data based, do not rest upon assumptions, and leave an audit trail that, in principle, could be used to verify the veracity of the figures reported from the model. (However, in practice not all of the data being used has been closely audited. At best most of the data has been subjected to a "desk audit" to check for internal consistency.)

A further analysis of the tactical models under the five categories identified above demonstrates further some of the strengths and weaknesses of these models as policy analytic tools.

<u>Context of Analysis</u>. In terms of audience, each of the major parties to school finance negotiations is an audience for one of the several tactical simulation models. In terms of the COEI question, the purpose of the models is to forecast changes in next year's aid distribution <u>if</u> a COEI were to be implemented. Such a model would use COEI figures generated by a statistical model (such as the one discussed below) or some other proxy measure (such as an index based on wage rates in the State's several labor market areas) and apply these indices to the current formula. Currently available data would be used to forecast the impact of a COEI on a locality-by-locality basis.

<u>Assumed Functional Form</u>. The functional form employed by the tactical simulation models is given in equation (4) above. Typically, these models do not attempt to forecast the vector of data items $\underline{D}_{i,t}$, into future time periods. Extensive efforts are not made to infer statistical relationships

between items in the data vector for purposes of creating a formula. Instead the data vector is assumed to be accurate and forecasts of impact evoke an implicit <u>ceteris</u> <u>paribus</u> assumption since extensive efforts are not made to forecast shifts in the basic data vector being used to estimate future aid levels.

<u>Model Specification</u>. Model specification reduces to specifying exactly the function, $\underline{F}$, that is used to compute final aid levels, $A_{i,t}$, from the data vector. This specification is most usually uncontroversial since it involves specifying the existing or proposed aid formula in equation form. No <u>a priori</u> or theoretical assumptions need to be evoked in this process.

<u>Model Estimation</u>. The tactical models do not involve technical questions of model estimation. Instead, analysts must assure that the data vector is current and accurate. Correctly put, these are measurement and data management problems, not estimation problems. Once again, analysts using tactical models are not required to make any theoretical assumptions or arguments. Disagreements over model results reduce to empirical questions that, in principle, can be resolved by an audit of the data in question.

<u>Model Interpretation</u>. At a first blush, the interpretation of results emerging from tactical simulations are neither controversial nor based upon <u>a priori</u> or theoretical assumptions. Simply put, tactical models produce highly accurate forecasts of next year's aid levels. Questions of interpretation, logically reduce to questions of data accuracy. Again, such questions are empirical, not speculative in nature.

However, a criticism of the interpretability of tactical models can be raised along the following line of reasoning. If policies under consideration will reinforce patterns of inequity implicit in the present data or produce effects that reverse themselves over time, then tactical models by focusing on first order predictions of next year's aid levels fail to consider important second order effects and hence unwittingly may lead to poor policy choices (poor in the sense that next year's distribution decisions may create inequities or unintended effects in future years).

The retort to this criticism is that policy choices made this year will be reviewed by the Legislature next year. Since the budget is reviewed annually, defects in distribution policies can be remedied before these deleterious second order effects take place. In a word, with a one year budget cycle, one year forecasts are sufficient.

However, the retort to this retort points out that next year the tools being used to analyze distribution policy will be the same and the constellation of analysts and decision makers essentially the same as this year. Hence the same pressures that make it difficult to look at inter-variable correlations and long term effects this year will be present next year. This view holds that by relying on one year forecasts with no explicit attention paid to statistical patterns within a year's data or possible long trends, the entire policy process becomes locked into a myopia dominated by a vision of only one year. According to this view, the policy process will drift from one year to the next, always intending to look at multi-year effects next year. Meanwhile, the unanalyzed second order effects (such as shifts toward greater inequity in distribution patterns) continue to accumulate in a compounding fashion.

In sum, because of their technical properties, tactical simulation models produce highly reliable estimates that are directly admissable as evidence in the annual budget debate, an essentially political process. However, the same restrictions on the analysis that make such models politically acceptable may have the longer run effect of locking the policy process into a myopia

dominated by one year forecasts that ignore important interactions between variables or effects over time.

Discussion of the Econometrics Model

The econometrics model selected for study here is an analysis of the COEI in New York State as published by Wendling.[11] The published work is based directly upon a report to the Governor's Task Force on Equity and Excellence in Education. This work, completed by the Educational Commission of the States, statistically estimated a COEI for adoption in the State of New York.

As technical staff for the Governor's task force in New York and for similar task forces in other states, the Educational Commission for the States has long been an advocate of COEIs as a mechanism for improving educational finance policies. The results computed by Wendling for ECS and examined here represent a fairly accepted "state of the art" with respect to the statistical computation of COEIs.

However, even though the Wendling study represents the current state of the art in COEI computation and was able to pass a peer review for publication in an academic journal, the overall technology is conceptually and empirically troubling and perhaps not directly useful to decision makers who must make allocation decisions. The limited utility of the statistical studies of COEIs can be traced to how the five levels of concerns sketched above have been handled within these COEI studies.

Context of Analysis. Wendling completed his study under direct commission from ECS with the Governor's task force serving as his immediate client. His work was funded by a direct state appropriation supplemented by support from private sources interested in questions of school finance reform (most notably the Ford Foundation). This study was another in a continuing series of studies by ECS centering on COEIs. The purpose of the study was to statistically estimate a COEI that counted for differences in "supply" factors while "holding constant" differences in a community's "tastes" for differential levels of educational expenditures.

Overall, the client relationships between Wendling, ECS and the State are complex. With minor exceptions, the Governor's task force was staffed by experts external to the State. This staffing pattern was designed to insure that the task force's final recommendations would not be dominated by established interests within the State. New ideas and fresh blood were to be infused into the policy process. However, this same maneuveur had the net effect of insuring low participation by key actors in the policy process including the Legislature, the State Education Department, and the Division of the Budget.

One could argue that the context for the study just described provided an ideal opportunity to launch a truly impartial study. On the other hand, one could argue that the task force would be doomed from its conception to be ineffectual because it lacked key linkages to the political environment within the State. In sharp contrast to the tactical simulation models just described, this econometric model was owned by a relatively impartial, academic analysis team with weak political connections. The tactical simulation models were developed and maintained by analytic units that were active participants in the political processes surrounding the annual budgetary debates.

Assumed Functional Form. The functional form of the model employed by Wendling is given in equations (1) and (2) above. The dependent variable was the salaries of a sample of teachers and administrators from across the State

and the independent variables were approximately fifty measures of both supply and taste variables that might explain the variance in teachers' and administrators' salaries across the State. Taste related variables were classified as personal characteristics of administrators and teachers (assuming that different communities preferred personnel with differing characteristics), professional environment (measured as teachers or administrators per 1000 students), and a community's fiscal capacity. Supply variables over which the communities were assumed to have no control were grouped into students' characteristics, school district characteristics, and regional characteristics.

The functional form employed by Wendling made several important statements about the school finance system. First, salaries derive from a production function type formulation and the inputs to this function must be determined by measurable attributes of individuals, communities, and regions. Second, the exact production function is a combination log-linear and log-log one with no lagged or non-linear effects. Finally, this functional form obeys certain stochastic assumptions (such as normal distribution of the residual error term and no measurement error in the independent variables) that allow the entire equation to be estimated by least squares techniques.

Indeed, this assumed functional form is restrictive in nature and empirical studies exist demonstrating that even when good statistical "fits" can be obtained from such forms, they may be quite poor representations of real world processes.[12] In actuality, the justification for using such a functional form rests on prior arguments that cite that such studies have been successfully completed in the past. Also, such a functional form has the convenient property of being able to be estimated easily by standard statistical techniques.

In the case of the Wendling study, this functional form was evoked implicitly and a priori with virtually no discussion of why such a functional form might be sensible.

Model Specification. For the Wendling study, the exact specification of the equation to estimate teacher and administrator salaries proved to be a critical issue that was settled by resorting to theoretical, common sensical, and other non-empirical forms of argument. The key issue reduced to what variables should be considered "taste" variables and hence held constant, and what variables should be considered supply variables and hence have an impact on the COEI. Specifically, the logic underlying the computation of the COEI ran as follows:

The basic equation estimated was:

$$\ln(SAL) = a + AT + BS + e \qquad (5)$$

where SAL is the salary being estimated, A is a vector of price coefficients associated with the taste vector T; B is a vector of price coefficients associated with the supply vector S; and e is a normally distributed residual term. The index for a given school district was computed by holding constant taste variables at their mean value as follows:

$$\ln(SAL_i) - \ln(\overline{SAL}) = (a + A\overline{T} + BS) - (a + A\overline{T} + B\overline{S}) \qquad (6)$$

where the bars indicate mean values. This equation reduces to:

$$\ln(SAL_i/\overline{SAL}) = B(S - \overline{S}) \qquad (7)$$

exponentiating each side of this equation:

$$SAL/\overline{SAL} = INDEX_i = e^{B(S - \overline{S})} \qquad (8)$$

An examination of the mathematics used to compute the COEI indicates that the key step in the entire process involves determining which variables to include in the taste vector, T, and which variables to include in the supply vector, S. In a review of COEI studies completed previous to the Wendling study, Johnson discovered that several key variables such as measures of personal characteristics of teachers have been classified as taste variables in one study and supply variables in another.[13] That is, the theoretical justification for placing variables into one category versus the other are ambiguous enough so that two researchers could make exactly the opposite decision. Furthermore, the magnitude of the price coefficients estimated by Wendling indicates that if several key variables were reclassified from the supply to the taste side of the equation (or vice versa), the actual indices computed would shift dramatically.

In other words, the absolute values of the actual indices computed by Wendling are sensitive to specification assumptions. Furthermore, the justification for such specifications rests upon rather ambiguous theoretical and common sensical arguments. As opposed to the tactical simulation model examined previously, the Wendling model rests critically upon non-empirical assumptions involving selection of the model's basic functional form and the specification of the model.

Model Estimation. The two vectors of price coefficients in the Wendling model were estimated using a standard least squares approach. The usual questions involving significance of estimates, robustness of estimates, the existence of colinearity among the independent variables and other violations of the implicit assumptions of the basic regression were involved in the estimation of the Wendling model's parameters. However, all of these questions are statistical in nature and require skillful treatment by the modeler rather than the invoking of additional assumptions without strict empirical grounding.

Model Interpretation. As with the tactical simulation models, the interpretation of Wendling's results does not appear to be a complex task on the surface. Wendling presents a cost of education index for every local school district in the State. These indices are proposed to be entered as arguments in the State's formula for distributing aid to localities. The only technical difficulty in interpreting his results centers around whether or not the basic assumptions underlying the computation of the indices are reasonable. These questions have been discussed previously.

In sum, the Wendling study begins with a common sensical argument that some school districts must pay more for similar educational inputs than others do and hence should be compensated more by the state aid formula for these additional costs. He employs a methodology that is commonly accepted within the academic community (the use of a production function type regression model) and develops results that have reasonable face validity (his indices vary from approximately .88 to 1.17 with a standard deviation of .083. These results are consistent with previously identified differences in the cost of living between localities within New York State and with other COEI studies. Although his a priori and theoretical arguments may be weak in some spots, they do not appear at any point to be clearly and definitively wrong. The study could be improved by increased sensitivity analysis as well as some additional field work to determine what are the important taste versus supply variables defining teachers' preferences.

However, even if these basic modifications were to be made, it seems clear that the analyst would ultimately have to base his conclusions upon some assumptions that are not strictly empirically founded. Given that key

decision makers within the State (that is state legislators) are locked into a zero sum distribution debate, it would appear unlikely that Wendling (or any other analyst for that matter) would be able to obtain concensus from all of the key decision makers because to agree to a set of assumptions that run counter to one's own interest is to surrender needlessly an additional chip in the political bargaining process. Even if they are constructed according to the best technical standards, models such as those employed by Wendling ultimately will not be admissable as direct evidence into a policy debate where key decision makers have conflicting interests concerning the ultimately desirable policy outcomes.

Discussion of the System Dynamics Model

A third model, a system dynamics model, has been constructed to examine the policy implications of COEIs. Two generations of system dynamics models were constructed (the EDFIN1 and EDFIN2 models) with the second generation being an elaboration of the first.[14] The EDFIN models group New York State's 700 local school districts into from four to eight aggregate representations of local districts. Each such local sector dynamically sets its annual budget and tax rates in response to shifting environmental factors (e.g. pupil enrollment shifts or inflation) and changes in state policy. The broad purpose of the models was to search for feedback effects in the school finance system that might produce unintended or adverse consequences from proposed policy changes.

As with the econometrics model discussed above, the system dynamics model of necessity evokes a series of analytic assumptions in order to examine school finance policies. Because such assumptions are theoretical or a priori in nature, the system dynamics model is also vulnerable to attack as being grounded at least in part in speculative versus empirical argument. However, as discussed below, the exact strengths and weaknesses of the system dynamics model differ considerably from those of the econometric model.

Context of Analysis. Work on the EDFIN1 model began in the summer of 1979 as a basic research project by a team of researchers at the Graduate School of Public Affairs, the State University of New York at Albany. The initial model development was guided by a belief that important feedback effects were operating in the school finance system and that unless such effects were explicitly analyzed, attempts at finance reform would either be ineffective or produce significant unintended and negative consequences. Initial model development was funded by two small institutional grants to SUNY at Albany, one by HEW and one by a private foundation.

Preliminary results from EDFIN1 indicated that potentially important feedback effects were in fact operating. Analyses focusing on policies designed to equalize per pupil expenditures found that policies producing dramatic progress toward equity in the short run (1 to 3 years) were defeated by self-adjusting pressures within the finance system in the longer run (5 to 7 years).[15] Preliminary results suggested that inflation contributed considerably to inequity in per pupil expenditures[16] and that central city compliance with special education mandates led to an unanticipated tax subsidy for rural and suburban taxpayers.[17]

In the fall of 1980, these preliminary results were presented to analysts within the State Education Department and the Division of Budget. Working in consultation with senior analysts in each of these units, the structure of an improved model was sketched broadly. The development of EDFIN2 was supported by the State Education Department over the summer of 1981.

Originally designed as a tool to analyze the policy design process for academic audiences, the EDFIN models have been drawn somewhat closer to the policy process in New York by the active interest and support from several senior analysts. However, the models are still in a preliminary stage, with preliminary results that are not directly coupled to the political policy process. As discussed below, the technical properties of the EDFIN models make them interesting tools for analysis, but as with the econometrics model their reliance on a priori and theoretical arguments inhibits them from producing forecasts that are directly admissable into the policy debate.

Assumed Functional Form. The basic functional form assumed by the system dynamics methodology is given in equation (3) above. This form assumes that the relevant reality under study can and should be seen as a system of n state variables whose rates of change are intercoupled by non-linear feedback effects. As with the functional form employed by the econometric model, this form must be justified on a priori grounds. The a priori arguments most commonly invoked by system dynamics analysts are collected under the general title of cybernetics, a body of literature suggesting that the important aspects of social systems may be represented by feedback theory.[18]

However, in contrast to the econometric functional form, the form assumed within equation (3) is a fairly flexible and more general form. This can be easily demonstrated by noting that the econometric functional form in equation (1) is a special case of equation (3) where there is one state variable, with no rates of change specified, and the overall form is linear or log-linear. In general, the prior functional forms assumed by system dynamics models are among the more flexible ones employed in the quantitative modeling of social processes.

Model Specification. The flexible functional form discussed above is both a blessing and a curse to system dynamics analysts. Such a flexible functional form demands extensive specification of causal feedback structure with such specification being grounded in a mixture of empirical, theoretical, and a priori arguments. Several of the major specification assumptions and their origins are sketched below.

A key specification within the EDFIN models centers on how to aggregate the 700 local school districts within New York into a small number of local sectors with similar fiscal and demographic characteristics. For the EDFIN1 model, this grouping was done using a rough grouping into four based upon a common sensical analysis of important cleavages within the State. The EDFIN2 model employed a more elaborate procedure to divide the State into eight sectors. Data on approximately fifty variables relating to local districts' fiscal and demographic characteristics were collected. These fifty variables were reduced to three factors (measuring relative size, wealth, and ruralness) using a standard factor analysis routine. A cluster analysis routine was then used to identify groups of districts with similar characteristics along the three factors identified above. Both the EDFIN1 and EDFIN2 models assume that the grouping of school districts into aggregate sectors is a sensible thing to do (that there are few behavioral differences within a given sector). However, the empirical basis for this assumption for the EDFIN2 model is much more carefully worked out.

Another example of the types of specification assumptions evoked by the EDFIN models centers on how localities and the State set their annual budgets. EDFIN1 assumed that desired expenditure levels were determined exogenously to the model and that the purpose of the model was to examine how the state aid allocation sector responded given various levels of desired expenditures.

However, the EDFIN2 model assumes that information from the availability of state and local revenues feeds back to influence the budget setting process. These assumptions, clearly more realistic in nature, were needed to get the EDFIN2 model to track available time series data in a reasonable fashion.

A final example treats how the EDFIN models simulate the response of the Legislature to shortfalls and windfalls in the amounts of money available in the state budget to provide aid to local districts. EDFIN1 assumed that if a windfall or shortfall existed, then the Legislature would distribute this differential strictly in proportion to how the rest of the aid was being distributed between the sectors (i.e., if sector I was receiving 20% of the total block of state aid, then sector I would bear 20% of any shortfall or windfall in a given year). Discussions with analysts within the State Education Department indicated that this assumption is a good first approximation, but that in acutality the Legislature allocated shortfalls and windfalls by manipulating several of the values in the State aid formula (such as the maximum or ceiling aid that the state was willing to pay per pupil). These observations were incorporated into the final version of the EDFIN2 model.

As the above three examples suggest, the specification of the EDFIN models is not an entirely straight-forward process. As with all system dynamics models, this specification process involves part empirical, part theoretical, and part common sensical (relying upon expert judgment) arguments. It also seems apparent that the specification of EDFIN2 is more thorough and carefully done than the specification of EDFIN1. However, as with the econometric model, insofar as the model's specification is not entirely uncontroversial, the model's structure and hence conclusions are vulnerable to disqualification in a political environment.

Model Estimation. As with structural specification, the estimation of parameters within the EDFIN models is based upon a mixture of empirical, theoretical, and a priori arguments. In fact, the system dynamics literature recognizes that questions of parameter estimation can not be divorced from questions of structural specification.[19]

For example, initial conditions can be empirically derived from data made available by the State Education Department. Similarly, parameters used in the formula to distribute aid are published by the State and are uncontroversial. However, estimates of how quickly the Legislature will respond to shifting patterns of local expenditures are not readily available and such estimates must be pieced together using arguments that rest upon expert judgment and common sense. For example, common sense rules out the possibility that legislatures react to local changes more quickly than one year (the length of one budget cycle) nor more slowly than 10 years since this time rivals or exceeds the average legislator's tenure in office.

As opposed to the econometrics model discussed above, the estimation of parameter and table function values within a system dynamics model is not a "scientific" question that can readily be settled by reference to single tests such as the statistical significance of proposed parameter estimates. In fact, Forrester and Senge have proposed that the estimation of model parameters is part of a larger process of validating and building confidence in a complex model.[20] They suggest that this overall confidence building process involves multiple tests of model structure, behavior, or both structure and behavior. Such tests involve close collaboration with the ultimate model users as key decision makers develop a concensus of confidence in the model's structure and behavior.

In the case reported here, key decision makers, due to the very nature of the allocation decision being made, are by definition at odds and hence the concensus building, model validation tests proposed by Forrester and Senge and others will not work.[21]

Model Interpretation. Output from the system dynamics model involves scores of variables that vary against time and must be digested and interpreted by key decision makers. Recent research suggests that key decision makers may not agree on the interpretation of model results first because the task of integrating such a mass of information may be psychologically taxing[22] and second because individual decision makers may place differential weights on various aspects of the model's output and hence arrive at differential interpretations of what the model actually means.[23] However, neither of these problems is reason for calling into doubt the basic assumptions that underpin model output.

In sum, as with the econometric model (and in opposition to the tactical model), the system dynamics model is based upon a blend of empirical, theoretical, and a priori evidence. The standard wisdom in the system dynamics field defines model validity in terms of concensus building tests involving key decision makers. However, in a highly politicized environment, arriving at concensus, by definition, means that one or more of the key actors must surrender all or part of his or her political position. Hence, concensus surrounding model assumptions is not likely.

Even though the relative strengths and weaknesses of the two modeling approaches are different, both the system dynamics and econometric approaches share a common fate of producing results that will not be directly admissable into a policy debate involving zero-sum questions.

Implications and Discussion

Taken together, the three models presented in this case study paint a rather gloomy picture with respect to the use of complex models to support political decisions of a zero-sum nature. The results of the study can be summarized in two rules of thumb and one immediate corollary to these rules.

The rules of thumb are: (1) Statistical models that explore interactions between more than two independent variables will not produce results that are admissable into zero-sum political decisions, and (2) dynamic models that project impacts more than two budget cycles into the future will not produce results that are admissable into zero-sum political debates.

As detailed in the case study above, the reasons for both of these rules are fundamentally the same. Complex statistical or dynamic models of social phenomena of necessity rely in part on a priori or theoretical arguments. Analysts must stray from rock solid empirical foundations if they are to design new policy options for future implementation (as opposed to evaluate policies implemented in the past). In a zero-sum distribution decision, if some key decision maker (i.e., legislator) experiences a relative gain, some other decision maker must experience a loss. Hence concensus around a priori or theoretical assumptions will not occur since some coalition of decision makers will always lose from such technical concensus.

An immediate corollary of these two rules of thumb is that the introduction of a full range of quantitative modeling capabilities to zero-sum policy choices will tend to produce a policy process that is more myopic and takes fewer factors into account than policy processes functioning without the support of mathematical models.

Complex models that take into account interactions between several variables or that project multi-year impacts will be inadmissable into the debate for the reasons noted above. Hence, only models that are solidly based empirically (such as the tactical model discussed above) will be admitted into the policy debate. By focusing on highly accurate forecasts of next year's distribution, these models will draw attention each year to the short run consequences of distribution decisions.

In the absence of accurate short term forecasts, legislators could previously attempt to implicitly balance short term with long term effects through purely verbal debate. In a word, the introduction of short run forecasts has lifted a "viel of ignorance" that had previously allowed legislators to act in a more statesman-like fashion in ignorance of the exact gains and losses that would accrue to their home constituencies from their actions.

A hopeful suggestion might be that the pessimistic conclusions of this case study are not necessarily always true, but merely artifacts observed in and unique to this one case. The unique nature of these conclusions could be due to two reasons. First, perhaps the tactical models were more admissable not because of their technical properties but because they were each used by politically "connected" teams of analysts. Second, perhaps the problem with the system dynamics and econometric models was not due to intrinsic properties of these classes of models in general, but rather to the details of the specific models built in the first place--perhaps they were not technically sound models. Each of these two suggestions is briefly discussed in turn.

If a model's utility is in fact tied to the "connectedness" of the analyst team that builds it, then why haven't such politically connected analysts adopted other modeling technologies? Econometric and system dynamics modeling techniques have been in existence for 20 years or more. Surely if such models were politically viable, they would have received some acceptance during this time span among analytic staff. The possible counter-argument is that political staff members are analytically unsophisticated or ignorant of the potential of such models. This does not seem likely in a state where analysts possess state of the art graduate training and do rapidly acquire technological innovations when appropriate in other fields.

The second suggestion that the econometric and system dynamics models were not technically the best possible efforts is undoubtedly in part true. However, even given a large budget and unlimited time for model development, it seems unlikely that a complex model could, in principle, be constructed without resort to some a priori or theoretical arguments. Perhaps the impact of unlimited time and budget would be to camouflague more cleverly critical assumptions behind layers of analysis thereby increasing the time and effort needed to critique the resultant models. The ultimate solution to the question of whether or not technically "perfect" models are or are not grounded in a priori and theoretical arguments could be reached if someone could point to at least one complex model of a distribution problem that is without controversy. No such model comes to mind easily.

In sum, results of this case study suggest that the most important issues in an analysis of zero-sum distribution issues are not technical ones. Instead, basic issues relating to selection of method and model conceptualization and specification are the most important issues in defining a model's ultimate utility in a political environment.

Modelers need to carefully manage their client relationships. Perhaps models constructed with no direct political client and making no pretenses at producing forecasts that are directly relevant to the policy process may

ultimately be the most valuable ones. Such models could allow key decision makers to discuss broader issues without the necessity of arriving at concensus publicly over key assumptions. Modelers must take care to express the results of their analyses in ordinary language without resort to jargon, higher mathematics or evoking the computer as a "black box" that gives solutions.

In the final analysis, the public policy process is a political dialogue that is carried out in ordinary language among parties with both private and public interests. Policy relevant models must strive to enrich that dialogue rather than displace it with assumptions and mathematics that obscure rather than inform the issue.

NOTES

[1]For an example of how forecasts based on the status quo can miss the mark by a factor of 10 or more, see David F. Andersen, "Using Feedback Simulation to Test Educational Finance Policies," Policy Sciences 12(October 1980) 315-331.

[2]Education Study Unit. New York State Division of the Budget. New York State's System of School Finance--with Examples from the 1975-76 School Year. Albany, New York, December 1976.

[3]For examples of such a simulation, see New York State Executive Department. Division of the Budget. SIMULBUD-L User's Instruction Manual. Albany, New York. Revised October 1978.

[4]Board of Education, Levittown Union Free District v. Nyquist. 94 Misc. 2d 466; 407 NYS 2d 606. Nassau County Supreme Court.

[5]For a review of some of the issues involved in COEIs, see Jay G. Chambers. "The Development of a Cost-of-Education Index: Some Empirical Estimates and Policy Issues." Journal of Education Finance 5 (Winter 1980)

[6]Peter G.W. Keen and David G. Clark. Simulation for School Finance, A Survey and Assessment. Denver, Colorado: Educational Commission of the States, 1979.

[7]Wayne Wendling. The Cost-of-Education Index: Measurement of Price Differences of Education Personnel Among New York State School Districts. Papers in Education Finance No. 26. Denver, Colorado: Education Finance Center, Education Commission of the States, April 1980.

[8]Wayne Wendling. "The Cost-of-Education Index: Measurement of Price Differences of Education Personnel Among New York State School Districts." Journal of Education Finance (Spring 1981) 485-504.

[9]Fiona Chen, David F. Andersen, and Tanette Nguyen. "The Impacts of Implementing a Cost-of-Education Index on the Equalization of Education Expenditures per Pupil." Proceedings of the 1981 International Conference on Cybernetics and Society. Atlanta, Georgia: IEEE Systems, Man and Cybernetics Society, October 1981.

[10]Educational Finance Unit. State Education Department. "File Format for Standard State Aid Data Tape." Internal memorandum. Albany, New York, February 18, 1981.

[11]Wendling (1981).

[12]Daniel F. Luecke and Noel F. McGinn. "Regression Analyses and Education Production Functions. Can They Be Trusted?" Harvard Education Review 45 (August 1975) 325-350.

[13]Gary P. Johnson. "Cost-of-Education Indices: State of the Art and Implications for School Finance Reform." Paper presented at the annual meeting of the American Educational Research Association, San Francisco, California, April 8-12, 1979.

[14]The most recent documentation of the EDFIN2 model is contained in a working memo from David F. Andersen to Robert Lamitie, New York State Education Department: "Progress Report on the EDFIN2 Modeling Project," September 30, 1981.

[15]Fiona Chen, David F. Andersen, and Tanette Nguyen. "A Preliminary System Dynamics Model of the Allocation of State Aid to Education." Dynamica (in press, Spring 1982).

[16]Tanette Nguyen and Fiona Chen. "Analysis of Impacts of Inflation on Equality of Educational Expenditures in N.Y.S. using the EDFIN1 Simulation Model." Memorandum to Robert Lamitie, New York State Education Department, July 1, 1981.

[17]Tanette Nguyen, David F. Andersen, and Fiona Chen. "The Dynamics of State Aid to Education: Interactions Between Special Education, Regular Education, and Non-Schooling Expenditures." Proceedings of the 1980 International Conference on Cybernetics and Society. Atlanta, Georgia: IEEE Systems, Man and Cybernetics Society, October 1981.

[18]For examples of these a priori justifications see: Norbert Weiner. Cybernetics; Karl W. Deutsch. The Nerves of Government: Models of Political Communication and Control. New York: The Free Press, 1966; or Jay W. Forrester. Industrial Dynamics and Principles of Systems.

[19]Alan K. Graham. "Parameter Estimation in System Dynamics Modeling." TIMS Studies in the Management Sciences 14 (1980) 125-142.

[20]Jay W. Forrester and Peter M. Senge. "Tests for Building Confidence in System Dynamics Models." TIMS Studies in the Management Sciences 14 (1980) 209-228.

[21]A more general discussion of the roles of concensus and refutation in building confidence in a model is given in James A. Bell and Peter M. Senge. "Methods for Enhancing Refutability in System Dynamics Modeling." TIMS Studies in the Management Sciences 14 (1980) 61-73.

[22]John W. Rohrbaugh and David F. Andersen. "Can Managers Make Sense Out of Management Science Models?" Unpublished manuscript, Graduate School of Public Affairs, SUNY-Albany.

[23]Peter C. Gardiner and Andrew Ford. "Which Policy Run is Best and Who Says So?" TIMS Studies in the Management Sciences 14 (1980) 241-257.