CONF-790852--1

THE QUANTIFICATION OF UNCERTAINTIES IN THE PARAMETERS OF A LONG-TERM ENERGY MODEL*

R. W. Peelle, D. J. Downing, and D. J. Bjornstad
Oak Ridge National Laboratory
Oak Ridge, Tennessee 37830

MASTER

ABSTRACT

Even if the form of an energy-economy model's equations can be assumed to specify correctly our technological processes and the relevant behaviors of our society over the necessary time range, there is uncertainty in model results induced by our imperfect knowledge of the numerical values of the model's parameters and input data. Some of this uncertainty is typically "covered" by provision of alternative scenarios with assumptions, but up to now modelers have rarely dealt in detail with the inherent uncertainty of input data. However, when model output or "response" can be represented by a first-order Taylor expansion in the input data about the nominal solution point, knowledge of the variance-covariance (uncertainty) matrix of the input data is sufficient to determine the uncertainty in the computed response induced by the input uncertainties. Some guidelines are given for the evaluation of the required input uncertainty matrices. Illustrative examples are given from our beginning efforts to develop an uncertainty matrix for the important parameters of the Long-term Energy Analysis Package used within the Energy Information Administration of the Department of Energy.

*Research sponsored by the Energy Information Administration, U. S. Department of Energy under contract W-7405-eng-26 with the Union Carbide Corporation.

INTRODUCTION

One way to increase the usefulness of energy economy models for policy analysis by decision makers will be to include uncertainty information with the model output. This goal is generally approached now by showing alternative scenarios for a range of values of a few key parameters and by comparing the projections of various model formulations for nearly-common input assumptions. This paper discusses the need to deal more completely with quantitative expression of the uncertainty of model results and sketches a path which may be followed to meet a portion of this need.

Particularly for scenarios which reach a generation into the future, serious uncertainty stands everywhere. Modeling current economic behavior is difficult, and one cannot be secure that our society will closely follow similar patterns under the altered stimuli of the future. In a rough way one may separate model output uncertainties into two categories: 1) the result of uncertainties in the model specification and 2) the propagated effect of uncertainties in the numerical values input by the model user. The first includes uncertainties from what is excluded from the model, and the second from what is included. This paper is concerned only with the latter category, the output uncertainties generated by uncertainties at the inputs, The assumption is made that the model specification is sufficiently valid that the indicated separation of classes of output uncertainties can be made. In this paper the terms "input data" and "model parameters" are used interchangably, and no distinction is made between endogenous and exogenous parameters even though in a completed work the two classes might be treated separately.

Beyond the considerable effort involved, some may feel that propagation of input uncertainties is unnecessary or imprudent. To claim that uncertainties in model parameters representing the future cannot be assessed would be to claim too much, for in this case the nominal values of the parameters would be equally inaccessible. Another thought has been that a model just plays out alternative scenarios based on hypothesized input parameters, so it is not necessary to consider uncertainties. This argument is weak if the model contains so many parameters that the user can hardly be cognizant of all the assumptions made. Moreover, the distinction that a model yields "projections" rather than "forecasts" gives little comfort to the user who must ultimately make decisions based at least in part on model outputs. The "sensitivity studies" often given to show the consequences of reasonably altered key input parameters are satisfactory for those parameters varied, but these few studies do not inform the user of the cumulative consequence of uncertainty in the hundreds of other parameters of presumably less import.

There is promise in confining the use of model outputs to ratios or differences which will be relatively insensitive to input uncertainty but appropriate for decision making. However, without the numerical values of the propagated uncertainties it is not possible to take advantage of this effect in an objective way. In the end, there is no way for the model user to escape the need to know quantitatively the impact of the uncertainties in the model inputs. He must know whether the uncertainty is so large that a resulting conclusion is in doubt.

As evaluation of the uncertainties in model inputs becomes accepted as part of the task of gathering a suitable data base, the necessarily increased scrutiny of the underlying information will tend to yield more accurate values for the model's input data. When uncertainties have been propagated to the model output, the modeler will know which classes of input uncertainties have the greatest impact on the outputs, so that future efforts to improve the accuracy of the data base can concentrate at the points of greatest leverage.

As part of a broader effort to validate the Long-run Energy Analysis Package (LEAP) developed by the Office of Applied Analysis of the Energy Information Administration, (6) the authors are beginning an attempt to develop uncertainty information appropriate for the parameters of LEAP. This paper indicates some of the approaches we expect to use in this work as well as the advantages we expect to obtain to the extent the effort is successful.

WHAT IS NEEDED TO PROPAGATE INPUT UNCERTAINTIES

Formulation of Uncertainty Propagation

Uncertainty propagation may be based on a first-order Taylor's expansion of a model's output about the solution point for the nominal or "best" set of input parameters. To fix ideas, let $R(\underline{a})$ be a specific response or result of the model when run with the set of input parameters which are the components of the parameter vector $\underline{a}$. The response of interest might be the fraction of GNP estimated to be spent for energy in the year 1995; and the a_i are input to represent in a specific model formulation our resources, demands, and economic behavior. (A model response is defined as any desired combination of output prices and commodity flows.) Each of the parameters is uncertain by an error or disturbance term δa_i. Since we do not know the values of these error quantities we will have to deal with their probability density functions.

The (unknown) error vector $\underline{\delta a}$ induces a corresponding error in the response R which may be estimated using a first order Taylor expansion about the expectation value of the parameter vector $E\{\underline{a}\} \equiv \underline{\hat{a}}$.

$$\delta R = R(\underline{\hat{a}} + \underline{\delta a}) - R(\underline{\hat{a}}) \simeq \sum_i (dR/da_i)_{\underline{\hat{a}}}\, \delta a_i \quad , \tag{1}$$

or in vector notation

$$\delta R \simeq S^t\, \underline{\delta a} \quad , \tag{1a}$$

where S is the vector of derivatives of the response with respect to the various parameters. The S_i, or sometimes the relative quantities $(a_i/R)(dR/da_i)$, are frequently called the sensitivity coefficients for R.

Since the modeler uses his best estimates of the input parameters, the best estimate of $\delta R = 0$. More precisely, the expectation value of δR or the average over its probability density function is zero.

$$E\{\delta R\} \simeq E\{S^t \underline{\delta a}\} = S^t\, E\{\underline{\delta a}\} = 0 \tag{2}$$

But the uncertainty in R is not zero, and the portion of this uncertainty which arises from uncertainties in the a_i's may conventionally be expressed in terms of the mean square deviation or variance of R:

$$Var(\delta R) \cong Var(R)$$

$$Var(\delta R) \cong E\{\delta R^2\} \cong E\{S^t \underline{\delta a}\ \delta a^t S\} = S^t E\{\underline{\delta a}\ \underline{\delta a}^t\} S \equiv S^t V(\underline{a}) S$$

The matrix V(a) is called the variance-covariance or uncertainty matrix of the input parameters a_i. Within the limits of linear uncertainty theory it contains all the information needed to obtain through Eq. (3) the output uncertainty for any response R provided that the sensitivity coefficients can be obtained. The diagonal elements of V(a) are the familiar variances of the input quantities; if the probability density function of a is boldly assumed to be "normal" then there is roughly a probability of 2/3 that the true value of a_i lies in the interval $\pm (V_{ij})^{1/2}$ about the best estimate. Usually, the off-diagonal elements of V(a) are best considered in terms of the correlation matrix defined as $Corr(\underline{a})_{ij} = V_{ij}/(V_{ii}\ V_{jj})^{1/2}$, so that all the elements have values in the range (-1,1). Note that for a model such as the current LEAP, which has over 2000 input parameters, the covariance and correlation matrices are very large indeed, and one may well wonder whether the information required for application of Eq. (3) could ever be assembled. Fortunately, since the parameters a_i are gathered from diverse sources, we can start with the assumption that most of the off-diagonal elements are null. Nonetheless, the task of obtaining the necessary correlations is as formidable as it is essential. The labor involved is attractive only because it is needed to assess the validity of model outputs.

Obstacles to be Overcome

Many problems must be solved to apply the uncertainty analysis of Eq. (3) to an energy economy model. Some of these are listed in the paragraphs below.

The complete set of sensitivity coefficients may be difficult and expensive to derive; moreover, the derivatives may not even be continuous over the important range of variation of the parameters. Research into mathematical methods is required in this area, and a companion paper addresses this important question. (5)

The uncertainties in some input quantities may be so large that the validity of the first-order expansion of Eq. (1) might be questioned. Initially, the reasonable course is to defer this question on the assumption that the first-order theory is likely to suffice for those cases where output uncertainties are acceptably low. If large uncertainties are found for several parameters with high sensitivity, reruns of the model would be appropriate with these input parameters separately varied over their likely ranges. If many parameters have marginally large uncertainties, an overall check of the uncertainty propagation should be considered employing Monte Carlo variation of the input parameters. Parameter values to be used would be perturbed by amounts drawn from the (probably multivariate normal) joint probability density function defined by the uncertainty matrix developed for the input data.

Perhaps the greatest obstacle to our current effort is that parameter uncertainties and correlations have not been given for the parameters of LEAP or for other comparable models, so a data base of uncertainty information must be constructed.

Assembling *a priori* Information to Simplify Uncertainty Analysis

The approach just discussed provides a format around which to organize the uncertainty analysis but the level of effort necessary to carry out such an analysis may be very large. Therefore, prior to undertaking a full blown uncertainty study, the analyst would do well to lay out the information available and, if necessary, draw simplifying assumptions to increase the tractability of his work. Such simplifications may hinge on three types of prior information: knowledge of the parameters of the variance-covariance matrix; knowledge of the range of relevant policy use to which the model will be put; and, knowledge as to the importance of individual parameters in generating model outputs.

Although it may appear trite to state it, the overall productivity of the uncertainty analysis hinges on the quality of the information contained in the variance-covariance matrix, and so the class of data available may guide the final work plan for the analysis. Frequently no information of this type will be available directly. When it is not the analyst has the option of constructing proxy uncertainty values from data similar to those contained in the model.

In planning the uncertainty analysis, two additional types of a priori knowledge should be considered. The first involves limiting the analysis to portions of the model actually used for critical analyses. To the extent that one knows in advance that all outputs are not used equally, it may be possible to limit the analysis, and by doing so concentrate additional resources on fewer areas. The second consideration is whether or not individual parameters can be ranked, if only crudely, as to their overall importance in generating the model solution. If, for example, only 10% of the parameters account for 90% of the model output variance, it should be possible to concentrate efforts on fewer parameters.

Thus, to the extent that the analyst can limit his activities to investigating portions of the variance-covariance matrix, it may be possible to bring more resources to bear on a relatively small number of parameters.

PRACTICAL EVALUATION OF PARAMETER UNCERTAINTY MATRICES

While we have only begun to consider the uncertainties in the LEAP input parameters, there is enough experience in related enterprises to indicate the main approaches to be used. Experience will lead to further development of methods.

Utilize Prior Estimates

A first step is to represent what is known, including correlations. The uncertainty evaluator can return later to refine estimates of uncertainties. For example, in his report developing a data base for LEAP, Bhagat et al. (2) give ranges of values for some cost parameters. These are understood to correspond roughly to 10%-90% of the cumulative probability distribution. (3) Whether these ranges represent the scatter among corresponding data for different plants the ambiguities the authors encountered in obtaining any value, it is reasonable to adopt as initial values the uncertainties derived from these ranges.

Use Statistical Techniques

Simple statistical methods may be used whenever several sources of corresponding data can be found. Suppose that cost data can be compiled for several (N) coal-burning electric plants and that all these data can be corrected for passage of time and adjusted as necessary to be appropriate for a nationally aggregated model like the current LEAP. For each plant, suppose there are several cost components parameters a_{in}. Then the portion of the parameter uncertainty matrix corresponding to the parameters for this specific activity can be estimated directly from the scatter among the collected data. The sample variance-covariance matrix element linking parameters a_i and a_j is defined as

$$VS_{ij} = \sum_n (a_{in} - \overline{a}_i)(a_{jn} - \overline{a}_j)/(N - 1) \quad , \tag{4}$$

if the a_i and a_j have been taken as an average over the same set of data. (Otherwise the denominator is more appropriately taken as N.) If these sample covariances are boldly taken as correct for representing the probability density function of the coal-burning electric plant cost parameters, then the uncertainty matrix components of the mean parameter set can be taken as

$$V_{ij} = VS_{ij}/N \quad . \tag{5}$$

This relation can be obtained for example by considering a weighted least-squares combination of the N parameter sets.

When portions of the overall parameter covariance matrix are estimated using Eqs. (4) and (5), there will be many small off-diagonal elements which will appear to have little pattern, as well as some correlations large enough to appear real. In this technique one accepts the helter-skelter appearance of the matrix elements with confidence that on the average the elements obtained will give a correct representation for uncertainty propagation.

Consider Common Uncertainty Components

The effects of common underlying uncertainties are best analyzed starting with linear error terms, not variances. A common uncertainty often systematicly affects the value chosen for more than one input datum. As a simple example we trace the covariance terms introduced when construction cost data are combined and represented, as in LEAP, by specific capital costs (SCC_i) and capital labor fraction ($CLFR_i$) for the ith type of facility. (1) Assume that the underlying data are the estimated or observed materials cost M_i and labor cost L_i (which might be further subdivided by classes of labor and types of material), where all these costs are normalized for a given sized plant. Then

$SCC_i = L_i + M_i$, and $CLFR_i = L_i/(L_i + M_i)$. The first step is to express these relations in terms of small error terms:

$$\delta SCC_i = \delta L_i + \delta M_i$$

$$\delta CLFR_i = (M_i \ \delta L_i - L_i \ \delta M_i)/SCC_i^2 \quad . \tag{6}$$

If the labor and materials estimates are for (incorrect) simplicity assumed not to be correlated, i.e., if $E\{\delta L_i \times \delta M_i\} = 0$, we can write the input covariance elements induced by the choice of variables:

$$V(SCC_i) = E\{(\delta SCC_i)^2\} = V(L_i) + V(M_i)$$

$$V(CLFR_i) = M_i^2 V(L_i) + L_i^2 V(M_i) \ /SCC_i^4$$

$$CoV(SCC_i, CLFR_i) = M_i V(L_i) - L_i V(M_i)/SCC_i^2 \quad . \tag{7}$$

These distinctions may be inconsequential for the current version of LEAP, since there is no balance of materials or labor use against supply and only the total capital cost should impact the model solution. However, the details should become more relevant for distant years because the uncertainty will become significant in the assumption that one choice of constant value dollars can be correct for both labor and materials through the year 2020, and the breakdown of specific capital cost into its components will be required to assess covariance elements $Cov(SCC_i, SCC_j)$ linking the cost estimates for the various types of facilities considered.

The above pattern of analysis, in which the behavior of uncertainty propagation is traced through the same process used to develop the parameter itself, is typical of the manipulations required in developing input parameter uncertainty matrices.

Review Implications of Common Parameter Values

A common numerical value for a series of parameters does not necessarily reduce the number of model parameters for uncertainty analysis. In the LEAP example, many corresponding parameters for different types of facilities are given common values, presumably for lack of sufficient cause to choose different values. (6) Examples are the discount rates for debt and for equity, which are given the same values for all industries to permit calculation of the present values of future profit flows. The approximation seems reasonable, since the performance of the financial markets may tend to influence all industries comparably; certainly if different discount rates were specified for the various industries the main component of the uncertainty in these rates would be strongly correlated for the various industries. Yet it is uncertain that the same rates will be appropriate for the various industries, and a relatively small difference between discount rates for competing industries would affect relative willingness to invest in new capacity. Since a prime goal in long-run models such as LEAP is to characterize adequately the market penetration of new energy technologies, the discount rates for the various technologies may need to be treated as separate parameters for uncertainty analysis.

A second example from LEAP is that few parameters are indexed in the time variable. Taking the inflation rate as zero is surely a reasonable simplification, yet we know that the mean inflation rates for various producer categories differ by as much as a percent, (4) so relative costs for competing industries may shift in the future for reasons not easily related to technological change. At this time we do not assume that early attention should be awarded to this complexity! The point is that the parameters given common values in the model should be scanned for cases where the parameters must be kept separate for uncertainty analysis.

Scan for Blunders

Blunders in assessment of parameter uncertainties may be sought using a variety of techniques. The simple ones depend on careful review of labeled arrays looking for unreasonable values. Finding mistakes is facilitated by scanning tables of absolute covariance elements, relative covariance elements, and correlation matrices. Mistakes are also found after uncertainty propagation if the impact of a parameter's uncertainty seems incredibly small or large.

A sophisticated test is based on the knowledge that every proper parameter uncertainty matrix must be positive definite; that is, all the eigenvalues of the matrix must be positive. This must be so, or a set of sensitivity coefficients could be found which would correspond to a negative variance for some response! So a strong test of a covariance matrix is to diagonalize it and scan for any negative eigenvalues. This may be practical if the overall uncertainty matrix can be arranged to consist of relatively small blocks of nonzero elements along the major diagonal.

CONCLUSIONS

One part of the uncertainty in the output of an energy economy model such as LEAP arises from the uncertainties in and correlations among the model's input parameters. The modeler's and user's confidence in a model, and therefore its utility, will be increased if it can be shown that the projected impact of the input uncertainties on the calculated responses of interest are sufficiently small as not to imperil the intended uses. Since a method exists which in principle can produce these projected output uncertainties, this method should be applied to models of potential importance. Procedures exist to help the evaluator define the input parameter variance-covariance or uncertainty matrix, but the effort involved in this evaluation will be substantial even if the effort is confined to those parameters be lieved to be most important. When the methodologies for evaluation of input uncertainties and propagation to output model responses have been demonstrated, it will be appropriate for the input uncertainties to be assessed in conjunction with the development of the input data base. Perhaps the greatest impact of uncertainty analysis on model outputs will be the numerical clarification of which questions can reasonably be asked of a model consistent with the information content of its data base.

REFERENCES CITED

1. Adler, R. J., Cagalet, E. G., Haas, S. M., Marshalla, R. A., Nesbitt, D. M., and Phillips, R. L., The DFI Energy Economy Modeling System, Decision Focus, Incorporated (December 1978).

2. Bhagat, N. Beller, M., Hermelee, A., Wagner, J., and Lamontagne, J., Evaluation of Technological Data in the DFI and PIES Models, BNL-50949, Brookhaven National Laboratory (April 1979).

3. Bhagat, N., Brookhaven National Laboratory, private communication (June 1979).

4. Council of Economic Advisors, Annual Report, U. S. Govt. Printing Office, p. 245 ff (January 1979).

5. Oblow, E. M. Weisbin, C. R., and Alsmiller, R. G., "A Sensitivity Approach to the Validation of Long-Range Energy-Economic Forecasts," Proceedings of Conference Energy Modeling II, August 13-17, 1979.

6. Office of Applied Analysis, EIA, DOE, private communication (1979). See also Annual Report to Congress for 1978, United States Department of Energy, Energy Information Administration, DOE/EIA-0173/1, p. 73.