MODEL VALIDATION:

AN OVERVIEW WITH SOME EMPHASIS ON RISK MODELS

Bruce A. McCarl and A. Gene Nelson

INTRODUCTION

Model validation according to Lady "[refers] to activities to determine how a model performs. In particular [validation refers] to...the degree of fit between the model [results] and 'reality'." However this definition is too broad to be useful and is perhaps too firmly grounded in the scientific tradition. A model cannot represent all of "reality," so the focus needs to be narrowed to that part of reality which the model is intended to represent. Further, a model may not need to represent the intended reality perfectly, rather a model needs to abstract reality "adequately" for the model's anticipated use.[1] Thus, model validation, for the purposes of this paper, refers to activities designed to determine the utility of a model, i.e., whether it is appropriate for its intended use(s); whether the benefits of improving model performance exceed the costs; whether the model contributes to making of "better" decisions, and possibly how well it performs compared to alternative models. Note that our primary concern is with the model's "validity in use." Therefore, the validation process relates to the application of the model and not to the model per se.

Models are heavily used in risk research. However, formal model validation has been infrequently attempted based on the published reports.[2] Nevertheless many studies contain strong conclusions relative to potential policy or other uses without assurance regarding model validity. Validation exercises can make our models more credible; provide a "better" platform from which to make policy or operational decision recommendations; allow us to amass evidence on the usefulness and applicability of our theory and models; provide insights into proper ways of modeling reality (for example, when our preconceived notions fail, validation testing allows us to develop other hypotheses); and increase the likelihood of our research results being used by decision makers.

Higher priority should be given to model validation. Thus this paper provides an overview of issues and approaches relating to model validation in the context of risk research within agricultural economics. This topic has not been extensively treated within the literature. Consequently we attempt to overview the relevant literature citing related discussions, review articles and examples.

Bruce A. McCarl is a professor of Agricultural and Resource Economics at Oregon State University. A. Gene Nelson is professor and head of Agricultural and Resource Economics at Oregon State University. Paper presented at the 1983 annual meeting of Regional Research Project S-180, San Antonio, Texas, March 28-30, 1983. We appreciate the comments of L.M. Eisgruber on an earlier draft of this paper.

The paper is structured into three major sections. First we discuss general issues regarding validation of risk models, i.e. forms of validation; risk models and their need for validation; validation issues; and issues regarding risk model validation. Second, procedures for validation are presented, including discussion of data requirements, tests, testing procedures, test criteria and ways of dealing with invalid models. The concept of operational validation is also briefly discussed. Finally, there are a few concluding comments.

VALIDATION AND MODELING BACKGROUND

FORMS OF VALIDATION

The ultimate validation test would involve making a model available and then observing whether it is, in fact, useful for its intended purpose. Clearly this is not always possible. Often measures of model validity are required for a model to even be considered for use. Moreover it is time consuming and expensive to give models such a trial by fire. Consequently validation is often done through a mixture of a priori tests and continuing evaluation.

Validation exercises can vary widely. Models may be validated by assumption (i.e., embodying statements such as the model was built "following such and such a theory" or "based upon perceptions of the problem"); they may be subjected to formal testing procedures; or a model's forecasts may be tested over a number of years, or situations.

In practice model validation procedures can consist of efforts designed to either <u>technically</u> and/or <u>operationally</u> validate a model. (Schellenberger apparently proposed these terms, however he defines the terms differently and also includes the term "dynamic validity"). Technical validation, as we will use the term, represents validation of a model for a particular use and embodies (a) testing of assumptions and data, (b) testing of technical equation logic and (c) testing of predictive ability. Operational validation represents validation of a model as it is implemented and involves (a) identifying constraints on appropriate uses, (b) tests of mechanisms which adapt the model to a particular use (i.e., data revision procedures), (c) tests of model updating procedures and possibly (d) repeated technical validation exercises for a number of potential usages. The extent to which these validation procedures are applied in the application of a particular model depends on the costs of the validation process versus the benefits of the validity information derived.

RISK MODELS: THEIR USE AND VALIDATION

Risk models can be used for three purposes: (a) structural exploration, (b) prediction, or (c) prescription.

Risk models are used in structural exploration to examine how risk enters into the formation of "reality." Such studies attempt to discover risk's role in the economic structure. Examples of such uses are Just; Lins et al. and Hazell et al. and Feder. Validation is of particular importance in this setting as invalid models lead to improper conclusions on the role of risk in influencing decisions and thereby reality.[3] However for many types of models developed for structural exploration, validation may only be possible by testing its predictive ability. An exception would be conventional statistical tests of model fit in an econometric study.

Risk models are used in prediction to forecast the consequences of decisions. These predictions may be of specific levels of variables (i.e., what will average income and income risk be under a policy) or of changes in variables in a comparative statics sense (i.e., how much will income and income risk change if a policy were implemented). Such predictions are intended for either policy makers as analytical input to the policy formation process (as in Musser and Stamoulis); or for decision makers as inputs into decisions regarding changes in external parameters (i.e., Barry, Baker and Sanint) and/or their resource base (i.e., budgeting of machinery changes as in Danok, McCarl and White). Validation is important in predictive settings as (a) validation indicates the degree to which predictions should be believed, (b) validation exercises help avoid prediction errors resulting from the use of invalid models, (c) validation permits statements regarding the models' ability to predict, thereby giving credence to the empirical model, the modeling approach and the underlying theory (see for example, Robison where empirical evidence is drawn together regarding the validity of Expected Utility Theory). This last use is particularly important given the wide variety of risk models which have been suggested, analytically and/or empirically used, and theoretically and/or empirically criticized.[4]

The prescriptive use of risk models involves solving specific problems for decision makers. In a prescriptive application a decision maker would utilize the model to determine the best strategy to be implemented. Validation is important in this setting to (a) give the model credibility so decision makers may use it, (b) evaluate the ability of the model to generate "best" solutions and (c) evaluate the ability of the model to adequately represent the decision process.

The presentation below will concentrate on the validation of models to be used for structural and predictive purposes. This is done because very few agricultural economics models are used for prescription. There are very few uses in agricultural economics where models are, in fact, directly used to make decisions. Models are most often used in either a comparative statics sense to predict the income or acreage consequences of external changes or a structural exploration sense to somehow discover whether risk is an important factor. Even when models are used with farmers, most seem to be used in comparing alternatives and giving the direction of change arising from those alternatives. We know of very few models which attempt to make decisions for farmers.

A GENERALIZED MODELING PROCESS

Before beginning a discussion of validation procedures is useful to have a general modeling concept. Risk models run the spectrum of quantative techniques used in agricultural economics including, for example, such diverse things as eonometric models used to study structure, mathematical programming models used to study policy decision making, stochastic simulation models used to study dynamic influences of parameters, and nonstochastic simulation models used to project the dynamic consequences of changes in "known" parameters.

Figure 1 gives a unified modeling structure common to most models. A modeler begins with a perception of the problem, then using theory and knowledge of the problem, derives a modeling approach. Subsequently data are collected and the model relationships fit. These model relationships may be fit formally in the case of econometrics or, less formally with averages or heuristically with single values constructed for simulation and programming models. (Some studies such as structural exploration studies in econometrics may stop once the relationships are fit.) The next phase is model construction in which the empirical relationships are integrated into a unified model; then a base case model is developed. Experiments are designed; model experimentation carried out; and problem results obtained. Validation occurs throughout this process.

The generalized process in Figure 1 allows validation to be discussed without a great deal of reference to the specific tool being validated. This generality has both costs and benefits. Those wishing details on specific techniques should examine the literature reviews relating to specific areas, i.e., econometrics [Dhrymes et al., Zellner], simulation [Johnson and Rausser, Anderson, 1974, Shannon, Sargent], and mathematical programming [McCarl]. Also Gass [1980, 1981] give a comprehensive overview and references. Finally the reader should note that the discussion below does not relate well to directly estimated equation econometric models including risk which are not used in forecasting. These are usually judged valid or invalid by statistical criteria.

ISSUES REGARDING VALIDATION OF ANY MODEL

While validation is presented as a desirable exercise, those considering undertaking validation exercises should be aware of a number of issues which arise. Perhaps the most fundamental issue deals with the

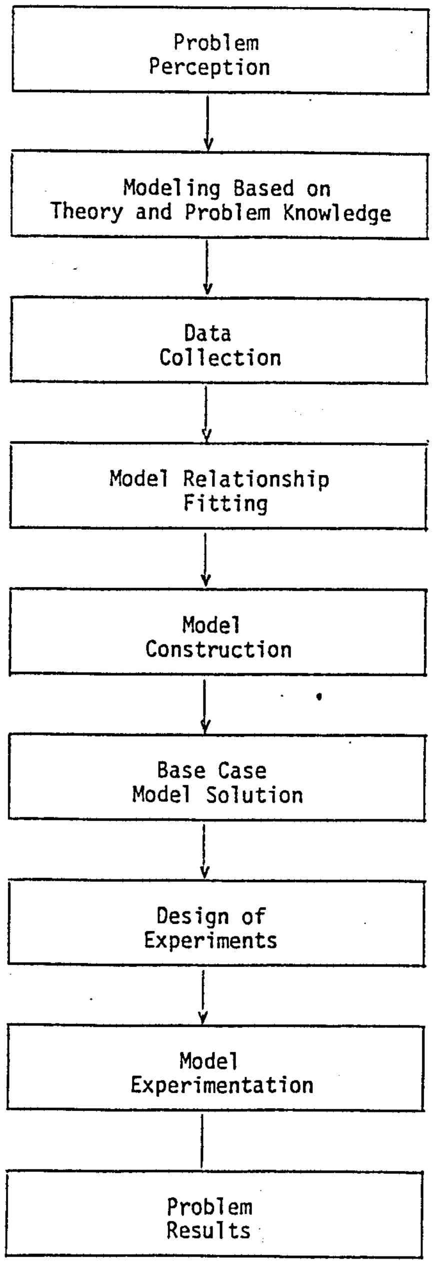

Figure 1. Generalized Modeling Process.

nature of the outcome of the model validation process. Models can never be validated, only invalidated. Further testing of a model in one or a few of its uses does not guarantee that it will perform satisfactorily in all uses. Thus the outcome of a model validation process is either a model that has been proved invalid, or a model about which one has an increased degree of confidence, not a valid model. Nevertheless we will use the term valid model from now on to denote a model which has passed validation tests.

Another fundamental issue is the subjectivity of the validation process. There is not, and there most likely never will be, a totally objective and accepted approach to model validation. Model validation is subjective in many ways. Modelers subjectively choose the tests with which they will validate, choose criterion to measure the validity/invalidity of their model, choose what to validate within their model, choose in which uses the model will be validated, choose what data to use in validating, etc. Thus, the statement "the model was judged valid" can mean almost anything. (See Anderson [1974] for more on subjectivity).

However, above and beyond these fundamental issues, there are a host of attendent issues. House and Ball were attempting to provide insight into whether validation was possible for policy models and summarized their arguments as follows:

> "Social science models often deal with phenomena at an empirical level where immutable natural laws cannot be ascertained; therefore, historical validation, even if possible, gives limited confidence that the resulting model has validity for predicting long-term future.
>
> The formal statistical techniques used for validation are based on assumptions about the nature of the sample, such as the assumption that the future has some known relation to the past; hence, they may have difficulties similar to those of historical validation.
>
> There is little agreement as to what it means to be able to predict the future, with or without formal models.
>
> We are still very much in our infancy when it comes to measuring the state-of-the-present, using such techniques as indicators; consequently, we are hard put to say whether we have reasonable gauges with which to measure the future.
>
> Complex models are harder to validate than simple ones; but for most modern day problems, more complex approaches are necessary for policy analyses.
>
> Validation must be considered in relation to the type of model and to the purpose for which the model is used. Each combination has different implications for the feasibility, appropriateness and specific technique of validation.

Models used to aid decisions and policy analysis should be judged on the basis of their utility in aiding decisions relative to alternative procedures, rather than on the same basis as models used in science.

There are risks in insistence on validation, since inappropriate application of validation could unfairly discredit models that have real utility."

There are several more points which can be made adding to, expanding upon, and clarifying the comments of House and Ball.

1. A model does not have to be valid in order to be used. There have been models used which have not been tested for validity and some of these models probably would not pass validation tests. This observation leads to an essential issue: To what extent is validity required in the use of models? Models which are invalid in terms of prediction of specific values may be valid (i.e., used) if they predict the right direction and magnitude of change. Models will also turn out to be useful if they properly identify what will change and what will not, even if they are horrible at predicting exact values. (Geoffrion amplifies on this point arguing that models are not used for numbers but rather for understanding situations.)

2. Even if a model is "valid," one should be cautious when predicting the future. The future is often not like the past, new factors or variables can enter changing data and invalidating assumptions, possibly invalidating the model. For example policies may change which change the structure of the situation and invalidate the model (as argued at length by Lucas).

3. Models may be improprerly judged valid or improperly judged invalid, allowing the possibility of both type I and type II errors. An important distinction in the validation process is that the model must be judged valid or invalid for its' specific use. Thus, models should be tested only within the bounds anticipated for the model's intended use. Similarly, models which are not judged to be valid for all anticipated uses should be qualified as to their validity.

4. Models may be judged valid for the wrong reasons due to specification error. For example, in a structural study a factor may be identified as important whereas it was by chance correlated with an important omitted factor. Specification error can also arise where excessive constraints are placed on the model, forcing it to validate. Similarly validating the model over the historical data which were used in model construction may lead to excessive confidence. (Howrey and Kelejian show this does no more than reproduce statistical test results when using econometric simulations.)

5. Models once validated are not valid for eternity, rather they must be continually validated in an operational sense.

ISSUES REGARDING VALIDATION OF RISK MODELS

The above general discussion has not focused on the problems specific to risk models which are emphasized in S-180. Such models have several underlying features which complicate validation.

1. Risk models by nature deal with stochastic realities. Ordinarily when doing validation tests reality is compared with model solutions. However, when reality is stochastic the possibiity exists that the observed reality is of low probability, i.e., suppose that a year with extraordinarily low yields and prices was chosen for model validation. Conceptually, one needs to test whether the output probability distribution of the model's output (or the parameters of output distribution) corresponds to the probability distribution of reality. This implies that risk models need a more complete definition of reality than do nonrisk models (i.e., more data).

2. In dealing with stochastic realities, decision makers do not know all the input parameters at the time decisions must be made. Consequently decisions may be made that are relatively robust to the stochastic parameter outcomes. This implies that model validation may be difficult if the data used do not exhibit the subjective probability distributions which led to the observed reality (even more data). Still, there are certainly cases which can arise in which stochastic outcomes alter the feasibility of plans.

3. In validating risk models, particularly those used for structural exploration, one must remember that it is possible to confuse profit maximizing behavior with risk averse behavior (for example, see the cases in Pope (1981); Roumasset; and Baker and McCarl).

4. Operational validation is a particular concern in agricultural economics risk models as these models are quite often utilized in quite diverse situations (i.e., with several different farmers each possessing different subjective probabilities and different resource endowments).

GENERAL APPROACHES TO VALIDATING MODELS

Models may be validated as a whole or in parts. Validation may also be done by assumption and/or by results. Also approaches to validation differ dependent on whether one is attempting technical validation or operational validation. The discussion below largely relates to technical validation. Some brief comments on operational validation will be offered later in the paper.

VALIDATION BY ASSUMPTION

Validation may be done by assumption, by results, or by both assumptions and results. All models which are considered valid go through a validation by assumption phase, initially. Validation by assumption involves a combination of several different approaches. Models may exhibit validity through expert opinion, antecedent, theory, data or logical structure. Validity through expert opinion is manifest in the statement, "Based on the modeling team's expert opinion the model is proper." Antecedent validity is called on when the modeling team says, "This model was used because so and so used it." Antecedent validity may also involve reliance on others previous validation efforts. Theoretical validity is called on when the modeling team states, "Based on theory the model represents the proper theoretical relationships, and therefore it is valid." Data validity is present when the modeling team says, "The procedures utilized to construct a model either adequately represent current data or are adequately deduced from history." Logical model structure validity is embodied in the statement, "The model equations have been checked and tested and they are properly specified. Thus, since the equations are good, the model is valid." All of these validation tests are characterized by one statement. Validation is <u>assumed</u> not tested.[5] Unfortunately far too many models, perhaps most risk models, are judged valid by assumption. (Many use validation by assumption along with the plausibility test below.)

VALIDATION BY RESULTS

The general scope of this manuscript along with time, space, and knowledge constraints precludes an extremely detailed discussion of the science of validation testing. Nevertheless, this section attempts to provide an overview of the procedures involved, along with an attempt to reference relevant discussions and examples.

Validation by results involves four components. First, one must have a parameter-output set with which the model results are compared. Second, one must conceptualize appropriate ways of generating results from the model for comparison, i.e. specific validation tests. Third, one must have a procedure for measuring degree of association including a criterion for accepting or rejecting the statement "the model is close enough." Fourth, one must have ideas on what to do if the model fails validation.

<u>Parameter-Outcome Sets</u>

Reality as measured numerically, consists of both parameters leading to a situation and outcomes arising from that situation and provides a basis for comparison in validation testing. Several things need to be considered in assembling these data.

1. A model should not be validated using only the data which were used in model construction. There is the definite possibility that the data implicit in the model are not representative of the parameters leading to situations for which the model is used.

2. The data themselves must be subjected to validation (a point made and expanded on by Gass [1981]). Inconsistent data (for example, when the "observed" output is inconsistent with the "observed" levels of the decision variables, and associated parameters) virtually guarantees the failure of validation tests.

3. The data needed for validation should be as comprehensive as possible with observations present for all decision variables, all output measures, all usages of resources and all prices for both product and resources. However, partial data sets can be used when complete data sets are not available.

4. Validation of risk models imposes special data requirements on the development of parameter-outcome sets. These sets should contain information on the distributions of the parameters-outcomes.

5. Parameter-output sets do not have to be formally established. Often formal parameter changes are utilized along with informal perceptions of the relevant outcome set, i.e., a 10% increase in production would lead to an increase in the cost of production. Formal data are preferred, but informal data may be used in determining if the model results are plausible.

6. The parameter-outcome sets employed should to the extent possible be representative of the model in use. While this statement is totally obvious, it has important implications. Data sets may be selected which, while they appear to be appropriate, are not, for example, models used for simulating long run equilibrium should not be compared with short run disequilibrium data. In addition, data for validation purposes should be chosen so that model use is tested in as many ways as possible.

Test Design and Implementation

Technical validation of results can be done via a number of tests. These tests may be crudely subdivided into seven categories, and are presented in order of their increasing complexity. Each of these tests may be done singly, or for many parameter-outcome sets, either historically generated or with new data. In addition the tests may be done on the whole model or on just portions of the model.[6] A general procedure for using these tests is given in Table 1 and examples of their use are given in Table 2.

A. Plausibility Test. The purpose of this test is to examine whether the model creates "plausible" results. The basic test involves sensitivity analysis of the solution under a situation about which the modeling team has opinions on appropriate direction and possibly magnitude of model

Table 1. A Procedure for Model Validation.

Step 1. Enter the parameters, constraints, alternatives, etc., which implement the particular validation test at hand.

Step 2. Obtain a solution to the model.

Step 3. Evaluate the results. There are two possibilities for the results. Either the model has solved with an answer or the model has somehow blown up.

a) If the model has blown up, discover why. Programming models maybe either unbounded or infeasible. Simulation models may exhibit numerical difficulties or may incorporate equations which cannot feasibly represent the particular variable values Fix the model and go to Step 2 or go to Step 6.

b) If the model has a solution, then perform association tests to discover the degree of correspondence between the numerical reality and the model solution. These tests should be done on all possible output variables, imputed prices, and on aggregates such as income, total acreage, etc. Go to Step 4 or 5 depending on the outcome of the tests.

Step 4. If the model results exhibit a sufficient degree of association, then do a more complex validation test, going to Step 1 or determine that the model is not invalid for use and terminate the validation procedure.

Step 5. If the model does not pass the validation test, consider whether a) the "reality" data are consistent and correctly calculated, b) the data are properly entered into the model structure appearing in the right equations, etc., c) the assumptions underlying the model structure are proper and correct. If the deficiencies in the model leading to the invalid solution are somehow corrected, go to Step 2 and repeat the validation test, otherwise go to Step 6.

Step 6. If the model is judged invalid then address the questions about whether the model needs to be fixed, discarded or qualified (discussed in invalid model section of the paper). If the model is revised then go back to either Step 2 or one of the earlier validation tests dependent upon the extent of model revision. If the model is qualified in terms of use, then either continue this validation test if there is anything remaining to be done (go to Step 2), move on to higher validation tests (go to Step 1), or accept the model for use (terminate validation).

Table 2. Technical Validation Examples.[a]

Validation Test	Type of Model: Econometric Simulation	Non Econometric Simulation	Mathematical Program
Plausibility	[d]	Anderson (1971) Miller and Halter	Delgado and McIntire
Possibility	[f]	[f]	Kutcher
Supply Function	[e]	[e]	Kutcher Rodriguez and Kunkel
Dual Supply Function	[e]	[e]	Kutcher Rodriguez and Kunkel
Prediction	Just and Rausser Lins, et al.[b] Just[b]	Miller and Halter Herath, Hardaker and Anderson	Barnett, Blake and McCarl Kutcher Brink and McCarl[c] Hazell and Pomerada
Predictive Change	[d]	[d]	Hazell, et al. Brink and McCarl[c]
Predictive Tracking	Just and Rausser	Jones and Brockington Singh	Pieri, Meilke and MacAulay

[a] This table was rather hastily drawn together. In addition we had great difficulty finding risk models. (We couldn't even find the word "validation" in the W149 proceedings.) Thus this table contains both risk-free and risk-including models.

[b] We are not sure these should be included. They are simple regressions and the only validation tests are those statistics arising from a regression.

[c] This study involves validation against another model.

[d] Undoubtedly has been done; can't find one.

[e] Appears possible, would probably be difficult.

[f] Probably has been done.

response. The model is run with a particular change and the changes in the outcomes are compared with the modeler's perceptions. Models failing to give plausible results are either invalid or the perceptions of "plausible" results may be improper. Formal association tests are not done (the change test below is the formal version). The plausibility test is the most commonly used validation-by-results test and along with the validation-by-assumption statements give the extent to which many, perhaps most, models are validated.

B. Possiblity Test. The purpose of this test is to examine whether it is possible for the model to duplicate a "reality" situation. The basic test involves fixing the parameters and decision variables at the reality levels and examining the model results.[7] The results may be consistent (statistically) with reality or they may be inconsistent. Mathematical programming models may be subjected to a "dual" possibility test in which the prices of various resources are fixed within the dual of the model and dual feasibility explored. Severe violations of these tests imply that the model could not possibly replicate the real world solution.

C. Supply Function Test. The purpose of this test is to examine the intersection of the model supply function and an inelastic demand function, i.e. whether the marginal cost of production in the model is anywhere close to the "reality" price of a good. This test is executed by first constraining production to equal the levels ascertained in reality. Second, observing the probability distribution of a) dual variables, or b) changes in cost measures for small variations in quantity. Third, comparing the observed cost changes and market prices one may see if the competitive price equals marginal cost criterion can and does hold. A model failing this test may still be proper if the modeled entity a) does not behave competitively, or b) expects different prices from those received.[8]

D. Dual Supply Function Test. The purpose of this test is to examine the intersection of the model supply function and an infinitely elastic demand function, i.e. whether production at the reality price is anywhere close to "reality production." Here any prices are fixed at the expected distribution of prices and output compared with actual output. Failures of this test would indicate a divergence between the rule--produce until price equals marginal cost--and reality. (This test is usually appropriate only to aggregate, price endogenous models.)

E. Prediction Test. The purpose of this test is to examine the ability of the model when specified with "identical" parameters to those leading to an observed reality is able to predict that reality. This is the most common of the more extensive validation efforts. A variant of this test, when the model does not predict well, involves comparing the results from the prediction test with the possibility test above to see how close the unconstrained model solution is to a model solution when constrained to reality (i.e., is it an alternative optimal). Models failing the prediction test when they have passed the possibility test are unsatisfactory predictors of absolute levels of the items tested against.

However, prediction of changes may be of equal or overriding importance as measured by the next test.

F. Predictive Change Test. Models may not need to predict exactly as long as they predict change or, in some cases, even the direction of change accurately. The predictive change test is designed to test a model's performance in doing such. The predictive change test is usually implemented by developing two parameter-outcome sets, running the model with the parameters of each, then testing the degree of association between the change in model predictions and the change in the relevant measures within the outcome sets. A model failing such a test may be felt to be unsatisfactory for any comparative statics uses.

G. Predictive Tracking Test. The ability to predict one time change may not be felt to be adequate in terms of validation. Rather the ability of a model to track behavior stimulated by changes in decision variables may be felt to be crucial. To test for adequacy in this type of an exercise a model would be subjected to a series of parameters which had occurred and the model's degree of association with respect to tracking changes compared. Models failing such tests may be able to be qualified to specific cases where they do track satisfactorily.

Association Measures

Implicit in any validation by results exercise is the measurement of degree of association between model predictions and a numerically measured reality. The acceptance or rejection of closeness is ultimately subjective (as discussed by Anderson [1974]). However this subjective judgment may be based on subjective or objective measures. The subjective measure of association is usually implicit in a statement such as: On balance the model output was not significantly different from the numerical reality (for example, see Barnett, Blake and McCarl). Objective measurement of association involves statistical tests between the model and numerical reality.

Statistical association measures for validation have been extensively used by simulators. The literature reviews by Anderson [1974]; Johnson and Rausser; and Shannon point to a number of these and discuss acceptance criterion. For example, Table 3 is reproduced from Johnson and Rausser and indicates a wide variety of tests which could be utilized.

Risk models pose some special problems in terms of validation tests. First, the items to be compared often consist of distributions of variables rather than point estimates. Thus, when validating risk models the point criteria may need to be modified to either compare the entire probability distribution of "reality" or at least its moments. Second, in many studies, it may not be possible to obtain the distribution of outputs which would occur. This clearly depends upon the modeling technique used. For example, repeatedly simulated models give distributions as do models such as MOTAD which give a simulated income distribution whereas models such as chance constrained programs do not give such information.

Table 3. Evaluation Criteria for Investigating the Explanatory Predictive Power of Systems Models

	Explanatory	Predictive
Point Criteria	1) Coefficient of multiple determination 2) Durbin-Watson coefficient 3) Graphical analysis of residuals 4) t statistic 5) Chi-square or F statistic 6) Aitchison-Silvey, test of a priori restrictions 7) Ramsey specifications error tests a) omitted variable test b) functional form test c) simultaneous equation test d) heteroscedasticity test e) Chi-square 'goodness-of-fit' test for normality 8) Sample mean squared error (changes and levels) 9) Information inaccuracy statistics for sample data	1) Mean forecast error 2) Mean absolute forecast error 3) Mean squared forecast error 4) Any of the above relative to a) the level or variability of the predicted variable b) a measure of "acceptable" forecast error for alternative forecasting needs and horizons 5) t statistic 6) Chi-square or F statistics 7) Theil's inequality coefficient 8) Information inaccuracy statistics for non-sample data
Tracking Criteria	1) Number of sample turning points missed 2) Number of turning points falsely explained 3) Number of sample under-or-over estimations 4) Rank correlation 5) Test of randomness for directional estimations 6) Test of randomness for explained turning points 7) Information theory statistics for sample data	1) Number of non-sample turning points missed 2) Number of turning points falsely predicted 3) Number of non-sample under-or-over predictions 4) Rank correlation 5) Test of randomness for directional predictions 6) Test of randomness for predicted turning points 7) Information theory statistics for non-sample data
Error Criteria	1) Bias and variance of explained error 2) Errors in start-up position versus errors in explained change 3) Comparison with various "Naive" explanations 4) Comparison with indicator qualitative errors	1) Bias and variance of forecast error 2) Errors in start-up position versus errors in predicted changes 3) Comparison with various "Naive" forecasts 4) Comparison with "judgmental," "consensus," or other non-econometric forecasts
Spectral Criteria	1) Comparison of power spectra for estimated and sample data series 2) Spectral serial correlation test of structural or reduced form sample disturbances 3) Cross spectral statistics of relationships between estimated and actual sample values	1) Comparison of power for predicted and non-sample data series 2) Spectral serial correlation test of structural or reduced form non-sample disturbances 3) Cross spectral statistics of relationships between predicted and actual non-sample values

Source: Johnson and Rausser

Dealing with Invalid Models

When a model is judged invalid during validation testing, then the modeler has one of four options. First, and most obviously, the model can be discarded. Second, the validation test results could be judged as improper and the result ignored (if for example the parameter-outcome set data were found to be inadequate for validation testing). Third, model use can be qualified, i.e., a model which did not pass the prediction test but did pass the change test could be qualified by saying that the model should not be used to predict specific levels of the output variables, but can be used to reliably predict changes (e.g., as in Brink and McCarl). Models may also need to be qualified in terms of ranges of data over which they are applicable. Fourth, the model could be "fixed." The purpose of this section is to discuss two issues: How can a model be fixed? Are there cases when models are not capable of being fixed?

The possessor of an "invalid" model should remember that models are abstractions of reality and, even after a competent modeling job, can be inadequate abstractions. During the modeling process many assumptions are made. These assumptions can be categorized into (a) assumptions regarding model specification, (i.e., variables, parameters, and data which are included/excluded) and (b) assumptions regarding the solution algorithm. (For example, when using linear programming, one assumes additivity, certainty, continuity and proportionality:) When a model is judged invalid and it is felt that the model can represent reality, then there is the failure in assumptions. Thus, the proper question arising when a model fails a validation test may well be: What assumption needs to be fixed? For example, the modeler must question whether there are important causal factors that have been omitted which should be included, or the modeler must consider whether to switch algorithms for model solution.[9] Finally one must question the validity of the modeling assumptions in conjunction with perceptions about model use and the comparability of the reality data. For example, in a large sector model, are invalid subregional acreages acceptable when valid national figures are obtained?

There may also be reasons why models will never validate. One may be comparing a full equilibrium - full adjustment model with a disequilibrium partial adjustment set of data. For example, when modeling a fallow-wheat system one could be looking at the long run and ignoring the short run constraints imposed by the number of acres that were in fallow or in production last year. In the face of such difficulties one must again refer to model purpose and question whether or not failure to validate on a particular parameter-outcome set is, in fact, catastrophic. In terms of model use, one may also need to recalculate the parameter-outcome sets being used to validate and attempting to validate on long-run averages rather than short-run situations. In general point numerical reality is not always the appropriate thing to validate a model with. For example, a model built to study the long-run effects of such things as the recent payment in kind program, may not be validatable due to a lack of relevant "reality."

TOWARD OPERATIONAL VALIDATION

The discussion above has concentrated on technical validation which is the first step in model validation. Operational validation refers to the validation of the model as it is implemented for use. The questions of operational validation appear to be more complex than those involved with technical validation especially in a risk modeling framework. Assuming that our models are ever to be used, they must be operationally valid.

In an operational context, the validation objective is to examine model performance across uses. In this sense operational validation involves more interaction with model users. Operational validation is important to give the model (or its results) credibility, so model users can interpret and utilize the results appropriately. Operational validation appears to be particularly relevant to certain activities which will be carried out under S-180. For example, objective 3 - analyzing and evaluating strategies which farmers can use in risk management programs - implies the need to validate models over a wide range of situations (enterprises, organization, debt structure, etc.), in which farmers operate.

As above, operational validation can proceed by some combination of assumptions and results. Validation-by-results procedures involve essentially the same considerations (parameter-outcome sets, tests, criteria, and dealing with invalid models) discussed previously. However a number of additional questions arise:

1. How can the input data be validated? Are the data obtained relevant to the decision to be analyzed in the model?

2. Does the model adequately represent the decision process and the major variables to be considered for each specific use? Is the structure of the model appropriate?

3. Are the results of the model appropriately presented so that they can be interpreted correctly by the model user? How does the format used for presenting the results affect their use?

4. Will the model continue to perform satisfactorily over time and across situations? Does the model easily adapt to the alternate anticipated usages?

Validation of input data is a tricky problem for models which are used in diverse situations, new data are potentially needed for every application. Validation by assumption is often called on, for example, Fishburn and Balch have invented "Super Genie" who guarantees us that our elicited utility functions will be proper. Input questionnaires may be necessary and the modeler needs to be concerned with questionnaire design so that meaningful data are obtained. Validation of these items is difficult, however agricultural economists have some experience relative to questionnaire design and farm management model use (for example, see the material in McCarl, et al. or the large psychological literature on questionnaire design). Data on subjective probabilities are of particular concern in operational validation. Hogarth argues that because man is a "selective, sequential information processing system with limited capacity, he is ill equipped for assessing probability distributions." Can we test whether the elicited probabilities are valid representations of true beliefs regarding the future? It is possible to develop a system of feedback and evaluation comparing actual occurrences with the decision makers' probabilistic predictions, for example, using scoring rules (Bessler and Moore)? Altman discusses a number of other difficulties regarding data development.

The extent to which model users should validate models for themselves is an issue in operational validation. If the model is to be useful, it will produce some results that are different from what the model user would have expected. The believability of such will depend on model credibility. This does not require review of the model's logical structure as is required for technical validation, but involves making the black box somewhat transparent so that the decision maker can grasp why a particular result occurred and may require modeling features which permit model users to conduct their own validation exercises (see McCarl et al. for a discussion of a model which permits model user validation).

The output presentation format must also be subjected to validation. An important question here is whether appropriate measures of performance and risk are being presented. This, of course, complicates the process of model validation because consideration must be given not only to what to predict and present, but also how well it is predicted and how it is presented. Conrath reveals that the format in which results are presented influences decision making. Payne [p. 440] argues, based on psychological studies, that "the way in which sources of information are displayed affects their utilization."

The above discussion has purposely not covered validation-by-results tests as this was done in the technical section above. However, there is one more relevant issue. Operational validation of models in use entails a continuing model evaluation process. Records may be kept on model predictions which may be compared with reality. This is a demanding validation test, and has been implemented occasionally in agricultural economics. For example, Just and Rausser examined several econometric models over time.

CONCLUDING COMMENTS

The process of thinking about and considering what to say within this paper has led the authors to one comment: Validation is very difficult, especially for risk models. There are many potential questions which have not been formally dealt with and we hope at the conclusion of the S-180 project, more can be said about validation of risk models than we are able to say at this point in time. Nevertheless we must forge ahead.

Risk models have been validated to a very limited extent. This poses large potential difficulties. We as researchers are faced with a very large number of risk models which we could use, yet very little empirical evidence about which kinds of risk models should be used. Simultaneously we have a lot of theory about risk models however much of this theory is not validated, in part, because few attempts have been made to validate risk models. We agree with Samuelson's comments that the validity of the theory rests in large part on its ability to be consistent with empirical fact. We therefore feel that validation is extremely important in the risk research area and that careful, competently done validation exercises will aid our research, directing us to those situations where our models can be most validly used. In addition we believe validation exercises will lend us more credibility to the use of our research by decision makers and policy makers. Therefore we think validation should be a high priority item in the conduct of risk research.

In terms of actually doing validation, Shannon gives a procedure for achieving "the greatest possible validity":

1. Using common sense and logic throughout the study.

2. Taking maximum advantage of the knowledge and insight of those most familiar with the system under study.

3. Empirically testing by the use of appropriate statistical techniques all of the assumptions, hypotheses, etc. that possibly can be tested.

4. Paying close attention to details, checking and rechecking each step of the model building process.

5. Assuring that the model performs the way it was intended by using test cases, etc. during the debugging phase.

6. Comparing the input-output transformation of the model and the real world system (whenever possible).

7. Running field tests or peripheral research where feasible.

8. Performing sensitivity analysis on input variables, parameters, etc.

9. Checking carefully the predictions of the model and actual results achieved with the real world system.

In addition we believe several other points need to be made. We believe that more microeconomic understanding of risk is needed as a precursor to more aggregate analysis. We believe this is an important justification for S-180. We believe that in doing our risk research we must spend more time understanding the problem including doing validation exercises, so that we have properly meshed risk modeling with the way decision makers incorporate risk and react to risk. We believe in studying our risk models we must more adequately address the question of how we expect these models to be used and we must adequately qualify the use of these models based on their ability to adequately perform. Finally we believe that there are substantial contributions that increased validation efforts can make relative to risk research and risk analysis, and we hope in five years a better paper can be written than this one has been.

FOOTNOTES

1. Based on scientific tradition, one might question how a model can be "valid for its use." We feel a model can be judged valid for use irregardless of whether it passes scientific validation tests. A model which is scientifically invalid may be used by a decision maker who has implicitly judged it "valid for use." Thus, scientific validity of the model may not always be required for a "valid" model.

2. This lack of attention to validation is not uncommon. Gass (1981) states "As best can be discerned the topic of model validation has not been of general interest to the OR [modeling] community" and "Papers that describe validation attempts of specific applications are not in plentiful supply."

3. Samuelson argues that in effect theories without empirical validity should be rejected. This certainly argues for validation tests which can be construed as tests of theory.

4. The profession is currently faced with an overwhelming number of proposals regarding risk modeling. For example, in using mathematical programming of choice decisions under net income risk, an individual could use game theory [Agrawal and Heady], E-V models [Freund]; MOTAD, [Hazell (1971)], marginal risk constrained linear programming [Chen and Baker], matrix diagonalization [McCarl and Tice], single or multiple index models [Sharpe], lexicographic utility [Lin, Dean and Moore], or one of many others. Very little information is available on what should be used. However, there is not a shortage of viewpoints on and criticisms of such models. For example, see the three recent papers by Hazell [1982], Pope [1982] and Robison.

5. Antecedent validity may rule out the need for tests in many cases, however, we do not feel the risk validation literature is yet strong enough for researchers to state that their models have already been validated.

6. In our reading we have seen it argued that complex models must be validated in smaller components first, then later subjected to overall validation. We do not feel validation of the whole model can be overlooked (i.e., a model is not necessarily valid even though each individual equation fits well).

7. The reader familiar with the simulation literature will note that we include what others call verification in our definition of validation.

8. There are forms of both this test and the following test relating to demand functions. Namely a demand function test could be executed by fixing factor use at an observed level and then comparing imputed price

of the factor with the observed factor price. Similarly a dual demand function test would involve fixing factor prices.

9. An attendant possibility is that the algorithm is working improperly.

REFERENCES

Agrawal, R. and E. Heady. Operations Research Methods for Agricultural Decisions. Ames: Iowa State University Press, 1972.

Altman, S. "Pitfalls of Data Analysis." Pitfalls of Analysis, G. Majone and E. Quade (eds.) New York: John Wiley and Sons, 1980.

Anderson, J. "Simulation: Methodology and Applications in Agricultural Economics." Review of Marketing and Agricultural Economics 43(1974): 3-55.

Anderson, J. "Spatial Diversification of High-Risk Sheep Farms." Systems Analysis in Agricultural Management, J. Dent and J. Anderson (eds.). New York: John Wiley and Sons, 1971.

Baker, T. and B. McCarl. "Representing Farm Resource Availability Over Time in Linear Programs: A Case Study." North Central Journal of Agricultural Economics 4(1982):59-68.

Barnett, D., B. Blake, and B. McCarl. "Goal Programming via Multi-dimensional Scaling Applied to Senegalese Subsistence Farms." American Journal of Agricultural Economics 64(1982):720-727.

Barry, P., C. Baker, and L. Sanint. "Farmers Credit Risks and Liquidity Management." American Journal of Agricultural Economics 63(1981): 216-227.

Bessler, D. and C. Moore. "Use of Probability Assessments and Scoring Rules for Agricultural Forecasts." Agricultural Economics Research 31(1979):44-47.

Brink, L. and B. McCarl. "The Adequacy of a Crop Planning Model for Determining Income, Income Change, and Crop Mix." Canadian Journal of Agricultural Economics 47(1979):13-25.

Chen, J. and C. Baker. "Marginal Risk Constraint Linear Program for Activity Analysis." American Journal of Agricultural Economics 56(1974):622-627.

Conrath, D. "From Statistical Decision Theory to Practice, Some Problems with the Transition." Management Science 19(1973):A873-A883.

Danok, A., B. McCarl, and T. White. "Machinery Selection Modeling: Incorporation of Weather Variability." American Journal of Agricultural Economics 63(1980):700-708.

Delgado, C. and J. McIntire. "Constraints on Oxen Cultivation in the Sahel." American Journal of Agricultural Economics 64(1982):188-196.

Dhrymes, P. et al. "Criteria for Evaluation of Econometric Models." Annuals of Economic and Social Measurement 1(1972):291-323.

Feder, G. "Adoption of Interrelated Agricultural Innovations: Complementarity and the Impacts of Risk Scale, and Credit." American Journal of Agricultural Economics 64(1982):94-101.

Fishburn, P. and M. Balch. "Subjective Expected Utility for Conditional Primitives." Essays on Economic Behavior Under Uncertainty, M. Balch, D. McFadden, and S. Wu (eds.) Amsterdam: North Holland Publishing Company, 1974.

Freund, R. "The Introduction of Risk into a Programming Model." Econometrica 24(1952):253-263.

Gass, S. "Decision-Aiding Models -- Validation, Assessment, and Related Issues." Working paper MS/S 81-022, Department of Management Sciences and Statistics, University of Maryland, 1981.

Gass, S. (ed.) Validation and Assessment Issues of Energy Models. Proceedings of a workshop, NBS Special Publication 564, Washington, D.C., 1980.

Geoffrion, A. "The Purpose of Mathematical Programming is Insight, not Numbers." Interfaces 7(1976):881-892.

Hazell, P. "A Linear Alternative to Quadratic and Semivariance Programming for Farm Planning Under Uncertainty." American Journal of Agricultural Economics 53(1971):53-62.

Hazell, P. "Application of Risk Preference Estimates in Firm-Household and Agricultural Sector Models." American Journal of Agricultural Economics 64(1982):384-393.

Hazell, P. and C. Pomareda. "Evaluating Price Stabilization Schemes with Mathematical Programming." American Journal of Agricultural Economics 63(1981):550-556.

Hazell, P., R. Norton, M. Parthasarthy, and C. Pomareda. "The Importance of Risk in Agricultural Planning Models." Programming Studies for Mexican Agricultural Policy, R. Norton and L. Solis (eds.). New York: Johns Hopkins Press, 1981.

Herath, H., J. Hardaker, and J. Anderson. "Choice of Varieties by Sri Lanka Rice Farmers: Comparing Alternative Decision Models." American Journal of Agricultural Economics 64(1982):88-93.

Hogarth, R. "Cognitive Processes and the Assessment of Subjective Probability Distributions." Journal of American Statistical Association 70(1975):271-294.

House, P. and R. Ball. "Validation: A Modern Day Snipe Hunt: Conceptual Difficulties of Validating Models." In: Gass, 1980.

Howrey, E. and H. Kelejian. "Simulation Versus Analytical Solutions: The Case of Econometric Models." Computer Simulation Experiments with Models, T. Naylor (ed.) New York: John Wiley and Sons, 1971.

Johnson, S. and G. Rausser. "Systems Analysis and Simulation: A Survey of Applications in Agricultural and Resource Economics." A Survey of Agricultural Economics Literature, Vol. 2, L. Martin (ed.) Minneapolis: University of Minnesota Press, 1977.

Jones, J. and N. Brockington. "Intensive Grazing Systems." Systems Analysis in Agricultural Management, J. Dent and J. Anderson (eds.). New York: John Wiley and Sons, 1971.

Just, R. "Investigation of the Importance of Risk in Farmer's Decisions." American Journal of Agricultural Economics 56(1974):14-25.

Just, R. and G. Rausser. "Commodity Price Forecasting with Large-Scale Econometric Models and Futures Market." American Journal of Agricultural Economics 63(1981):197-208.

Kutcher, G. "Agricultural Planning at the Regional Level: A Programming Model of Mexico's Pacific Northwest." Programming Studies for Agricultural Sector Policy, R. Norton and L. Solis (eds.). New York: Johns Hopkins Press, 1981.

Lady, G. "Model Assessment and Validation Issues, Structure and Energy Information Administration Program Goals." Gass, 1980.

Lin, W., G. Dean, and C. Moore. "An Empirical Test of Utility vs. Profit Maximization in Agricultural Production." American Journal of Agricultural Economics 56(1974):497-508.

Lins, D., S. Gabriel, and S. Sonka. "An Analysis of the Risk Aversion of Farm Operators: An Asset Profolio Approach." Western Journal of Agricultural Economics 6(1981):15-29.

Lucas, R. "Econometric Policy Evaluation: A Critique." The Phillips Curve and Labor Markets, K. Brunner and A. Meltzer (eds.). New York: North Holland Publishing Co., 1976.

McCarl, B. "Validation of Linear Programming Models." Unpublished paper, Department of Agricultural and Resource Economics, Oregon State University. 1982.

McCarl, B., W. Candler, H. Doster, and P. Robins. "Purdue Top Farmer Automatic Crop Budget: Design, Application and Experience." Canadian Journal of Agricultural Economics 25(1977):17-30.

McCarl, B. and T. Tice. "Linearizing Quadratic Programs Through Matrix Diagonalization." American Journal of Agricultural Economics 62-3 (1980):571-573.

Miller, S. and A. Halter. "Systems-Simulation in a Practical Policy-Making Setting: The Venezuelan Cattle Industry." American Journal of Agricultural Economics 55(1973):420-432.

Musser, W. and K. Stamoulis. "Evaluating the Food and Agriculture Act of 1977 with Firm Quadratic Risk Programming." American Journal of Agricultural Economics 63(1981):447-456.

Payne, J. "Alternative Approaches to Decision Making Under Risk: Moments Versus Risk Dimensions." Psychology Bulletin 80(1973):439-453.

Pieri, R., K. Meilke and T. MacAulay. "North American-Japanese Pork Trade: An Application of Quadratic Programming." Canadian Journal of Agricultural Economics 25(1977):61-79.

Pope, R. "Empirical Estimation and Use of Risk Preferences in Apparaisal of Estimation Methods that Use Actual Economic Decisions." American Journal of Agricultural Economics 64(1982):376-383.

Pope, R. "Supply Response and the Dispersion of Price Expectations." American Journal of Agricultural Economics 63(1981):161-163.

Robison, L. "An Appraisal of Expected Utility Hypothesis Tests Constructed from Responses to Hypothetical Questions and Experimental Choices." American Journal of Agricultural Economics 64(1982):367-383.

Rodriguez, G. and D. Kunkel. "Model Validation and the Philippine Programming Model." Agricultural Economics Research 32(1980):17-25.

Roumasset, J. "Risk Aversion, Indirect Utility Functions, and Market Failure." Risk, Uncertainty and Agricultural Development, J. Roumasset, J-M. Boussard, and I. Singh (eds.) New York: Agricultural Development Council. 1979.

Samuelson, P. "Professor Samuelson on Theory and Realism: Reply." American Economic Review 55(1965):1164-1172.

Sargent, R. "Verification and Validation of Simulation Models." Progress in Modelling and Simulation, F. Cellier (ed.), Academic, London, England (in press).

Schellenberger, R. "Criteria for Assessing Model Validity for Managerial Purposes." Decision Sciences 5(1974):644-653.

Shannon, R. "Simulation: A Survey with Research Suggestions." AIIE Transactions 7(1975):289-301.

Sharpe, W. Portfolio Theory and Capital Markets. New York: McGraw-Hill Book Company, 1970, 316 p.

Singh, I. "Recursive Programming Models of Agricultural Development." Contributions to Economic Analysis: Studies in Economic Planning Over Space and Time, G. Judge and T. Takayama (eds.) Amsterdam: North Holland Publishing Company, 1973.

Zellner, A. "Statistical Analysis of Econometric Models." Journal of the American Statistical Association 74(1979):No. 367.