REVIEW OF MARKETING AND AGRICULTURAL ECONOMICS
VOL. 52, NO. 3 (December, 1984)

Model Validation: An Overview with some Emphasis on Risk Models

Bruce A. McCarl*

Model validation is a topic which has a long history in the literature. This article attempts to focus much of the previous discussion upon agricultural economics models. True model validation is not possible. Nevertheless, model validation exercises improve the relevancy of models and strengthen the theoretical basis for modelling. Procedures are given for conducting validation exercises and for dealing with models which fail such exercises. In addition, comments appear on considerations involved with the validation of risk models.

1.0 Introduction

Model validation according to Gass (1983, p. 611) refers to activities "to establish how closely the model mirrors the perceived reality of the model user/developer team". However this definition is too broad to be useful. A model cannot represent all of "perceived reality," so attention should be narrowed to that part of reality which the model is intended to represent. Further, a model need not mirror the perceived reality perfectly; rather, it needs to abstract reality "adequately" for the model's anticipated use. Thus, model validation[1], for the purposes of this paper, refers to activities designed to determine the usefulness of a model, i.e., whether it is appropriate for its intended use(s); whether the benefits of improving model usefulness exceed the costs; whether the model contributes to making "better" decisions; and possibly how well the particular model 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, when one reviews the published literature one discovers that formal risk model validation has been attempted infrequently. Nevertheless many studies contain strong conclusions relative to potential policy or other usages without assurance regarding model validity. Validation exercises can improve model credibility; provide a "better" platform from which to make policy or operational decision recommendations; contribute to evidence on the usefulness and applicability of theory and modelling methods; provide insights into proper ways of

---

* Bruce A. McCarl is a professor of Agricultural and Resource Economics at Oregon State University. An earlier draft of this paper was presented at the 1983 annual meeting of Regional Research Project S-180, San Antonio, Texas, 28-30 March, 1983, with A. Gene Nelson as a co-author. The author would like to thank L. M. Eisgruber, Scott Hamilton, Wes Musser, Nancy Bockstael and especially A. Gene Nelson for comments on earlier drafts of this paper. Technical paper No. 7427 of the Oregon Agricultural Experiment Station.

[1] The author is aware of the discussion involving verification versus validation appearing in the literature, however, he chooses to use the term validation as all encompassing.

modelling (for example, when preconceived notions fail validation testing this stimulates development of alternative hypotheses); and increase the likelihood of models and their results being used by decision makers.

Model validation has received considerable attention, particularly with respect to simulation models (Sargent 1982; Oren 1981; Mirham 1972; Naylor and Finger 1967; Balci and Sargent 1981). However, there has not been as much attention devoted to validation of the full spectrum of models used in agricultural economics (House 1974, and Gass 1980, 1983 are exceptions). Further, risk model validation has been sparsely treated. Thus, the purpose of this paper is to discuss the issues of model validation separately and in the context of risk modelling, bringing both subjects into a higher position of prominence. The material presented relates to economic model validation (utilizing econometric, mathematical programming and simulation methods) along with a number of comments on risk models. A review and synthesis of the previous literature is used and procedures are drawn together for the validation of agricultural economics models.

The paper is organized into four sections. Following the introduction (section 1), general comments on validation as it applies to models in use are presented in section 2. In section 3 attention is given to procedures for validating models. This is followed by concluding comments in section 4.

2.0 Validation and Modelling Background

2.1 Models: Their Use and Validation

Models can be used for three purposes: (a) structural exploration, (b) prediction, and/or (c) prescription.

Models are used in structural exploration to examine how phenomena enter into the formation of "reality". Such studies attempt to discover the role of various phenomena in influencing economic behaviour. Examples of such uses, involving risk models, are given in Just (1974), Lins *et al.* (1981), Hazell *et al.* (1983) and Feder (1982). Validation is of particular importance in this setting as invalid models lead to improper conclusions on the role of the phenomena studied in influencing decisions and thereby reality[2]. However, in models developed for structural exploration, validation may be possible only by testing predictive ability. An exception would be conventional statistical tests of model fit in an econometric study.

Models are used in prediction to forecast the consequences of decisions. These predictions may involve specific levels of variables (i.e., average income and income risk under a policy) or changes in variables (i.e., how much will income and income risk change if a policy were implemented). Again using risk models as an example, such predictions are intended as analytical input to the policy formation process (as in Hazell and Pomereda

[2] This lack of attention to validation is not uncommon. Gass in an earlier draft of his 1983 paper states "As best can be discerned the topic of model validation has not been of general interest to the OR (modelling) community" and "Papers that describe validation attempts of specific applications are not in plentiful supply."

1981, or Musser and Stamoulis 1981); or as inputs into decisions regarding how to posture oneself in the face of changes in the economic environment (e.g., Barry, Baker and Sanint 1981) and/or changes in the resource base (i.e., budgeting of machinery changes as in Danok, McCarl and White 1980). 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 invalid models; and (c) validation permits statements regarding the model's ability to predict, thereby giving credibility to the model, the modelling approach and the underlying theory[2] (see for example, Robison (1982) where empirical evidence is drawn together regarding the validity of the Expected Utility Theory). This last use is particularly important given the wide variety of models which have been suggested, analytically and/or empirically used, and theoretically and/or empirically criticized.[3]

The prescriptive use of models involves the solution of specific decision maker problems. 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 will 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 problem.

The following discussion concentrates on the validation of models to be used for structural and predictive purposes. Few agricultural economics models are used for prescription. Models are most often used in either a comparative statics sense to predict the income or area consequences of external changes or in a structural exploration sense to somehow discover important factors. Even when models are used with decision makers, most seem to be used in comparing alternatives and giving the direction of change arising from those alternatives.

2.2 Forms of Validation

Validation exercises vary widely. For example, 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 empirical knowledge)." At the other end of the spectrum, models may be subjected to a continuing series of formal testing procedures, or model results.

---

[2] Samuelson (1965) argues that in effect theories without empirical validity should be rejected. This certainly argues for validation tests. In turn such tests often can be construed as tests of theory.

[3] Taking risk models as an example, many have been proposed. For example, in using mathematical programming of choice decisions under net income risk, an individual could use game theory (Agrawal and Heady 1972), E-V models (Freund 1956), MOTAD (Hazell 1971), marginal risk constrained linear programming (Chen and Baker 1974), single or multiple index models (Sharpe 1970), lexicographic utility (Lin, Dean and Moore 1974), 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 (1982).

The ultimate validation test would involve observing whether an available model is used for its intended purpose. Clearly this is not always possible. Evidence of model validity is often required for a model to be considered for use. Moreover it is time consuming and expensive to give models such as "trial by fire." Consequently, validation most often involves a mixture of *a priori* tests and continuing evaluation.

In practice model validation procedures can consist of efforts designed to either *technically* and/or *operationally* validate a model. Technical validation, as we will use the term, represents validation of a model for a particular (single) use and embodies (a) testing of assumptions and data, (b) testing of technical equation logic, and (c) testing of predictive-prescriptive ability. Operational validation represents validation of a model as it is implemented for its intended uses and as it is used, 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 utilized depends on the costs of the validation process versus the benefits of the validity information derived.

2.3 A Generalized Modelling Process

A general view of modelling is useful to facilitate the discussion of validation procedures across the spectrum of agricultural economics models. Models to be validated can involve, for example, such diverse settings as econometric models used to study structure, mathematical programming models used to study policy alternatives, Monte Carlo simulation models used to study decision consequences, and deterministic simulation models used to project system behaviour.

Modellers follow virtually the same process in any modelling exercise. A modeller begins with a perception of problem then, using theory and knowledge of the problem, derives a modelling approach. Subsequently data are collected and the model relationships fitted. These model relationships may be fitted formally in the case of econometrics or less formally with averages or "reasonable" values in simulation and programming models. Some studies, such as econometric structural exploration studies, may stop once the relationships are fitted. However, when doing a more complete modelling exercise, the next phase is model construction in which the empirical relationships are integrated into a unified empirical model; then a base case model is developed. Experiments are designed; model experimentation carried out; and model results obtained. Validation occurs throughout this process.

This generalized modelling process 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.* 1972; Zellner 1979), simulation (Johnson and Rausser 1977; Anderson 1974; Shannon 1975; Mirham 1972; Oren 1981; Balci and Sargent 1982), and mathematical programming (Kutcher 1983; McCarl 1982). Also Gass (1980, 1983) gives a comprehensive overview and references. Finally note that directly estimated econometric models are not extensively treated here. These are usually judged valid or invalid by well known statistical criteria.

2.4 Issues Regarding Validation of Any Model

Those undertaking validation exercises should consider a number of issues. The most fundamental issue relates to the outcome of the model validation process. *Models can never be validated, only invalidated.* Testing of a model in one or several of its uses does not guarantee that it will perform satisfactorily in all uses. 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. The outcome is not a valid model. Nevertheless we will use the term "valid model" to denote a model which has not failed validation tests.

Another fundamental issue is the subjectivity of the validation process. There is not, and never will be, a totally objective and accepted approach to model validation. Model validation is subjective in many ways. Modellers subjectively choose the uses to validate, what to validate, the tests to validate with, the data to serve as a basis of comparison, and criteria to measure the validity/invalidity of the model. Thus, the statement "the model was judged valid" can mean almost anything. (See Anderson (1974) for more on subjectivity).

There are a host of attendant considerations to those above. Summarizing them briefly:

1. The types of models used in agricultural economics research are generally not models which deal with phenomena where immutable natural laws are present. Consequently, one cannot say that validity has been found (House 1974).

2. Numerical validation exercises normally use historical data. Such exercises obviously embody the assumption that the future will be like the past (House 1974). However, 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 be implemented which change the structure of the situation invalidating the model (Lucas 1976). Consequently, validity cannot be shown using historical data.

3. Models may be judged valid improperly or judged invalid improperly, allowing the possibility of both type I and type II errors. An important distinction in the validation process is that the models must be judged valid or invalid for its specific uses. Thus, models should be tested only within the bounds anticipated for the model's intended uses. Similarly, models which are judged valid for some, but not all, anticipated uses should be qualified as to their validity (See Brink and McCarl (1979) for an example).

4. A model does not have to be valid in order to be used. Undoubtedly models have been used which would not pass validation tests. Preoccupation with validation may discredit useful models. This observation leads to a question: To what extent is validity required for a model to be useful? Models which are invalid predictors of specific numerical values may be valid (i.e., useful) if they predict the right direction and magnitude of change. Models may also be useful if they properly identify outcomes that will change and those which will not, even if they are poor predictors of the magnitude of change. Further, a model may be valid if it provides insight, not quantitative information (Geoffrion 1976).

5. Models may be judged valid for the wrong reasons due to specification error. For example, a programming model may be forced to validate using such devices as flexibility constraints, but the presence of these constraints may limit the usefulness of the model for its purpose. Similarly validating the model over historical data which were used in model construction may lead to excessive confidence (Howrey and Kelejian 1971).

6. Models once validated are not valid forever, rather they must be continually validated in an operational sense.

2.5 Issues Regarding Validation Of Risk Models

The above general discussion does not cover problems specific to risk models for use in agricultural economics. 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 the distribution of reality is stochastic the possibility 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 probability distribution of the model's output corresponds to the probability distribution of reality. This implies that models of stochastic phenomena need a more complete definition of reality than do other models. In other words, more data are needed. Ackoff (1962) and Mihram (1972) present discussions of this topic.

2. Decision makers do not know all the parameters (i.e., weather) at the time decisions are made when dealing with stochastic realities. Consequently, decisions are made considering the probability distribution of the parameters. Model validation may be difficult if the data used do not exhibit the decision maker's subjective probability distributions at the time the decision was made. This requires even more data. Further, stochastic parameter outcomes can alter the feasibility or desirability of plans causing adaptive decision making. Thus, validation may require information which changes over time.

3. In validating risk models, particularly those used for structural exploration, profit maximizing behaviour may be confused with risk averse behaviour thus incurring specification error (for example, see the cases in Pope 1981; Roumasset 1979; and Baker and McCarl 1982).

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

3.0 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. Additionally, approaches to validation differ depending on whether one is attempting technical or operational validation. The discussion below largely relates to technical validation, although some brief comments on operational validation will be offered later.

3.1 Validation by Assumption

Validation may be done by assumption, by results, or by both assumptions and results. Virtually all models go through a validation by assumption phase, which involves a combination of several different approaches. Models may be judged valid through expert opinion, antecedent, theory, data or logical structure. Validity through expert opinion is manifest in the statement, "Based on the modelling team's expert opinion the model is proper." Antecedent validity is called on when the modelling team says, "This model was used because it was used before." Antecedent validity may also involve reliance on other previous validation efforts[5]. Theoretical validity is called on when the modelling team states, "Based on theory the model relationships are proper, and therefore the model is valid." Data validity is present when the modelling team says, "The procedures utilized to estimate the 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 *assumed*, not tested. Unfortunately far too many models, perhaps most risk models, are judged valid by assumption.

3.2 Validation by Results

Validation by results involves four steps. First, a parameter-output set must be assembled with which the model results are compared. Second, specific validation tests must be conceptualized. Third, a procedure for measuring degree of association including a criterion for accepting or rejecting the statement "the model is close enough" should be chosen

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

Fourth, procedures for dealing with models which fail validation tests often are required.

Space limitations preclude 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.

3.2.1 Parameter-Outcome Sets

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 be validated using data which were not used in model construction.

2. The data themselves must be subjected to validation tests (a point made and expanded on by Gass (1983)). Inconsistent data virtually guarantee the failure of validation tests (for example, when the "observed" output is inconsistent with the "observed" levels of the decision variables, and associated parameters).

3. The data needed for validation should be as comprehensive as possible with observations present for all decision variables, output measures, usages of resources and prices for both product and resources. 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 and 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 ten per cent increase in production would require 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, to the extent possible, should represent the model in use. While this statement is obvious, it has important implications. Data sets may be selected which appear appropriate but which 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.

7. The parameter-outcome sets will be of predominantly historical origin. However, subjective and/or current data may also be employed.

3.2.2 Test Design and Implementation

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

Table 1: A Procedure for Model Validation.

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

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 failed.

(a) If the model has "failed," discover why. Programming models may be either unbounded or infeasible. Simulation models may exhibit numerical difficulties or may incorporate equations which cannot feasibly represent the particular variable values. Repair the model and go to Step 2; otherwise to go Step 6.

(b) If the model has a solution, then utilize association measures to discover the degree of correspondence between the outcome set and the model solution. These measures should be applied to all possible output variables, and imputed prices. Aggregates, such as income and total land area, should be examined. Go to Step 4 if the measures indicate a sufficient degree of association. Go to Step 5 otherwise.

Step 4. Prepare to 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. Consider whether (a) the "reality" parameter-outcome set data are consistent and correctly calculated, (b) the data are properly entered into the model structure, and (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 consider whether the model needs to be revised, discarded or qualified. 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 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 the validation exercise with qualifications).

---

[6] It has been argued that complex models must be validated in smaller components first, then perhaps later subjected to overall validation. Validation of the whole model should not be overlooked (i.e., a model is not necessarily valid even though each individual equation fits well).

Table 2: Technical Validation Examples.[a]

Validation Test	Type of Model: Econometric Simulation	Type of Model: Non Econometric Simulation	Type of Model: Mathematical Programme
Plausibility	Rich (1981) Lins *et al.*[b] (1981) Just[b] (1974) Helmberger and Akinyosye (1984)	Anderson (1971) Miller and Halter (1973)	Delgado and McIntire (1982)
Possibility	[e]	[e]	Kutcher (1983)
Supply Function	[e]	[e]	Kutcher (1983) Rodriguez and Kunkel (1980)
Dual Supply Function	[e]	[e]	Kutcher (1983) Rodriguez and Kunkel (1980)
Prediction	Just and Rausser (1981) Basiotis *et al.* (1983)	Miller and Halter (1973) Herath, Hardaker and Anderson (1982)	Barnett, Blake and McCarl (1982) Kutcher (1983) Brink and McCarl[c] (1979) Hazell and Pomerada (1981)
Predictive Change	[d]	[d]	Hazell, *et al.* (1983) Brink and McCarl[c] (1979) Amos and Timmons (1983)
Predictive Tracking	Just and Rausser (1981) Lee and Bui-Lan (1982) Reynolds and Gardiner (1980)	Jones and Brockington (1971) Singh (1972)	Pieri, Meilke and MacAulay (1977)

[a] This table contains recent examples of both risk-free and risk-including models which have been validated using the various tests.

[b] There is doubt whether 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 of one model against another.

[d] Special case of Predictive Tracking.

[e] Appears possible, however, is most amenable to use with mathematical programming.

3.2.2.1 *Plausibility Test.* This test examines whether the model creates "plausible" results. The test involves sensitivity analysis to determine whether model results under situation conform to the modelling team's biases. The model is run with a particular parameter change, and the resultant outcome changes are examined for plausibility. Formal association tests are not done. Models failing to give plausible results are either invalid or the perceptions of "plausible" results may be improper. The plausibility test is the most commonly used validation-by-results test and, along with the validation-by-assumption statements, represents the extent to which many, perhaps most, models are validated.

3.2.2.2. *Possibility Test.* This test examines whether it is possible for the model to duplicate a "reality" situation. The test involves fixing the parameters and decision variables at the parameter-outcome set levels and examining the model results. 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.

3.2.2.3 *Supply Function Test.* This test examines the intersection of the internal model supply function and an inelastic demand function, i.e., whether the model marginal cost of production is close to the observed price (Kutcher 1983). This test is executed by first constraining production to equal the output levels. In turn, one observes the (a) dual variables, or (b) changes in total cost for small variations in quantity. Subsequently, one compares the observed cost changes and market prices to see if the criterion of competitive price equals marginal cost can and does hold. A model failing this test may still be acceptable if the modelled entity (a) does not behave competitively, or (b) expected prices are different from those received.[7]

3.2.2.4 *Dual Supply Function Test.* This test examines the intersection of the model supply function and an infinitely elastic demand function, testing whether production at the observed price is close to observed production. Here, prices are fixed at the expected prices, and output compared with actual output. Failures of this test would indicate a divergence between the rule of produce until price equals marginal cost and reality.[8]

3.2.2.5 *Prediction Test.* This test examines the ability of the model to predict outcomes, when specified with parameters "identical" to those leading to that outcome. This is the most common more extensive validation effort. When the model does not predict well, compare the results from the prediction test with those from the possibility test to see whether the reality solution is virtually as good as the prediction test solution (i.e., an alternative optima). Models failing the prediction test when they have passed the possibility test are unsatisfactory predictors of actual levels. However, prediction of changes may be of equal or overriding importance as measured by the next test.

---

[7] There are forms of both this test and the following test relating to demand functions which 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.

[8] This test and the one above were originally proposed by Kutcher (1983) and are most easily implemented within mathematical programming models.

41016-5817-2¶

3.2.2.6 *Predictive Change Test.* Models may not need to predict exactly as long as they predict the magnitude or possibly even direction of change accurately. The predictive change test is implemented by developing two separate parameter-outcome sets, running the model with the parameters of each, then testing the degree of association between the first differences in the model predictions versus those observed in the outcome sets. A model failing such a test would be unsatisfactory for comparative statics usage.

3.2.2.7 *Predictive Tracking Test.* The ability to predict one time change may not be adequate in terms of validation. Rather, the ability of a model to track behaviour over time may be crucial. To test for tracking adequacy, a model is subjected to a historical series of parameter values and the model's degree of association with the associated outcome series compared in terms of either actual levels or changes (first differences). Models failing such tests may be able to be qualified to specific cases where they do track satisfactorily.

3.2.3. Association Measures

Implicit in any validation-by-results exercise is the measurement of the 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 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 1982). Objective measurement of association involves statistical tests between the model predictions and numerical reality.

Statistical association measures for validation have been extensively used by simulators. The literature reviews by Anderson (1974), Gass (1983), Johnson and Rausser (1977), Mirham (1972), and Shannon (1976) point to a number of these and discuss acceptance criterion. For example, Table 3 is reproduced from Johnson and Rausser (1977, pages following 187) and indicates a wide variety of tests which could be utilized.

Risk models pose 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 compare either the entire probability distribution of "reality" or at least its moments. Second, it is not always possible to obtain the distribution of outputs from the model. This clearly depends upon the modelling technique used. For example, distribution of outputs may be obtained from repeatedly simulated models and from models such as MOTAD (which produce a simulated income distribution), whereas models such as chance constrained programmes do not produce such information.

3.2.4 Dealing with Invalid Models

When a model is judged invalid during validation testing, then the modeller has one of four options. First, and most obviously, the model can be discarded. Second, the validation test results could be judged improper and the results ignored. This may be appropriate if 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

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

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

Source: Johnson and Rausser (1977, pages following 187).

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 (Brink and McCarl 1979). Models may also need to be qualified in terms of ranges of data over which they are applicable. Fourth, the model could be repaired. The purpose of this section is to discuss two issues: How can a model be repaired?[9] and examples of cases when models cannot be repaired.

Those confronted with "invalid" models should remember that models are abstractions of reality and, even after a competent modelling job, can be inadequate abstractions. During the modelling 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 fails a validation test, assumptions may need to be repaired.[10] Modellers should question whether there are important causal factors that have been omitted which should be included; whether a different model solution algorithm should be used; or whether the validation test comparison is consistent with anticipated model usage. For example, in a large sector model, one may accept invalid subregional crop area allocations 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 modelling a fallow-wheat system one could be looking at the long-run and ignoring the short-run constraints imposed by the areas that were in fallow or in production last year. In the face of such difficulties, one must consider 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, attempting to validate on long-run averages rather than short-run situations. Finally, one may not be able to find appropriate outcome sets. For example, a model built to study the long-run effects of new developments, such as the recent U.S. Payment In Kind program, may not be able to be validated due to a lack of relevant "reality" data.

3.3. Towards 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 are more complex than those involved with technical validation especially in a risk modelling framework. However, if models are ever to be routinely used, they must be operationally valid.

---

[9] A difficult situation, which will not be discussed, involves the notion of repeated validation tests using the same data and "repaired" models. Perhaps one should reserve two sets of data for validation testing. One for initial testing and model repair and the second for testing once a model is judged "valid." One needs to be careful in terms of both validating a model on data from which it is built and validating the model on data from which it has been extensively "repaired."

[10] An attendant possibility is that the algorithm is working improperly.

In an operational context, validation exercises examine model performance across uses. In this sense operational validation involves a series of technical validations and more interaction with model users. Operational validation is important to give the model credibility. Operational validation can proceed by some combination of assumptions and results. Validation-by-results procedures involve the same considerations (parameter-outcome sets, tests, criteria, and dealing with invalid models) discussed previously. However, issues arise with respect to input data, decision maker validation efforts and output.

Validation of input data is difficult for models which are used in diverse situations, as new data are potentially needed for every application. Validation by assumption is often called on (for example, Fishburn and Balch (1974) have invented "Super Genie" which guarantees that elicited utility functions will be proper). Input questionnaires may be necessary, and the modellers need be concerned with questionnaire design so that meaningful data are obtained. Rigid models again impose their own special difficulties. Data on subjective probabilities are of particular concern in operational validation. Hogarth (1975) argues that because man is a "selective, sequential information processing system with limited capacity, he is ill equipped for assessing probability distributions." It is difficult to test whether the elicited probabilities are valid representations of true beliefs regarding the future. Feedback and evaluation procedures can be designed to compare actual occurrences with the decision makers' probabilistic predictions, for example, using scoring rules (Bessler and Moore 1979). Altman (1980) discusses a number of other difficulties regarding data development.

The extent to which nontechnical 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 credibility of these results will depend on the credibility of the models. 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 modelling features which permit model users to conduct their own validation exercises (see McCarl *et al.* (1977) for a discussion of a model which permits model user validation).

The output presentation format should 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 to how well it is predicted and to how it is presented. Conrath (1973) reveals that the format in which results are presented influences decision making. Payne (1973, p.440) argues, based on psychological studies, that "the way in which sources of information are displayed affects their utilization."

One must also be concerned with the questions: Does the model adequately represent the decision problem and the major variables to be considered for each specific use? Is the structure of the model appropriate? Will the model continue to perform satisfactorily over time and across situations? Does the model easily adapt to the alternate anticipated usages?

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 can be compared with reality. This is a demanding validation process, and has been implemented occasionally in agricultural economics. For example, Just and Rausser (1981) examined several econometric models over time.

4.0 Concluding Comments

The material presented and reviewed above yields several conclusions with respect to model validation. First, models should be validated in terms of use. Conversely, model use, in part, determines validation procedures. Second, validation is difficult, especially for risk models. Third, model validation should be a high priority research area. Fourth, and finally, more attention should be placed on risk model validation.

Amplifying on these comments, first take validation in the context of use. It appears obvious that models should be validated in terms of their uses. Nevertheless, this has only been an implicit theme within the validation literature. Further, most validation exercises have been technical, not operational, and most statements about model validity have not been qualified with respect to use. The validation in use concept has major implications for model validation exercises. Validation in use (1) allows models to be validated only in terms of selected outputs; (2) leads to models being qualified in terms of their use; and (3) implies that more effort and attention should be devoted to operational validation.

Turning attention to the difficulty of validation, note that true validation cannot be achieved. This is inevitable since agricultural economics models generally depict behaviour, not phenomena governed by invariant natural laws. Additionally, the historical and/or subjective data used in validation exercises can never be unequivocally judged representative of all model usages. Thus, modellers must be satisfied with acceptance of the null hypothesis that a model is not invalid. Furthermore, validation exercises involve subjective choices of items validated, procedures and association tests. The situation is further complicated in risk models and/or models of stochastic phenomena since observed "reality" is really a random draw from a probabilistic reality.

The difficulties above not withstanding, increased attention to validation is important and needed. Validation exercises will improve the quality and relevancy of modelling. Samuelson (1965) comments that the validity of the theory rests in large part on its ability to be consistent with empirical fact. Therefore, validation is extremely important in research. Carefully, competently done validation exercises will not only improve model usefulness but will also strengthen the theoretical foundations for model building. Systematic attention to validity can be achieved following procedures such as those suggested by Shannon (1975, p.294) who states "the greatest possible validity is achieved by:

(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."

The tests described in this review provide additional validation procedures.

Finally, regarding the conclusion on risk model validation. Validation of risk models has received little attention. This poses many potential difficulties. Agricultural economists are faced with many alternative risk models which could be used, yet there is little empirical evidence about which kinds of risk models should be used. In addition, a sizable body of theory exists in risky decision making. However, much of this theory is not validated, in part, because few attempts have been made to validate risk models. Researchers should spend more time validating risk models.

References

ACKOFF, R. L. (1962), *Scientific Method: Optimizing Applied Research Decisions*, New York: John Wiley and Sons.

AGRAWAL, R. C. and HEADY, E. O. (1972), *Operations Research Methods for Agricultural Decisions*, Ames: Iowa State University Press.

ALTMAN, S. (1980), "Pitfalls of data analysis", in G. MAJONE and E. QUADE (eds.), *Pitfalls of Analysis*, New York: John Wiley and Sons.

AMOS, O. M. and TIMMONS, J. F. (1983), "Iowa crop production and soil erosion with cropland expansion", *American Journal of Agricultural Economics* 65 (3), 486-92.

ANDERSON, J. R. (1971), "Spatial diversification of high-risk sheep farms", in J. B. DENT and J. R. ANDERSON (eds.), *Systems Analysis in Agricultural Management*, New York: John Wiley and Sons.

————, (1974), "Simulation: methodology and applications in agricultural economics", *Review of Marketing and Agricultural Economics* 43 (1), 3-55.

BAKER, T. G. and MCCARL, B. A. (1982), "Representing farm resource availability over time in linear programs: a case study", *North Central Journal of Agricultural Economics* 4 (1), 59-68.

BALCI, O. and SARGENT, R. G. (1981), "Bibliography on validation of simulation models", *Newsletter TIMS College on Simulation*, New York, July.

BARNETT, D. A., BLAKE, B. and MCCARL, B. A. (1982), "Goal programming via multi-dimensional scaling applied to Senegalese subsistence farms", *American Journal of Agricultural Economics* 64 (4), 720-7.

BARRY, P. J., BAKER, C. B. and SANINT, L. R. (1981), "Farmers credit risks and liquidity management", *American Journal of Agricultural Economics* 63 (2), 216-27.

BASIOTIS, P., BROWN, M., JOHNSON, S. R. and MORGAN, K. J. (1983), "Nutrient availability, food costs and food stamps," *American Journal of Agricultural Economics* 65 (4), 685-93.

BESSLER, D. A. and MOORE, C. V. (1979), "Use of probability assessments and scoring rules for agricultural forecasts", *Agricultural Economics Research* 31 (1), 44-7.

BRINK, L. G. and MCCARL, B. A. (1979), "The adequacy of a crop planning model for determining income, income change, and crop mix", *Canadian Journal of Agricultural Economics* 47 (1), 13-25.

CHEN, J. T. and BAKER, C. B. (1974), "Marginal risk constraint linear program for activity analysis", *American Journal of Agricultural Economics* 56 (3), 622-7.

CONRATH, D. (1973), "From statistical decision theory to practice, some problems with the transition", *Management Science* 19 (12), A873-A883.

DANOK, A. B., MCCARL, B. A. and WHITE, T. K. (1980), "Machinery selection modelling: incorporation of weather variability", *American Journal of Agricultural Economics* 63 (4), 700-8.

DELGADO, C. L. and MCINTYRE, J. (1982), "Constraints on oxen cultivation in the Sahel", *American Journal of Agricultural Economics* 64 (2), 188-96.

DHRYMES, P. J., HOWREY, E. P., HYMANS, S. H., KMENTA, J., LEAMER, E. E., QUANDT, R. E., RAMSEY, J. B., SAPIRO, H. T. and ZARNOWITZ, V. (1972), "Criteria for evaluation of econometric models", *Annals of Economic and Social Measurement* 1 (3), 291-323.

FEDER, G. (1982), "Adoption of interrelated agricultural innovations: complementarity and the impacts of risk-scale, and credit", *American Journal of Agricultural Economics* 64 (1), 94-101.

FISHBURN, P. C. and BALCH, M. (1974), "Subjective expected utility for conditional primitives", in M. BALCH, D. McFADDEN, and S. WU (eds.), *Essays on Economic Behaviour Under Uncertainty*, Amsterdam: North Holland Publishing Company.

FREUND, R. J. (1956), "The introduction of risk into a programming model", *Econometrica* 24 (2), 253-63.

GASS, S. I. (ed.) (1980), *Validation and Assessment Issues of Energy Models*, Proceedings of a Workshop, National Bureau of Standards Special Publication 564, Washington, D.C.

————(1983), "Decision-aiding-models — validation, assessment, and related issues", *Operations Research* 31 (4), 603-31.

GEOFFRION, A. M. (1976), "The purpose of mathematical programming is insight, not numbers", *Interfaces* 7 (1), 81-92.

HAZELL, P. B. R. (1971), "A linear alternative to quadratic and semivariance programming for farm planning under uncertainty", *American Journal of Agricultural Economics* 53 (1), 53-62.

————(1982), "Application of risk preference estimates in firm-household and agricultural sector models", *American Journal of Agricultural Economics* 64 (2), 384-93.

HAZELL, P. B. R. and POMAREDA, C. (1981), "Evaluating price stabilization schemes with mathematical programming", *American Journal of Agricultural Economics* 63 (3), 550-6.

HAZELL, P. B. R., NORTON, R., PARTHASARTHY, M. and POMAREDA, C. (1983), "The importance of risk in agricultural planning models", in R. NORTON and L. SOLIS (eds.), *The Book of Chac: Programming Studies for Mexican Agriculture*, New York: John Hopkins Press, pp. 225-49.

HELMBERGER, P. G. and AKINYOSYE, V. O. (1984), "Competitive pricing and storage under uncertainty with an application to the U.S. soybean market", *American Journal of Agricultural Economics* 66 (2), 119-30.

HERATH, H. M. G., HARDAKER, J. B. and ANDERSON, J. R. (1982), "Choice of varieties by Sri Lanka rice farmers: comparing alternative decision models", *American Journal of Agricultural Economics* 64 (1), 88-93.

HOGARTH, R. (1975), "Cognitive processes and the assessment of subjective probability distributions, *Journal of American Statistical Association* 70 (350), 271-94.

HOUSE, P. W. (1974), "Diogenes revisited — the search for a valid model", *Simulation* 23 (4), 117-25.

HOWREY, E. P. and KELEJIAN, H. H. (1971), "Simulation versus analytical solutions: the case of econometric models", in T. H. NAYLOR (ed.), *Computer Simulation Experiments with Models*, New York: John Wiley and Sons.

JOHNSON, S. R. and RAUSSER, G. C. (1977), "Systems analysis and simulation: a survey of applications in agricultural and resource economics", in L. R. MARTIN, (ed.), *A Survey of Agricultural Economics Literature, Vol. 2*, Minneapolis: University of Minnesota Press.

JONES, J. G. W. and BROCKINGTON, N. R. (1971), "Intensive grazing systems", in J. B. DENT and J. R. ANDERSON (eds.), *Systems Analysis in Agricultural Management*, New York: John Wiley and Sons.

JUST, R. E. (1974), "Investigation of the importance of risk in farmer's decisions", *American Journal of Agricultural Economics* 56 (1), 14-25.

JUST, R. E. and RAUSSER, G. C. (1981), "Commodity price forecasting with large-scale econometric models and futures markets", *American Journal of Agricultural Economics* 63 (2), 197-208.

KUTCHER, G. P. (1983), "Agricultural planning at the regional level: a programming model of Mexico's pacific northwest", in R. M. NORTON and L. SOLIS (eds.), *The Book of Chac: Programming Studies for Mexican Agriculture*, New York: Johns Hopkins Press.

LEE, B. M. S. and BUI-LIN, A. (1982), "Use of errors of prediction in improving forecast accuracy: an application to wool in Australia", *Australian Journal of Agricultural Economics* 26 (1), 49-62.

LIN, W., DEAN, D. W. and MOORE, C. V. (1974), "An empirical test of utility vs. profit maximization in agricultural production", *American Journal of Agricultural Economics* 56 (3), 497-508.

LINS, D. A., GABRIEL, S. C. and SONKA, S. T. (1981), "An analysis of the risk aversion of farm operators: an asset portfolio approach", *Western Journal of Agricultural Economics* 6 (1), 15-29.

LUCAS, R. (1976), "Econometric policy evaluation: a critique", in K. BRUNNER and A. MELTZER (eds.), *The Phillips Curve and Labor Markets*, New York: North Holland Publishing Company.

McCARL, B. A. (1982), Validation of linear programming models. Unpublished paper, Department of Agricultural and Resource Economics, Oregon State University.

McCARL, B. A., CANDLER, W. V., DOSTER, H. and ROBBINS, P. (1977), "Purdue top farmer automatic crop budget: design, application and experience", *Canadian Journal of Agricultural Economics* 25 (1), 17-30.

MILLER, S. R. and HALTER, A. (1973), "Systems-simulation in a practical policy-making setting: the Venezuelan cattle industry", *American Journal of Agricultural Economics* 55 (3), 420-32.

MIRHAM, G. A. (1972), "Some practical aspects of the verification and validation of simulation models", *Operational Research Quarterly* 23 (1), 17-29.

MUSSER, W. N. and STAMOULIS, K. G. (1981), "Evaluating the Food and Agriculture Act of 1977 with firm quadratic risk programming", *American Journal of Agricultural Economics* 63 (3), 447-56.

NAYLOR, T. A. and FINGER, J. M. (1967), "Verification of computer simulation models", *Management Science* 14 (2), B92-B101.

OREN, T. I. (1981), "Concepts and criteria to assess acceptability of simulation studies: a frame of reference", *Communications of the ACM* 24 (4), 180-9.

PAYNE, J. (1973), "Alternative approaches to decision making under risk: moments versus risk dimensions", *Psychology Bulletin* 80 (6), 439-53.

PIERI, R. G., MEILKE, K. D., and MACAULAY, T. G. (1977), "North American-Japanese pork trade: an application of quadratic programming", *Canadian Journal of Agricultural Economics* 25 (1), 61-79.

POPE, R. D. (1981), "Supply response and the dispersion of price expectations", *American Journal of Agricultural Economics* 63 (1), 161-3.

———, (1982), "Empirical estimation and use of risk preferences in appraisal of estimation methods that use actual economic decisions", *American Journal of Agricultural Economics* 64 (2), 376-83.

REYNOLDS, R. G. and GARDINER, B. (1980), "Supply response in the Australian sheep industry: a case for disaggregation and dynamics", *Australian Journal of Agricultural Economics* 24 (3), 196-209.

RICH, M. M. (1981), "The policy implications of possible New Zealand wheat pricing schemes", *Australian Journal of Agricultural Economics* 25 (3), 266-77.

ROBISON, L. J. (1980), "An appraisal of expected utility hypothesis tests constructed from responses to hypothetical questions and experimental choices", *American Journal of Agricultural Economics* 64 (2), 367-83.

RODRIGUEZ, G. and KUNKEL, D. (1980), "Model validation and the Philippine programming model", *Agricultural Economics Research* 32 (1), 17-25.

ROUMASSET, J. (1979), "Risk aversion, indirect utility functions, and market failure", in J. ROUMASSET, J-M. BOUSSARD, and I. SINGH (eds.), *Risk, Uncertainty and Agricultural Development*, New York: Agricultural Development Council.

SAMUELSON, P. A. (1965), "Professor Samuelson on theory and realism: reply", *American Economic Review* 55 (5), 1164-72.

SARGENT, R. G. (1982), "Verification and validation of simulation models", in F. CELLIER (ed.), *Progress in Modelling and Simulation*, London, England: Academic Press.

SHANNON, R. E. (1975), "Simulation: a survey with research suggestions", *AIIE Transactions* 7 (5), 289-301.

SHARPE, W. F. (1970), *Portfolio Theory and Capital Markets*, New York: McGraw-Hill Book Company.

SINGH, I. (1972), "Recursive programming models of agricultural development", in G. G. JUDGE and T. TAKAYAMA (eds.), *Contributions to Economic Analysis: Studies in Economic Planning Over Space and Time*, Amsterdam: North Holland Publishing Company.

ZELLNER, A. (1979), "Statistical analysis of econometric models", *Journal of the American Statistical Association* 74 (367), 628-51.