copyright 1997 Bruce A. McCarl and Thomas H. Spreen

CHAPTER XVIII: VALIDATION OF PROGRAMMING MODELS

Model validation is important in any empirical analysis[1]. Programming models frequently are superficially validated. However, validation is necessary for both predictive and prescriptive model use[2]. Validation exercises almost always improve model performance and problem insight.

This chapter presents procedures for programming model validation and cites examples. The discussion will be most relevant to predictive model validation, however, the procedures may also be used with prescriptive models.

18.1 Background

Before beginning the presentation, a model structure is needed. Let the model contain demand(X), production(Y) and input purchase variables(Z) with the following structure.

$$
\begin{array}{llllll}
Max & f(X) & & -g(Z) & & \\
s.t. & X & -DY & & \leq & 0 \\
 & & GY & -Z & \leq & 0 \\
 & & AY & & \leq & b \\
 & X, & Y, & Z & \geq & 0
\end{array}
$$

Let us denote the optimal values of these variables as X^*, Y^*, Z^*. Now suppose these variables are assumed to correspond to real world observations $\overline{X}, \overline{Y}$, and $\overline{Z}$. The model also has associated shadow prices, U, V, and W which at optimality are U^*, V^*, and W^* and correspond to real world observations $\overline{U}, \overline{V}, \overline{W}$.

---

1 The material in this chapter is largely drawn from McCarl (1984) and McCarl and Apland.

2 The word validate is controversial. Some prefer to use verify. Within this text, validate refers to exercises determining whether the model user and or modeling team feels the model behavior is close enough to real world behavior.

copyright 1997 Bruce A. McCarl and Thomas H. Spreen

18.2 General Approaches to Validation

Validation approaches vary widely. The overall purpose is to test how well a model serves its intended purpose. For predictive models, validation tests can involve comparing model predictions to real world results. For prescriptive models, decision maker reliance is the ultimate validation test. Unfortunately, these tests can rarely be used because they are expensive and time-consuming (this is often the reason for modeling in the first place). Thus, models are frequently validated using historical events. Although a model may have a broad range of potential uses, it may be valid only for a few of those uses. The validation process usually results in identification of valid applications.

Model validation is fundamentally subjective. Modelers choose the validity tests, the criteria for passing those tests, what model outputs to validate, what setting to test in, what data to use, etc. Thus, the assertion "the model was judged valid" can mean almost anything (See Anderson; and House and Ball for elaboration). Nonetheless, a model validation effort will reveal model strengths and weaknesses which is valuable to users and those who extract information from model results.

Two validation approaches may be used: validation by construct and validation by results. Validation by construct asserts the model was built properly therefore it is valid. Validation by results refers to exercises where the model outputs are systematically compared against real world observations.

18.3 Validation by Construct

Validation by construct is always used in modeling, but it is also the end of most of the programming model validation exercises. Validation by construct, as the sole method of validation, is justified by one of several assertions about modeling .

The right procedures were used by the model builder. Usually this involves the assertion that the approach is consistent with industry, previous research and/or theory; and that the data were specified using reasonable scientific estimation or accounting procedures (deducing the model data from real world observations).

Trial results indicate the model is behaving satisfactorily. This arises from a nominal examination

copyright 1997 Bruce A. McCarl and Thomas H. Spreen

of model results which indicates they do not contradict the modeler's, user's, and/or associated "experts" perceptions of reality.

Constraints were imposed which restrict the model to realistic solutions. Some exercises use constraints to limit adjustment possibilities and force the model to give results very close to historically observed outcomes. The application of "flexibility" constraints (Day and Cigno; Sahi and Craddock [1975, 1974]) as in the recursive programming example is such an approach.

The data were set up in a manner so that the real world outcome had to be replicated. In some models one can assure replication of a real world outcome through the model structure and data calculation procedures. This approach is manifest in input-output modeling (Leontief, 1936) where procedures insure that the base solution will always arise. A similar approach has appeared in price endogenous programming applications (see Miller and Millar; Fajardo et al.).

Fundamentally, validation by construct suffers from the shortcoming that validation of a particular model is assumed, not tested. If a model plays an integral part in a study, going forth with a model that is only assumed valid does not appear to be totally satisfying. However, validation by construct is a necessary precursor to any validation by results testing.

18.4 Validation by Results

Validation by results involves comparison of model solutions with real world outcomes. Models used in such a comparison will always have been built relying on experience, precedence, theory, appropriate data estimation and measurement procedures. Thus, validation by construct will always precede validation by results[3]. Testing whether the model output reasonably reproduces real world results is the next validation step. That determination involves five phases: first, a set of real world outcomes and the data causing that outcome is gathered; second, a validation experiment is selected; third, the model is set up with the appropriate data, the experiment is implemented and a solution is generated; fourth, the degree

[3] We do not recommend imposing constraints or model structure to force validation unless they are absolutely necessary and certainly not until initial testing has been done

copyright 1997 Bruce A. McCarl and Thomas H. Spreen

of association between model output and the real world outcome is tested; and, finally, a decision is made regarding model validity. Comments relative to these steps appear below.

18.4.1 Parameter Outcome Sets

Data describing a real world observation contain both the values for the model input parameters and model output measures. Thus, when predicting corn acreage one needs the prices, costs and resource endowments that led to the acreage decision. Tests of the model beyond the original data set will generally be more representative of model accuracy in applications (Anderson and Shannon). While complete input parameter-outcome sets are most desirable, partial sets giving aggregate measures (giving total corn acreage - not acreage by planting date) can be useful.

18.4.2 Validation Experiments

A set of validation experiments is described below. These experiments are not mutually exclusive; rather they are a set of sequential experiments which should be performed (or at least considered) in a given order. Five general validation experiments will be presented: a feasibility experiment, a quantity experiment, a price experiment, a prediction experiment, and a change experiment.

18.4.2.1 Feasibility Experiment

The feasibility experiment has primal and dual forms. The basic idea involves setting up the model equations with the variables held at their observed levels, then examining solution feasibility.

The primal test involves addition of the constraints:

$$\begin{aligned} X &= \overline{X} \\ Y &= \overline{Y} \\ Z &= \overline{Z} \end{aligned}$$

This experiment tests internal model consistency. Often, the feasibility experiment is neglected in favor of, for example, seeing if the model can replicate X, Y, Z. However such a solution can never be replicated if it is not feasible. The feasibility experiment often determines needed data, data calculation or model structure revisions. Such an experiment also finds errors arising due to faulty model equation specification.

The dual feasibility experiment involves testing whether the observed shadow prices are feasible in

copyright 1997 Bruce A. McCarl and Thomas H. Spreen

the dual or the Kuhn-Tucker conditions. For the example above this involves seeing whether:

$$\begin{array}{rcccccl} U & & & & & \leq & f'(X) \\ -UD & + & VG & + & WA & \leq & 0 \\ & & -V & & & \leq & -g'(Z) \\ U & & & & & = & \overline{U} \\ & & V & & & = & \overline{V} \\ & & & & W & = & \overline{W} \end{array}$$

is feasible[4]. This procedure tests whether the solution is dual feasible and therefore primal optimal. Non-zero variables in the observed outcome should, because of complementary slackness, lead to equality Kuhn-Tucker conditions. Zero variables should ordinarily be associated with strict inequalities. Careful execution of this experiment quite often reveals inadequacies in structure, data, or the objective function. Again, there is the attendant possibility of an inconsistent "real world outcome" which requires correction.

The data requirements of these feasibility conditions are rather strong -- they assume knowledge of a complete solution. Often, one may know output and input levels (X, Z) and aggregate sums of production variables (sums within Y) but not individual variable values. Thus, tests involving totals may be in order. Second, the experiments may require artificial variables to both allow and help find infeasibilities as discussed in the last chapter.

18.4.2.2 Quantity Experiment

The quantity experiment involves constraining the outputs supplied or inputs demanded at their actual levels and removing f(X) or g(Z) then observing the shadow prices. The output variant (developed by Kutcher) involves adding the constraint

$$X = \overline{X}$$

with the objective function term f(X) dropped. The correspondence of $\overline{Y}, \overline{Z}, \overline{U}, \overline{V}, \overline{W}$ and the shadow prices on the equation pegging the X value can then be tested.

Such a test examines the consistency between the optimal and observed levels of the production (Y)

---

4 Note the ' sign of f(x) denotes the first derivative.

copyright 1997 Bruce A. McCarl and Thomas H. Spreen

and input supply (Z). In addition, the imputed values of the resources (V and W) may be examined for consistency with the observed values. Further, the dual values associated with the quantity constraints ($X = \overline{X}$) should be sufficiently close to the market price for the outputs.

These shadow prices give an indication of the marginal cost of production at the observed quantities (Figure 18.1). This procedure is a test of the economic assumption of perfect competition (Kutcher) since, under perfect competition, the shadow price should equal the market price.

The input version of the test is essentially identical. The model is augmented by the constraint $Z = \overline{Z}$ with the g(Z) term dropped from the objective function. In this case, the experiment should generate dual variables which can be compared to prevailing market prices of inputs (Figure 18.2).

18.4.2.3 Price Experiment

A third type of model validation experiment is the price experiment. This type of experiment is relevant in price endogenous models or models with fixed demand requirements. This experiment involves fixing the objective function coefficients at existing real world prices (U, W), then observing quantities (the dual of the quantity experiment). The optimal quantities (X^*, Z^*) are then compared to the observed levels (X, Z). The output price experiment is illustrated in Figure 18.3 where the fixed output price is equated with the model supply schedule to get a value of X^*. One may also examine how implicit fixed resource values are influenced in the experiment.

18.4.2.4 Prediction Experiment

The prediction experiment is the most common validation by results test. Examples can be found in Barnett, et al. (1982); Brink and McCarl (1979); and Hazell and Pomareda. The prediction experiment involves fixing the problem data at real world values and solving to get X, Y, Z. In turn we test whether the linear programming model output is close enough to the real world outcomes.

18.4.2.5 Change Experiment

The prediction experiment is to some degree the ultimate validation experiment in that it tests whether a model can replicate reality. However, most programming models are used for comparative statics

copyright 1997 Bruce A. McCarl and Thomas H. Spreen

analysis. This implies a need for an experiment which tests whether models accurately predict change.

To test a model's ability to predict change, one must have data on two real world situations and the resultant model solutions. Then, a comparison is made between the change in the model solution variables (e.g.,X_1^*, X_2^*) and the change observed in the real world solution X_1, X_2 as done in Hazell et al. (1981).

<u>18.4.2.6 Tracking Experiment</u>

Even a model which satisfactorily predicts a one time change may not be adequate. One may wish to track adjustments through time. For validation of such applications the model can be solved using a series of parameter sets. The focus of the validation would then be on how well the model "tracks" over time with respect to the corresponding observed adjustments in the system. Again, comparisons are made between changes in the model solution and observed changes in the real world solution (for example, see Pieri et al.).

<u>18.4.2.7 Partial Tests</u>

The above experiments are discussed for a model as a whole. Obviously, in any particular validation exercise, it may be desirable to perform experiments with some of the variables fixed at real world levels with other variables left unconstrained -- an attempt to validate portions of the model. Often, this type of experiment will be necessary with large models because observations on all decision variables and/or shadow prices may not be readily available. Validation experiments may then be performed to require sums of variables to equal observed real world levels, for example.

18.4.3 Employing a Validation Test

There are several identifiable stages for conducting one of the experiments given the model.

Step 1. Alter the model variables, equations and data to reflect the validation experiment.

Step 2. Solve the model(s).

Step 3. Evaluate the solution(s). Is it infeasible, unbounded, or optimal?

(a) If the model solution is infeasible, examine the results to find the cause of infeasibility. Use the artificial variable based method in the last chapter. Once the cause is found, go to Step 5.

copyright 1997 Bruce A. McCarl and Thomas H. Spreen

(b) If the model is unbounded, use the large upper bound method from the last chapter. Once the cause is found, go to Step 6.

(c) If the solution is optimal, perform association tests (as discussed below) to discover the degree of correspondence between the "real world" and the model solutions (except for the feasibility experiment). These tests should be conducted upon both the primal and dual variables.

Step 4. If the model variables exhibit a sufficient degree of association, then:

(a) do higher level validation experiments, or

(b) determine the model is valid and proceed to use it.

Step 5. If the model does not pass the validation tests, consider whether:

(a) the data are consistent and correctly calculated,

(b) the model structure provides an adequate representation of the real world system, and

(c) the objective function is correctly specified.

Step 6. Fix the Model --Procedures for recalculating model parameters will be problem specific. If, for example, all the variables have been fixed at "real world" levels and infeasibilities occur, then the units of the observed input parameters and outputs may be inconsistent. If the data are accurate and model structure problems are suspected, one should consider whether: errors have been made in constructing the matrix; additional constraints are needed; or such factors as risk and/or willingness to adjust (i.e., flexibility constraints) should be entered into the model. If the model has been respecified either structurally or through its data, proceed back to Step 3 and repeat the validation test. If not, go to Step 7.

Step 7. If the preceding steps do not lead to a valid model, one must decide whether to:

(a) do demonstrations with an invalid model -- assuming this is an approximately correct structure,

(b) abandon the project, or

copyright 1997 Bruce A. McCarl and Thomas H. Spreen

(c) limit the scope of validation to a lesser set of variables (aiming at a less strict level of validation), subsequently qualifying model use. This may happen in many cases due to some considerations discussed subsequently.

18.4.4 Evaluation Criteria for Validation

Association tests can be used to measure whether a set of model results are similar to observed results. Quite a number of association tests are available as reviewed by Shannon; Anderson; Gass (1983); or Johnson and Rausser, for example. These tests have been well presented elsewhere and their theoretical roots are well outside our scope, so only a brief discussion will be given.

Regression techniques have been used to measure the association of model solutions with observed values (for examples see Nugent; Rodriguez and Kunkel). In that case, model results are regressed on observed values with perfect association indicated by an intercept of zero and a slope of one. The Theil U test has also been used (Leuthold; Pieri, et al.). This is a nonparametric "goodness of fit" test. Garret and Woodworth suggest the use of the G Index for validation -- a procedure for comparing sets of basic variables (an example can be found in Keith). Simple measures such as means, sums, mean absolute deviations, and correlation coefficients, have been used (Nugent; Kutcher; Hazell, et al., 1981). The authors have not found applications of Kolmogrow-Smirnov, Chi Squared or various other "goodness of fit" tests in a programming context. However, these techniques have been applied in simulation settings (see Anderson; Johnson and Rausser; Shannon; and Gass).

18.5 What if the Model Does not Validate

From a practical standpoint, models do not always pass validation tests. Since models always involve many assumptions, failure to validate, likely indicates that improper assumptions have been used. Consequently, when models fail validation tests, modelers often ask: What assumptions should be corrected?

As discussed above, programming models embody assumptions about both mathematical structure and the model structure. The mathematical structure assumptions involve additivity, divisibility, certainty,

copyright 1997 Bruce A. McCarl and Thomas H. Spreen

and proportionality. These assumptions, when severely violated, will cause validation tests to fail. The model designer then must consider whether these are the cause of failure. If so, the use of techniques such as separable, integer, nonlinear, or stochastic programming may be desirable to construct a new model.

Modeling assumptions may also lead to validation test failure. These assumptions embody the correctness of the objective function, variables, equations included, coefficients, and equation specification. Programming algorithms are quite useful in discovering assumption violations. Given an optimal solution, one may easily discover what resources were used, how they were used, and their marginal values. Thus, when presented with an invalid solution, resource usage and resource valuation should be investigated. Models are most often invalid because of inconsistent data, bad coefficient calculation, bad equation specification, or an incorrect objective function. Thus, common fixes for a model failing validation involve data respecification and/or structural corrections.

copyright 1997 Bruce A. McCarl and Thomas H. Spreen

When dealing with linear programming, there are several other properties which can lead to validation failures. An optimal LP solution is characterized by the term basic, i.e., no more activities can be in the model than the number of constraints. For example, if a disaggregated regional model is constructed with a single constraint in each region, at most one activity will be produced in each region (if other constraints are not present in the model). This is ordinarily inconsistent with real world performance. Models then may be judged invalid because they overspecialize in production due to the nature of basic solutions. Several approaches may be taken when faced with this sort of inadequacy in a model solution. First, one may be satisfied with validating only aggregate results and not worrying about individual production results. Second, one may constrain the model to the observed solution and investigate whether this solution is an alternative optimal solution (which, as argued by Paris, may commonly occur). Third, one may recognize that a basic solution will not validate and enter constraints that limit the adjustment process of the activities within the model (flexibility constraints (Day) or aggregation procedures (Onal and McCarl [1991, 1989]) as discussed in the price endogenous chapter). Fourth, the model may be expanded by including risk considerations. Fifth, one may feel the model is structurally inadequate in that many of the factors that constrain production may be inadequately portrayed in the model (see the arguments in Baker and McCarl). Such a situation leads to either one of two fixes: more constraints can be added or the activities within the model may be respecified so they represent feasible solutions within omitted constraints as in the price endogenous chapter (Onal and McCarl [1991, 1989]).

Models may also fail validation because of the objective function. Specification of the constraints identifies the set of possible solutions, while the objective function determines the single optimal solution. Thus, the objective function must be carefully specified and reviewed (with the dual feasibility test used if possible). Finally, the objective function may generate alternative optimal solutions, one of which is the desired solution (see Paris or Burton et al. for discussion).

Another phenomena may cause models to fail validation tests. Operations, quite often, are performed over several time periods. An annual model depicting operations of this type may well be invalid

copyright 1997 Bruce A. McCarl and Thomas H. Spreen

because it ignores initial conditions or does not recognize that parameter expectations may change over time. Thus, unless the model has initial conditions identical to those in the "real world," it may be very difficult to validate.

18.6 Comments

Validation is an important concern within any programming exercise. A well validated model will have gone through both validation by construct and validation by results phases. Unfortunately, true validation will never occur as models can only be proved invalid. However, through satisfactory completion of the above experiments, the level of satisfaction may be increased.

The ultimate test of validity deals with adoption of the model by the decision maker. Satisfactory validation via the procedure given may not be sufficient for acceptance. A numerically valid model may solve the wrong problem and thus, will never be valid from the decision maker's viewpoint. Clearly, under these circumstances, validation in the broadest sense is only achievable by redefining the model so it takes into account the true problem.

copyright 1997 Bruce A. McCarl and Thomas H. Spreen

References

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

Barnett, D., B. Blake, and B. McCarl. "Goal Programming via Multidimensional Scaling Applied to Senegalese Subsistence Farms," American Journal of Agricultural Economics, 64,4(1982):720-7.

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.

Burton, R.O., J.S. Gidley, B.S. Baker and K.J. Reda-Wilson. "Nearly Optimal Linear Programming Solutions: Some Conceptual Issues and a Farm Management Application." American Journal of Agricultural Economics. 69(1987):813-818.

Day, R. Recursive Programming and Production. North Holland Publishing Company, Amsterdam. 1963.

Day, R. and A. Cigno, editors. Modeling Economic Change: The Recursive Programming Approach. Amsterdam: North-Holland Publishing Co., 1978.

Fajardo, D., B. McCarl, and R. Thompson. "A Multicommodity Analysis of Trade Policy Effects: The Case of Nicaraguan Agriculture." American Journal of Agricultural Economics, 63,1(1981):23-31.

Garret H. and R. Woodworth. Statistics in Psychology and Education. New York: David McKay Co., Inc., 1964.

Gass, S.I. "Decision-Aiding Models: Validation, Assessment, and Related Issues for Policy Analysis." Operations Research, 31(1983):603-31.

Hazell, P. and C. Pomareda. "Evaluating Price Stabilization Schemes with Mathematical Programming." American Journal of Agricultural Economics. 63,3(1981):550-6.

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.

Henderson, J. "The Utilization of Agricultural Land: A Theoretical and Empirical Inquiry. " Review Economics and Statistics, 41(1959):242-59.

Hillier, F. and G. Lieberman. Introduction to Operations Research. San Francisco: Holden-Day, Inc., 1967.

House, P. and R. Ball, "Validation: A Modern Day Snipe Hunt: Conceptual Difficulties of Validating Models," in S.I. Gass (ed.), Validation and Assessment of Issues in Energy Models, Proceedings of a Workshop, National Bureau of Standards, Spec. Pub. 564, Washington, D.C.;1980.

Johnson, S. and G. Rausser, "Systems Analysis and Simulation: A Survey of Applications in Agricultural and Resource Economics," in L. Martin (ed.), A Survey of Agricultural Economics Literature, 2(1977):157-301.

copyright 1997 Bruce A. McCarl and Thomas H. Spreen

Keith, N. "Aggregation in Large-Scale Distribution Systems." Ph.D. thesis, Purdue University, 1978.

Kutcher, G., "A Regional Agriculture Planning Model for Mexico's Pacific Northwest." in R. Norton and L. Solis (eds.), Programming Studies for Mexican Agricultural Policy, Chapter 11, New York: Johns Hopkins Press, 1983.

Leontief, W. "Quantitative Input-Output Relations in the Economic System of the United States," Review of Economic Studies, 18(1936):105-25.

Leuthold, R.M., "On the Use of Theil's Inequality Coefficients." American Journal of Agricultural Economics, 57,3(1975):344-6.

McCarl, B.A. "Model Validation: An Overview with Some Emphasis on Risk Models." Review of Marketing and Agricultural Economics. 52,3(1984):153-173.

McCarl, B. "Cropping Activities in Agricultural Sector Models: A Methodological Proposal." American Journal of Agricultural Economics, 64,4(1982): 768-72.

McCarl, B.A. and J.D. Apland. "Validation of Linear Programming Models." Southern Journal of Agricultural Economics. 68,5(1986):155-164.

McCarl, B. and T. Spreen. "Price Endogenous Mathematical Programming as a Tool for Sector Analysis." American Journal of Agricultural Economics, 62,1(1980):87-102.

Miller, T. and R. Millar, "A Prototype Quadratic Programming Model of the U.S. Food and Fiber System." mimeo, CES-ERS-USDA and Department of Economics, Colorado State University, 1976.

Nugent, J., "Linear Programming Models for National Planning: Demonstration of a Testing Procedure." Econometrica. 38(1970):831-55.

Onal, H. and B.A. McCarl. "Exact Aggregation in Mathematical Programming Sector Models." Canadian Journal of Agricultural Economics. 39(1991):319-334.

Onal, H. and B.A. McCarl. "Aggregation of Heterogeneous Firms in Mathematical Programming Models." European Journal of Agricultural Economics. 16,4(1989):499-513.

Paris, Q. "Multiple Optimal Solutions in Linear Programming Models." American Journal of Agricultural Economics. 63,4(1981):724-7.

Pieri, R., K. Meilke, and T. MacAuley, "North American-Japanese Pork Trade: An Application of Quadratic Programming." Canadian Journal of Agricultural Economics, 251977):61-79.

Rodriguez, G. and D. Kunkel, "Estimation of Flexibility Coefficients for Recursive Programming Models--Alternative Approaches." American Journal of Agricultural Economics, 56,2(1974):344-50.

Sahi, R. and W. Craddock. "Estimating Crop Acreages in the Prairie Provinces: Application of Recursive Programming." Canadian Journal of Agricultural Economics. 23(1975):1-16.

copyright 1997 Bruce A. McCarl and Thomas H. Spreen

__________. "Estimation of Flexibility Coefficients for Recursive Programming Models - Alternative Approaches." American Journal of Agricultural Economics. 56(1974):344-50.

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

copyright 1997 Bruce A. McCarl and Thomas H. Spreen

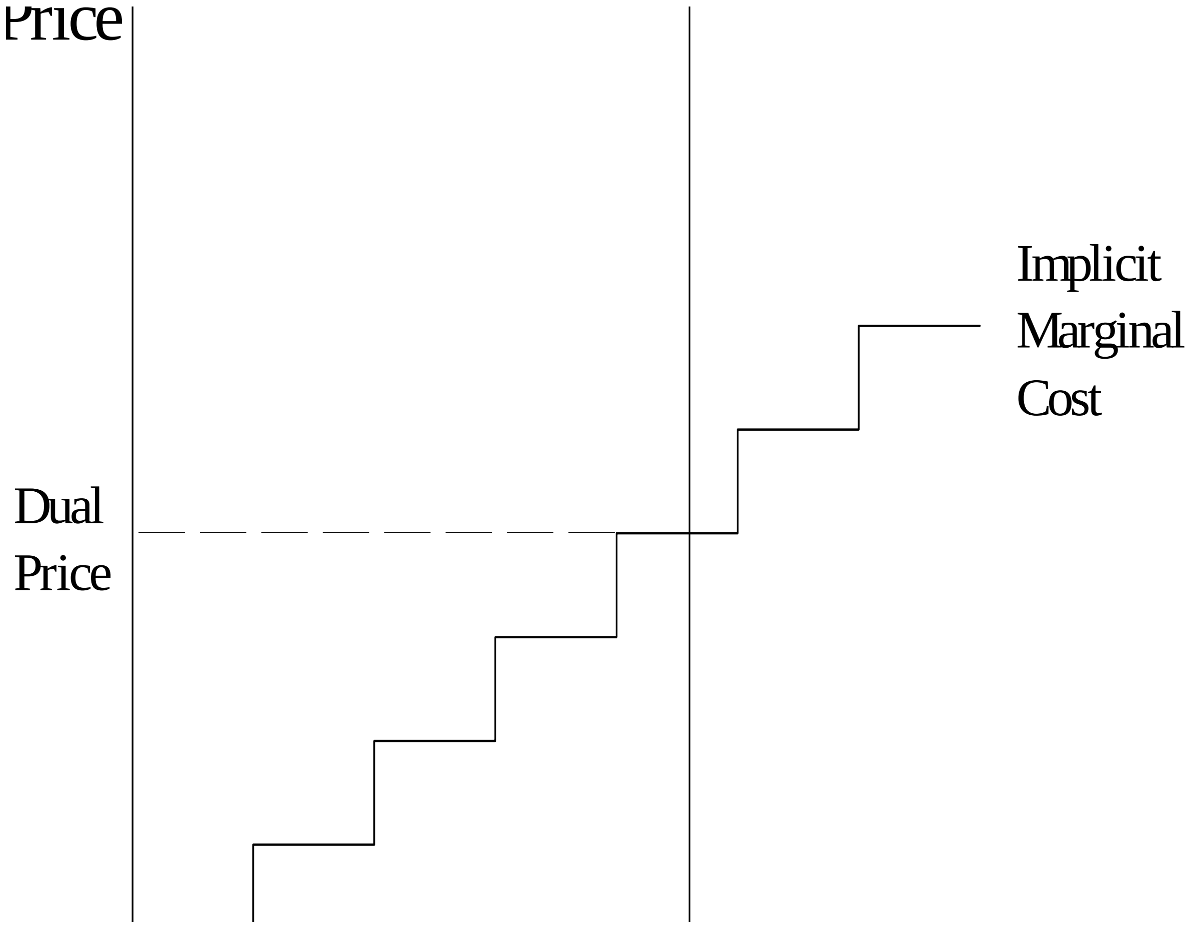
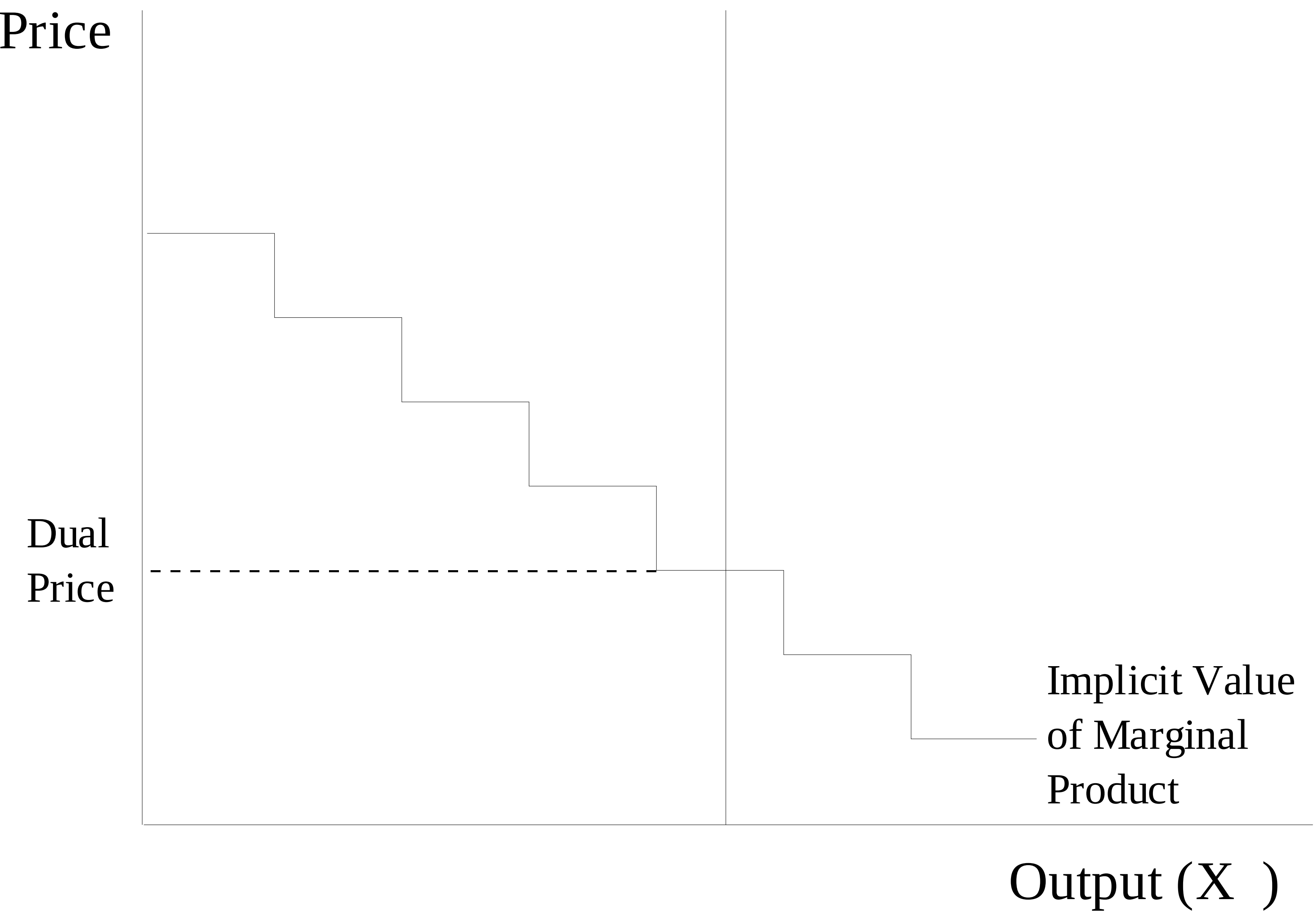


Figure 18.1 Output Quantity Experiment

copyright 1997 Bruce A. McCarl and Thomas H. Spreen



copyright 1997 Bruce A. McCarl and Thomas H. Spreen