1984

Pseudo Data: A Tool for Teaching Production Economics

Thomas W. Hertel

and

Lance McKinzie*

Education

AAEA paper presented at its annual meetings, 1984

*Assistant Professor and Research Associate, respectively, Department of Agricultural Economics, Purdue University, West Lafayette, Indiana 47907.

Pseudo Data: A Tool for Teaching Production Economics

by Thomas W. Hertel and Lance McKinzie

Abstract

It is argued that a "laboratory" data set would greatly facilitate the teaching of graduate-level production economics. Development of a process model and an optimal experimental design for generating pseudo data are outlined. A translog, multiproduct profit function is estimated and the resulting net and gross elasticities are discussed.

I. INTRODUCTION

Recent developments in duality theory and the concept of flexible functional forms has led to a resurgence of interest in production economics. Whether the topic is factor substitution, income distribution, technical change, economies of scale, or any of the other traditional problems in production theory, the current literature draws heavily on these new methods. While they have been available for over a dozen years (e.g., Diewert, 1969), widespread application and incorporation of these methods into graduate production economics curricula is much more recent.

As is generally the case, more sophisticated methods have left greater room for misapplication. Thus, the challenge in teaching this newer material to students in production economics lies not only in conveying the theoretical concepts, but also in teaching their responsible application. When is it appropriate to apply these more sophisticated methods, and what can go wrong when they are applied? The traditional approach to this type of teaching challenge has been to give students an empirical problem to work with. In the case at hand, this would involve giving them a data set with which to estimate (e.g.) a profit function, which could in turn be interpreted and perhaps criticized. A second-best alternative might be to assign a set of empirical articles to be read and evaluated.

Unfortunately, many of the data sets in use are not "well-behaved", i.e., estimation of a dual function using these observations does not result in a set of parameters which satisfy the required neoclassical restrictions. This problem is particularly severe in the multiproduct setting (e.g., Shumway, 1983). Strictly speaking the duality results do not apply to these ill-behaved functions, and we are left with something which cannot be readily interpreted. This problem may, or may not, be acknowledged by the authors of journal articles. In most cases the amount of information provided is insufficient for readers to check for themselves whether or not the function is well-behaved.

When a profit function is found to be ill-behaved, explanations generally turn to problems with the data set. Poor quality data and excessive aggregation (over commodities and/or firms) are commonly cited sources of difficulty. Another potential pitfall is that the underlying behavioral axioms (e.g., profit maximization or cost minimization) may not be satisfied. In some cases the latter are posed as testable hypotheses (Appelbaum, 1978), although they are generally maintained. Studies which attribute poor results to one of these causes can serve a valuable purpose in graduate courses, illustrating the point that not all research succeeds. However, they do not provide students with an adequate feeling for what the methods _are_ good for.

Consistently attributing bad results to poor data encourages a certain cynicism and sometimes sloppiness on the part of students who resign themselves to the fact that "this never works out in practice anyway". What is needed is a "laboratory" data set which permits teachers to abstract from data deficiencies, thus enabling students to focus attention on the method -- how it is used, and what its strengths and weaknesses are. Of course such a data set could also be selectively "disrupted" (e.g., via inappropriate aggregation) to illustrate the potential damage which can result.

There are several properties which a laboratory data set should satisfy. These include:

(i) The underlying technology is sufficiently well-understood to permit formulation of preliminary hypotheses. These should in turn facilitate interpretation and discussion of the estimated model.

(ii) The behavioral axioms guiding producers are known, and conform with neoclassical postulates.

(iii) The observed data is accurate and not aggregated.

(iv) Price variability is sufficient to permit measurement of all distinct substitution effects.

This paper reports on the development of a "pseudo" data set which meets (i)-(iv), and illustrates its use in the estimation of a flexible, multiproduct profit function. While some of our results are informative in their own right, their greatest value has been in the teaching of a graduate-level production economics course. Section II discusses the linear programming model from which the pseudo data are generated, based on the experimental design outlined in Section III. The next section details the estimation and interpretation of a translog, multiproduct profit function, while Section V provides a summary and conclusions.

II. THE VEHICLE: A PROCESS MODEL

The process model used to generate the pseudo data is a modified version of the Purdue Crop Budget Model (B-9). It is among the most extensively validated of all process models, having been used daily by extension and research staff, graduate and undergraduate students, as well as by thousands of midwest farmers over the course of its 15-year evolution. It is a linear programming formulation of a profit maximizing farm firm. The formulation utilizes highly detailed information including the farm's machinery working rates, available time for working in the field during different periods of the production year, and cultivation practices.

Timing of production activities is given particular attention in the B-9 model. Expected crop yields are generally acknowledged to decline as planting (and harvesting) of the crop are delayed. However, it is not economical to maintain the necessary machinery set to plant (harvest) all of the crop at one time. The B-9 model captures tradeoffs between the cost of larger, more expensive machinery sets and the benefits associated with improved yields due to timeliness of planting and harvesting. The latter also serves to promote diversification among crop outputs. While corn is often the most profitable crop to be planted during late April and early May, soybeans may be the preferred alternative in late May. This occurs since soybean yields decline at a slower percentage rate than do corn yields as planting is delayed.

As currently formulated, B-9 takes a farm's machinery complement, available full-time labor, and drying and storage capacities as fixed. Decision variables to be optimized determine the crop output mix. In order to accommodate the long-run situation, where capital, labor, drying, and storage are continuously variable inputs, modifications have been made. (In abstracting from the discrete nature in which machinery must be employed we assume the existence of an active rental market.) Land is held fixed since constant returns to scale would otherwise imply an unbounded solution.

The continuously variable machinery inputs deserve additional comment. There are eight different machinery choice variables -- four sizes of combines and four sizes of machinery complements which include all other machinery. The four sizes correspond roughly to 200, 600, 800, and 1200 acre operations. The different working rates for each machinery complement and combine are taken from Edelman (1981). A typical solution for machinery might be .3 units of machinery complement #2 plus .5 units of machinery complement #3. These solution values are weighted by the respective annualized costs for each machinery set (Leatham and Baker, 1982), to arrive at a dollar value for machinery.

Both input requirements and expected yields depend upon which crop was grown on the land in the previous year. There are significant economies from rotating corn and soybeans. Costs rise for corn grown continuously on the same land. Yields decline for both continuous corn and continuous soybean crops. These phenomena are not modeled in B-9. Hence, optimal cropping patterns prescribed by solutions to that model do not accurately reflect the true costs over time from extreme shifts between these two crops. The effect of including the complementarities arising from crop rotation is to give the product transformation curve for corn and soybeans more curvature in the region of equal acreages. Other things equal, a greater change in relative prices is required to achieve a given amount of substitution between these crops. Rotation corn-soybeans was added to the model as an additional crop alternative with greater yields and lower fertilizer and chemical inputs, compared to continuous cropping (<u>Farm Planning and Financial Management</u>, 1980).

III. GENERATING PSEUDO DATA

The concept of pseudodata was introduced into the economics literature by Griffin (1977a, 1977b, 1978) as a means of summarizing the information embodied in industry process models. The resulting cost or profit functions may, in turn, be employed to summarize an individual sector's price responsiveness in large econometric models. Critics of this method (e.g., Madala and Roberts, 1980) point out that the estimated coefficients may be quite sensitive to sample design and the number of basis changes which result from the price configurations chosen. Griffin himself (1982) has noted the sensitivity of his results to the frequency and range over which sample points are selected. It is somewhat surprising then, that none of these authors make reference to the extensive literature on experimental design.

The literature on experimental design draws an explicit link between choice of design and the form of the regression model. Thus, if the objective is to estimate a second-order Taylor series approximation, such as the translog, it is important to choose a compatible experimental design. For ease of reference, consider the following response surface, where three factors (x_1, x_2, x_3) determine the level of the dependent variable (y):

$$(1) \quad y = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 + \alpha_{11} x_1^2 + \alpha_{22} x_2^2 + \alpha_{33} x_3^2 + \alpha_{12} x_1 x_2 + \alpha_{13} x_1 x_3 + \alpha_{23} x_2 x_3.$$

This can be viewed as a second order Taylor's expansion. Furthermore, by redefining $y = \ln \Pi$ and $x_i = \ln p_i$, it becomes a translog profit function.

An important question in designing the experiment is: Over how many levels must each factor (x) be varied (at a minimum)? Two levels for each price is sufficient to estimate α_0, the linear terms and the interaction effects (α_{12}, α_{13}, α_{23}) (Anderson and McLean, p. 353). A two-level factorial for three factors (prices) generates $2^3 = 8$ observations. In order to capture the quadratic effects (α_{11}, α_{22}, α_{33}) we need to go to three levels for each price (Anderson and McLean, p. 353). A full three level factorial generates $3^3 = 27$ observations. The problem with a full three level factorial is that, as the number of factors increases, the number of observations required becomes rapidly unmanageable. Thus, for the eleven price profit function considered here, $3^{11} = 177,147$ data points result.

Fortunately, it is not necessary to carry out a full three level factorial in order to estimate all of the coefficients in the regression model (1). A composite design will permit measurement of both interaction and quadratic effects. This design consists of three groups of observations.

(i) A two-level factorial.

(ii) Points at the extreme of each factor, while at the center of the others.

(iii) The centerpoint itself.

In the three price case (i)-(iii) total 8 + 6 + 1 = 15 observations.

The required number of sample points for a composite design rises much more slowly, as the number of factors increases, than is the case with the three level factorial. Thus, for the eleven price case "only" $2^{11} + (2x11) + 1 = 2{,}071$ sample points are required. This number may be further reduced by utilizing a fractional two-level factorial in part (i) of the composite design. In this paper a 1/16th fractional factorial is employed. This results in $2^{11}/16 = 128$ points for (i). Adding the 22 factor extrema (ii) and the center point (iii) yields a sample of 151 observations.

IV. ESTIMATION OF A TRANSLOG MULTIPRODUCT PROFIT FUNCTION

Based on the pseudo data set generated by confronting our process model with prices from the composite experimental design, we were able to estimate a multi-product translog profit function. The results, based on single equation estimation of this profit function, were compared to those from estimation of a system of supply and demand equations. The extra information (in the form of Hotelling's lemma) embodied in the system approach, resulted in a better approximation of the underlying model.

Sakai (1974) has shown how to decompose the gross price effects, resulting from differentiation of the profit function, into an expansion effect and a pure substitution (compensated) effect. Lopez (1981) has provided a practical means for extracting compensated (net) and gross price elasticities from an estimated profit function. Using these results, and the system estimates of the translog profit function, we have computed the elasticities in Tables 1 and 2.

Compensated Elasticities: Table 1 summarizes our estimates of net elasticities evaluated at the base case (1983 prices). The input-compensated, output supply elasticities are generally less than one, with positive own-price effects and very large cross-price effects. Corn and wheat are found to be net complements in production. That is, an increase in the price of wheat leads to an increase in the optimal supply of corn, and vice-versa. (Remember that land is being held constant throughout this analysis.) This can be explained by focusing the major reason for crop diversification -- namely the timing of production activities. Consider what happens when wheat production increases, in response to improved wheat prices, while input availability is held constant. The shift of land into wheat lessens demands on labor and machinery during the spring and fall. With these resources less constrained, it is profitable to shift land from soybeans to the more input-intensive corn production. Hence the wheat-corn complementarity. On the other hand, corn and soybeans are strong net substitutes. They compete keenly for fixed inputs during the planting and harvesting periods.

Output-constant, input demand elasticities are quite small, indicating that most of the "action" in this model comes from changes in output mix. While there are many different activities with which to produce (e.g.) corn, they involve very similar input mixes. Thus the own-price effects are quite inelastic. The only cross-price elasticity exceeding 0.10 (in absolute value) is the compensated demand elasticity for combines with respect to the price of labor (0.22). This arises due to the fact that, as labor becomes more costly, a larger combine can be purchased to cover the same amount of land in fewer hours.

Table 1. Net Elasticities

QUANTITY	PRICE Corn	Soybeans	Wheat	Labor	Machinery	Combine	Drying	Storage	Fertilizer	Chemicals	Other Inputs
Corn	.5808	-.6929	.1116	0	0	0	0	0	0	0	0
Soybeans	-.7988	1.0104	-.2111	0	0	0	0	0	0	0	0
Wheat	.7302	-1.1983	.4678	0	0	0	0	0	0	0	0
Labor	0	0	0	-.1326	-.0116	.0507	.0030	.0034	.0310	.0120	.0442
Machinery	0	0	0	-.0862	-.1182	-.0083	.0181	.0156	.0772	.0271	.0753
Combine	0	0	0	.2213	-.0049	-.3201	-.0006	-.0096	.0253	.0304	.0592
Drying	0	0	0	.0218	.0177	-.0010	-.1039	.0164	.0220	.0284	.0004
Storage	0	0	0	.0113	.0070	-.0073	.0075	-.0295	-.0149	.0170	.0086
Fertilizer	0	0	0	.0207	.0070	.0039	.0020	-.0030	-.0333	.0090	-.0061
Chemicals	0	0	0	.0198	.0060	.0115	.0065	.0084	.0222	-.0425	-.0319
Other Inputs	0	0	0	.0332	.0076	.0102	.0000	.0019	-.0069	-.0145	-.0316

Table 2. Gross Elasticities

QUANTITY	PRICE										
	Corn	Soybeans	Wheat	Labor	Machinery	Combine	Drying	Storage	Fertilizer	Chemicals	Other Inputs
Corn	1.5055	-1.0772	-.1196	-.0369	-.0087	-.0145	-.0525	-.0349	-.1625	-.0095	.0106
Soybeans	-1.2418	.7895	-.0562	.0625	.0110	.0035	.0426	.0407	.2243	.0400	.0837
Wheat	-.7823	-.3189	1.8754	-.1526	-.0291	.0631	.0261	.0462	-.2975	.0084	-.4387
Labor	.1970	-.2896	.1246	-.1427	-.0149	.0514	-.0043	.0077	.0070	.0261	.0379
Machinery	.3425	-.3768	.1761	-.1105	-.1237	-.0061	.0059	.0136	.0084	.0315	.0398
Combine	.3379	-.0716	-.2252	.2244	-.0036	-.3279	-.0119	-.0265	.0060	.0140	.0853
Drying	2.0268	-1.4279	-.1540	-.0311	.0058	-.0198	-.1745	-.0326	-.2068	.0108	.0049
Storage	.6186	-.6263	-.1252	.0255	.0061	-.0202	-.0150	-.0268	.0075	.0541	.1015
Fertilizer	.5806	-.6952	.1624	.0047	.0008	.0009	-.0191	.0015	-.0806	.0391	.0052
Chemicals	.0837	-.3064	-.0113	.0431	.0070	.0053	.0025	.0269	.0966	.0035	.0491
Other Inputs	-.0426	-.2909	.2686	.0284	.0040	.0146	.0005	.0229	.0058	.0223	-.0338

Gross Elasticities: As expected, when all choice variables are permitted to adjust optimality, the firm's price responses are more elastic. In other words, the gross elasticities shown in Table 2 include an expansion effect, in addition to the previous pure substitution effect. Turning first to the supply elasticities, we find that all of the outputs are gross substitutes. The cross-price effects between corn and soybeans dominate the soybean own-price elasticity, indicating strong substitution possibilities. The gross supply elasticity for wheat is very large, in part due to its relatively small share in the base case. (Wheat is a minor crop in the eastern corn belt.)

Gross input demand elasticities are generally quite close to their output-compensated counterparts in Table 1. However, a few changes merit comment. The own-price elasticity of demand for chemicals is now positive, although not distinguishable from zero. Another notable point is the gross complementarity between fertilizer and drying inputs. This arises from the fact that they are both important inputs in the production of corn. An increase in the price of fertilizer causes a drop in the optimal supply of corn. This in turn dampens the demand for the complementary drying input which is also intensively employed in corn production.

Note that all of the inputs are regressive against soybeans. That is, an increase in the price of any input results in an increase in the optimal supply of soybeans. Symmetrically, an increase in the price of soybeans results in a drop in the demand for any of the inputs. The opposite is true for corn and input demands. These results follow from two facts: (i) total land area is fixed, and (ii) soybeans are relatively less input-intensive than corn.

V. SUMMARY AND CONCLUSIONS

In this paper we have outlined the development and use of a pseudo data set for purposes of exploring the use of a translog, multiproduct profit function. The vehicle used to generate pseudo data is a modified version of the B-9 linear programming model for a representative Indiana farm producing corn, soybeans and wheat. A fractional factorial, composite experimental design was employed in creating the data set. It provides an efficient design for estimation of the parameters in any of the flexible functional forms corresponding to second-order Taylor approximations. Knowledge of the underlying process model permits an extensive discussion of the resulting compensated and uncompensated elasticities.

In sum, there are numerous advantages to using pseudo data as a teaching tool in production economics. In particular, we feel that it enables students to focus on the particular method being taught, learning its strengths and its drawbacks. In abstracting from the inevitable problems of data quality and aggregation it is also hoped that some of the cynicism, which frequently develops with regard to the use of potentially valuable methods, can be avoided.

References

Andersen, V. and McLean. Design of Experiments: A Realistic Approach. New York: Marcel Dekker, 1974.

Appelbaum, E. "Testing Neoclassical Production Theory." J. of Econometrics 7(1978):87-102.

Diewert, W.E. "Exact and Superlative Index Numbers." J. of Econometrics 4(1976):115-145.

Edelman, Mark A., "Alternative Farm Commodity Program Options for the Eighties: An Analysis of Economic Implications with Emphasis on Growth of Indiana Crop Farms," Ph.D. Thesis, Purdue University, 1981.

Farm Planning and Financial Management, ID-68, Cooperative Extension Service, Purdue University, West Lafayette, Indiana, 1980.

Griffin, James M. "The Econometrics of Joint Production: Another Approach." Rev. Econ. & Stat. 59(Nov. 1977a):389-397.

________. "Long-run Production Modelling with Pseudo-data: Electric Power Generation." Bell J. Econ. 8(Spring 1977b):112-127.

________. "Joint Production Technology: The Case of Petrochemicals." Econometrica 46(1978):379-396.

________. "Pseudo-data Estimation with Alternative Functional Forms," in V.K. Smith (ed.) Advances in Applied Microeconomics, Vol. II, JAI Press, 1982.

Lau, L.J. "A Characterization of the Normalized Restricted Profit Function." Journal of Economic Theory 12(1976):131-163.

________. "Applications of Profit Functions," Chapter I.3 and "Testing and Imposing Monotonicity, Convexity, and Quasi-convexity Constraints," Appendix A.4 in Fuss and McFadden (eds.) Production Economics: A Dual Approach to Theory and Applications. Amsterdam: New Holland, 1978.

Leatham, David J. and Timothy G. Baker. "Real Capital Recovery Charges, and Cost-Minimizing Replacement Ages of Tractors and Combines," mimeo, Department of Agricultural Economics, Purdue University, revised draft, March 1982.

Lopez, R. "Estimating Substitution and Expansion Effects Using a Profit Function Framework." Unpublished paper, Department of Ag. Economics, University of British Columbia, 1981.

Sakai, Y. "Substitution and Expansion Effects in Production Theory: The Case of Joint Production," J. Econ. Theory, 9(1974):255-274, 1974.

Shumway, C.R. "Supply, Demand, and Technology in a Multiproduct Industry: Texas Field Crops." Amer. J. Agr. Econ. 65(1983):748-760.