THE HOUSEHOLD PRODUCTION FUNCTION: DATA AND ESTIMATION PROBLEMS

W. Keith Bryant, Cornell University[1]

Data and estimation problems of household production functions are discussed. Both are interrelated and dependent on the structure of the underlying model of the household in which the production function is situated. The household production functions that should be the focus of the research of consumer as opposed to family economists relate to the production of quality and price information, the home production of safety and health, and other functions that undergird consumer policy discussions.

While the concept of a household production function has been widely used in interpreting a variety of household behavior from consumer demand and household time use to marriage, estimates of the parameters of household production functions have been few and far between.[2] There have been three reasons for the relative paucity of such estimates: data problems, estimation problems and the fact that such estimates have not been needed to solve the problems studied. This paper focuses on some data and estimation problems involved in estimating household production functions.

HOUSEHOLD PRODUCTION FUNCTIONS: THE CONCEPT

The concept of a household production is deceptively simple. A production function describes the relationship amongst the inputs and outputs of a production process. For a simple one output-several input production function, it specifies the level of output produced by each combination of inputs used. Furthermore, as used in most economic models, it is a frontier concept in the sense that the function specifies the maximum output produced for each combination of inputs used.

The information required to estimate the parameters of the function depends upon some subset of six items:

a. data on output;
b. data on inputs;
c. data on input prices and households' unearned income;
d. knowledge of the production function's algebraic form;
e. knowledge of the objectives of the households engaged in the production process under study; and
f. the constraints under which they operate.

Simply to focus on input and output data problems is to miss a significant part of the data problems. This is so because the data problems are relative to the algebraic form of the function and to the model in which the household production function is embedded. In consequence, we discuss the latter three issues first.

ESTIMATION POSSIBILITIES AND PROBLEMS

The objectives of households engaged in a production process and the constraints within which they operate determine the range of estimation possibilities open to the analyst. The modern interest in the parameters of household production functions derives from the Beckerian view that households derive satisfaction from household produced and consumed goods and services and that household members' time and purchased goods and services are the inputs into household production. Consequently, I will assume that satisfaction maximization is the objective of households. This reduces the discussion to differences in the constraints households face. Two cases will be discussed.

A. Households Have Choice over the Hours Spent in Paid Employment - Typically Assumed by Labor Economists:

Suppose for simplicity that the households under study are one person households who maximize satisfaction by engaging in two household activities,

(1) $$G = g(X_g, H_g) \text{ and}$$

(2) $$R = r(X_r, H_r),$$

where: G and R are outputs and X_i and H_i ($i = g, r$) are goods and time inputs into the respective processes. We are interested in estimates of the technical parameters of g(.). Good R is viewed as a composite commodity representing all other

[1] Professor and Chair of Consumer Economics and Housing

[2] See: W. Keith Bryant, J. L. Gerner & U. Henze, "Estimating Household Production Functions: A Case Study," in Karen P. Goebel (ed.) Proceedings of the 29th Annual ACCI Conference, March 16-19, 1983, Kansas City, MO, pp. 179-183; M. Sanik and K. Stafford, "Household Production Functions for Food Preparation," in Karen P. Goebel (ed.) Proceedings of the 29th Annual ACCI Conference, March 16-19, 1983, Kansas City, MO, pp. 168-172; and, R. Gronau (1980), "Household Production -- A Forgotten Industry," Review of Economics & Statistics, 62(3):408-416.

household production. Suppose that the household's utility function is:

$$U = u(G, R; P_o, P_u); \tag{3}$$

where P_o is a vector of observed variables that affect household preferences and P_u is a vector of unobserved, exogenous variables that also affect household preferences;[3] the time constraint is:

$$T = H_g + H_r + M, \tag{4}$$

where M = time spent in paid employment; the expenditure constraint is:

$$px(X_g + X_r) = Y \tag{5}$$

where: p_x = market price of purchased inputs, and Y = household income; and, household income is:

$$wM + V = Y, \tag{6}$$

where: w = wage rate at which paid employment is recompensed and V = unearned income. Finally, suppose that households fail randomly to achieve maximum satisfaction given the constraints, either because they make mistakes in choosing the optimal levels of G and R to produce, or because they make mistakes in optimal input use. These mistakes may be termed management error.[4]

Given the assumptions above, the production function can be written as:

$$G_i = g(X_{gi}, H_{gi}) + e_{gi} \tag{7}$$

where e_{gi} = combined effects of the random management error on household i's output of G and the unobserved vector of exogenous variables affecting household preferences.

---

[3] An alternative, and today quite common, treatment of the unobserved variables affecting satisfaction regards them not as preference shifters but as objects of satisfaction on an equal footing in the utility function as R and G. Consequently, rather than taking P_u as given, households choose levels of G, R and P_u that maximize satisfaction. This treatment makes P_u endogenous, not exogenous, and causes statistical problems referred to in the literature as "heterogeneity of preferences." See M. Rosenzweig and K. Wolpin (1980), "Life-cycle Labor Supply and Fertility: Causal Inferences from Household Models," Journal of Political Economy, 88(2):328-348, April, for instance. Since the solution to heterogeneity is the same as for the management error that is assumed below, it is not discussed further.

[4] The assumption of random management error is one of many assumptions that justify the addition of a stochastic error term to what would otherwise be deterministic demand or production functions. The assumption of unobserved exogenous or endogenous variables affecting the preferences of households is another way to justify stochastic error terms.

On the assumption of a specific functional form for g(.) and data on both outputs and inputs, the analyst might incorrectly proceed with its estimation via OLS. The reason it would be incorrect is that the model assumes that households choose levels of X_g and H_g to produce a level of G that will maximize satisfaction along with R. But, the random management error affects not only G but also X_g and H_g: management errors in choosing levels of G that don't maximize satisfaction imply choosing the wrong levels of X_g and H_g to produce G; management errors in choosing the non-optimal combinations of X_g and H_g to produce G will result in non-optimal levels of G. So, regardless of where the management errors occur, observed levels of outputs and inputs will be affected. This means that the e_{gi} will be correlated with X_g and H_g, violating the assumptions of OLS. Without dealing with these correlations, the resulting estimates of the production function can be badly biased.

Given the assumptions, there is an indirect way of estimating the parameters of the household production function. Given the explicit recognition of household production, the utility function is weakly separable in input space. Consequently, one can estimate the derived input demand functions for X_g and H_g conditioned on the level of G the household demands. These determine the least cost amounts of X_g and H_g needed to produce the optimal level of G. These derived input demand functions can be written as:

$$X_g = x(p_g, w, G) + e_{xi} \tag{8}$$

$$H_g = h(p_g, w, G) + e_{hi} \tag{9}$$

where: e_{ji} (j = x, h) = random error component of input j. Since these functions show how households will change their level of input use as input prices and output change, they must describe movement along iso-quants as input prices change and between iso-quants as G changes. Therefore, the coefficients of the functions must be functions of the underlying parameters of the production function. With suitable mathematical manipulation of these functions, estimates of the parameters of the underlying production function can be extracted.

The unwary analyst might again be inclined to estimate (8) and (9) with OLS. But, the same problem of bias arises and for the same reason. Random management mistakes imply that the e_{ji} are correlated with G, thus violating the assumptions of OLS.

A solution to these similar problems, of course, is instrumental variables. If one wants to estimate the production function (i.e., equation (7)) directly, then one can first estimate the household's demand functions for X_g and H_g (i.e.,

$$X_g = d_x(p_g, w, V, P_o) + u_{gi} \tag{10}$$

$$H_g = d_h(p_g, w, V, Po) + u_{hi} \tag{11}$$

where u_{ji} ($j = x, h$) = random error terms)[5] and use the predicted values of X_{gi} and H_{gi}, $\hat{H}_{gi}$ and $\hat{X}_{gi}$, as right-hand side variables in (7). Alternatively, one can estimate the household's demand for G (i.e.,

$$(12) \qquad G_i = d_g(p_g, w, V, P_o) + u_{gi}$$

where u_{gi} = random error term) and use the predicted value, $\hat{G}_i$, in place of G_i in the estimation of (8) and (9).

In sum, given the labor economist's usual assumption that individuals have control over their hours of paid employment, there are two ways of estimating some or all of the parameters of a household production function: directly or indirectly by estimating derived input demand functions. Either way requires data on the vector of preference shifters, the prices of inputs and on unearned household income as well as output and inputs. The data requirements are, therefore, substantial.

B. Households Have No Choice Over Hours of Paid Employment - Typically Assumed by Consumer Economists:

In contrast with the usual assumption made by labor economists that individuals choose the number of hours they spend in paid employment, many consumer economists maintain that hours of work are set by employers or are otherwise out of the control of individuals. This imposes an added constraint (i.e.,

$$(13) \qquad M = M^*$$

where M^* = hours of paid employment set by employer) to the model contained in equations (1) through (7) above. Such an assumption means that the market earnings and, therefore, total household income are exogenous and given to the household. Further, the household allocates the remaining time (i.e., non-market time) and purchased inputs, $X = X_g + X_r$, (that can be purchased with its given income) between the production of G and R.

The direct estimation of the production function, g(.) in equation (7), is made somewhat more difficult by such an assumption. Since M is determined exogenously, the price of H_g is not w[6] and the set of exogenous variables with which to estimate $\hat{X}_g$ and $\hat{H}_g$ (i.e., p_g, V, and P_o) no longer includes w. Further, since Y is now exogenous, it can replace V as a right-hand side variable. Because total household income is usually measured with smaller error than unearned income, V, measurement error bias may be reduced.

Similarly, if one estimates the derived input demand functions, estimation of the instrumental variable, $\hat{G}_i$ from equation (12), involves one fewer right-hand side variable. w is no longer available for use as an exogenous variable because it is no longer a component of the prices of G and R. The market wage rate is equally not available as a regressor in the derived input demand functions because it is not the price of nonmarket time. Fewer regressors but somewhat less measurement error may, therefore, result from the added assumption that paid employment is fixed to the household.

If the assumption is incorrect, then the analyst is denied the use of the wage rate as an exogenous variable that would help to determine the instrumental variables for either input use or output levels. And, because Y is endogenous, its use as a regressor causes statistically biased coefficients. Of course, the question is an empirical one: are individuals' market hours fixed or not? To date, neither labor economists or consumer economists have been convinced by the other's evidence and arguments. I believe that the empirical evidence presented by labor economists is too strong to deny.

Specifying the Form of the Production Function

The implications for data of the algebraic form of the production function are simple: the fewer parameters the function has, the smaller the data set required to estimate it. This point can be made quite simply with the Cobb-Douglas function:

$$(14) \qquad G = \theta H_g^{\beta} X_g^{\alpha}$$

Here in its simplest form, there is one more parameter to be estimated than there are inputs. If household production exhibits constant returns to scale, then $\beta = 1 - \alpha$ and there are the same number of parameters to be estimated as there are inputs in the production process. The data implication is straightforward: under constant returns to scale, data on one input need not be available so long as data exist on its price. Furthermore, if it can also be assumed that all households in the sample face the same price for this input, then neither data on the use of the input nor its price are needed to estimate the parameters of the household production function. Additionally, the problem of endogeneity caused by random management errors evaporates.[7] Strong

[5] Note that these functions are quite different from the derived input demand functions conditional on output. Equations (10) and (11) are the input demand functions unconditioned on output. Because G does not appear as a right-hand side variable in (10) and (11), there is no problem of correlation between the u_{ji} and p_g, w, V, or P_o).

[6] The price of H_g is its opportunity cost in the production of R which is $u_r r_h$ where: u_r = marginal utility of R and r_h = marginal product of time in the production of R. Since $u_r r_h$ presupposes the estimation of both the utility function and the production function for R, searching for an exogenous measure of the "price of H_g" is fruitless.

[7] Assume, for instance, that the production function is Cobb-Douglas and does exhibit constant returns to scale. Then equation (14) can

assumptions can, thus, solve both data and statistical estimation problems. If the assumptions are not warranted, however, they can lead to great error.

But, one need not assume a specific form for the production function. One can, instead, estimate derived input demand functions conditioned on output that are derived from translog cost functions. Such functions are approximations to any production function and some of the production parameters of the underlying production function can be extracted from estimates of their coefficients.[8]

DATA PROBLEMS PER SE

We come, finally, to the data problems themselves. It has previously been said that in order to estimate household production functions, one needs data on inputs and output along with the prices of the inputs, data on unearned income or total family income, and on family preference shifters. This is a tall order and one that no public use data set produced by the U.S. Government meets. Most conspicuous by their absence are data on the time use of family members and on the outputs of household activities. Even the time-use data sets produced by the Survey Research Center lack output data. And, even if such data were available, the analyst then faces the same problems that demand analysts face in not having access to the prices of goods and services. While we may assume that consumers in cross-section face the same prices of goods and services, no consumer economist readily believes it. Furthermore, I am sceptical of the ability of the techniques recently developed by Deaton to extract information on price effects from advanced country, cross-section expenditure samples containing no price data.[9] Finally, the sample selection and missing data problems underlying the data on the price of family members' time use are daunting.[10]

SO WHAT

The conclusion must be that public use data collected by U.S. Governmental agencies contain monumental data problems if they are to be used to estimate household production functions. Are we immediately to plead with BLS, the Census Bureau, the Labor and Agriculture Departments to beef up their data collection efforts so that we can estimate household production functions? Before doing so, we need to decide which household production functions we as consumer economists want to estimate and for what purpose. Intellectual curiosity is not a sufficient reason for lobbying Congress to expand the nation's statistical system. The information on household production relations must be necessary for the solution of pressing consumer policy problems. Three that come immediately to mind are the household production of safety and health, and, to focus on environmental policy, the household production of garbage and trash.

If there are data problems that truly beset consumer economists, they are the data problems that prevent us from saying much that is empirical about the consequences for the consumer of the lack of price and product quality information, the extent of product-related injury and death, the incidence of medical malpractice, the effectiveness of our product liability laws, the effectiveness of deregulation, and the host of other consumer policy issues that confront us. The estimation of household production functions, I suspect, will contribute little to the research on these issues. Until we seriously wrestle with and overcome these data problems, what few consumer policy analyses we are able to conduct will lack empirical cogency, and our promise to solve consumer problems will be largely unmet.

---

be written as

$$(15) \qquad G = \theta H_g^{1-\alpha} X_g^{\alpha}$$

and the analagous derived demand function for H_g conditioned on output is

$$(16) \quad \ln H_g = \ln[A^{-1}(\alpha/(1-\alpha))^{-\alpha}] - \alpha \ln(w/p_x) + \ln G$$

Because of constant returns to scale, the coefficient on $\ln G_i$ is known to be one. $\ln G_i$ can be carried to the left-hand side of the equation, therefore, and subtracted from $\ln H_g$. This avoids the necessity of finding an instrumental variable for G because $\ln G_i$ is no longer a right-hand side variable. Furthermore, assume that households in the sample face the same prices of purchased inputs, r. Such an assumption is common in cross-section analyses. Provided that data on output, G, time input, H_g, and the price of time, w, are available, then the derived input demand function for H_g can be written as

$$(\ln H_{gi} - \ln G_i) = a_o + a_1 \ln w + e_{hi}$$

and estimated by OLS where:

$$\alpha = -a_1; \text{ and}$$

$$A = [\text{antilog } a_o]^{-1}[-a_1/(1+a_1)]^{a1}.$$

[8] See B. R. Beattie and C. R. Taylor (1985), The Economics of Production, New York: J. Wiley & Sons, or R. G. Chambers (1988), Applied Production Analysis, Cambridge: Cambridge U. Press.

---

[9] See A. Deaton (1987), "Estimation of Own- and Cross-Price Elasticities from Household Survey Data," Journal of Econometrics, 36(1-2):7-30, Sept-Oct. His technique assumes large transportation costs between spatial markets and no quality-adjusted price variation within spatial markets. Experience in the US suggests that the reverse is the case.

[10] For a good introduction to these problems see M. Killingsworth (1983), Labor Supply, Cambridge: Cambridge U. Press, pp. 78-100.