SPECIFICATION AND ESTIMATION OF THE DEMAND FOR GOODS WITHIN THE HOUSEHOLD

Mark M. Pitt
Brown University

May 1996

There is a large empirical literature in which household or individual data are used to estimate the demands for both market goods and nonmarket goods, such as health and human capital. For example, Pitt and Rosenzweig (1985) have estimated the effect of food prices and access to health programs and clean water on the health of individuals (as measured by recent morbidity) and on the quantity of food nutrients consumed by households in Indonesia. Studies such as these are useful to policymakers in that they inform them how changes in food prices and program provision affect the health of individuals. If separate demand equations are estimated for different types of individuals, perhaps differentiated by age and gender, even more is learned about the distribution of the effects of policy changes. For example, Pitt and Rosenzweig found that the effects of price changes on recent morbidity differed between male heads of household and their spouses in Indonesia. As useful as such studies may be, they tell almost nothing about how households allocate resources among their members. As demonstrated below, it is quite difficult to estimate the demand for goods within the household.

A useful way of considering the problem of the demand for goods within the household is to formulate it in terms of intrahousehold conditional demand equations. Such equations ask the question, how do the allocations provided one household member affect the allocations of others? For example, how does one person's health, time allocation, or food consumption affect that of another? And how does the allocation of each of these goods affect the other?

This paper begins with a restatement of the simple model of demand in which a single consumer chooses among a set of market goods. The concept of a conditional demand equation, first considered by Pollak (1969), is introduced in this simple framework before proceeding on to the case of multiperson households and nonmarket goods. The paper stresses the problem of finding believable and theoretically justified restrictions that enable the researcher to statistically identify cross-person demand relationships within the household and shows why it is, in general, impossible to identify these demand relationships based upon the usual exclusion restrictions. Some approaches to estimation are considered and examples drawn from the author's work with Mark Rosenzweig are used to illustrate these methods.

CONDITIONAL DEMAND IN A SIMPLE ONE-PERSON HOUSEHOLD

Consider the simplest possible case of a one-person household with fixed (exogenous) income, M, and a utility function having only market goods, x_i, as arguments. The household's problem is

$$\max U = U(x_1, x_2, \ldots, x_k), \quad \text{subject to} \quad \sum_{i=1}^{K} p_1 x_i = M, \tag{1}$$

where p_i is the market price of good i, taken parametrically by the household. The uncompensated (Marshallian) demand equations resulting from this problem are

$$x_i = d_i(p_1, p_2, \ldots, p_k, M), i = 1, \ldots, K. \tag{2}$$

Dual to this problem is the problem of minimizing the cost, c, of obtaining a given level of utility, U:

$$\min c = c(p_1, p_2, \ldots, p_k, U). \tag{3}$$

The partial derivatives of the cost function are the compensated (Hicksian) demand equations:

$$\frac{\partial c}{\partial p_i} = x_i = q_i(p_1, p_2, \ldots, p_k, U), i = 1, \ldots, K. \tag{4}$$

Consider the problem of this single-person household if the allocation of one or more goods are rationed and the ration is binding. That is, given the household's income and the prices it faces, it would wish to consume at least as much as the rationed quantity if it could freely choose consumption levels. In this case, consumption of the rationed good exactly equals the ration amount. For simplicity, good x_1 will be treated as the rationed good, and the rationed quantity is $\bar{x}_1$ The consumer's problem is now

$$\min c^1 = c^1(\bar{x}_1, p_1, p_2, \ldots p_k, U), \tag{5}$$

where c^1 denotes the cost of utility, conditional on consumption of the rationed quantity $\bar{x}_1$. One can view the consumer's problem as choosing consumption levels of all goods except x_1 so as to maximize utility subject to consuming $x_1 = \bar{x}_1$ and to having income of M - $p_1\bar{x}_1$ to spend on the other K-1 goods. Formally, the rationed household's problem is

$$\begin{aligned} &\max U = U(\bar{x}_1, x_2, \ldots, x_k), \\ &\text{subject to} \\ &\sum_{i=2}^{K} p_i x_i = M - p_1\bar{x}_1. \end{aligned} \tag{6}$$

From the rationed problem equation (6), it can be seen that the only way in which the price of the rationed good p_1 affects demands is through the term $p_1\bar{x}_1$ on the left-hand side of the budget constraint. A fall in the price of the rationed good just increases the income available for purchasing all other goods, that is, the price change only induces an income effect. There is no substitution effect resulting from changing the price p_1 as long as the ration remains binding. Thus the derivatives of the conditional cost function, equation (5), which are the conditional compensated (Hicksian) demand equations, do not depend on p_1:

$$\frac{\partial c^1}{\partial p_i} = x_i = q_i^1(\bar{x}_1, p_2, p_3, \ldots, p_k, U), \; i = 2, \ldots, K, \tag{7}$$

where q_i^1 is the demand for good i, conditional on the ration $\bar{x}_1$.

Estimation of Conditional Demand Equations for the Single-Person Household

If the rationed good is "food," these conditional demand equations tell how a change in the consumption of food alters the consumption of the other goods, such as time allocation and nonfood goods consumption. In the absence of actual rationing, the conditional demand equations (7) would never have to be estimated. The integrability of demand systems means that everything about preferences that can be learned is learned by estimating the unconditional demand equations (4). The parameters of the conditional demand equations can be constructed from the parameters of the unconditional demand equations. Furthermore, the reverse is also true: the unconditional cost function can be recovered from the conditional cost function (Browning 1983).

Nonetheless, consider how to empirically estimate the effect of changing the level of consumption of one good in a single-person household on the demand for all other goods. This is exactly the problem of estimating the conditional demands, equation (7), in the absence of rationing. To make the problem realistic, assume that estimation will use data from single-person households having heterogeneous preferences. If the preference heterogeneity results in additive stochastic terms appended to the conditional demand equations, least squares estimation will result in heterogeneity bias.

The level of observed consumption of the good conditioned upon x_1 is a regressor that will likely be correlated with this preference-based error. Consumers with above average preferences for good x_1 will consume more of it and consequently have less income remaining to spend on all other goods. One obvious approach to estimating models with endogenous regressors is to use instrumental variable methods. And from equations (4) and (7), there is a single, theoretically justified exclusion restriction, the price p_1, which is a determinant of x_1 in equation (4) but does not appear in the demand equation (7) conditional on x_1. It is straightforward to extend this example by conditioning on more than one good. Conditioning on the quantity consumed by more than one good results in conditional demand equations that exclude the prices of all conditioned goods as arguments (regressors), thus assuring exactly identifying exclusion restrictions for instrumental variable estimation.

CONDITIONAL DEMAND IN THE MULTIPERSON HOUSEHOLD

The problem for a multimember household analogous to the one described in equation (1) is:

$$\max U = U(x_{11},x_{12},\ldots,x_{1J},x_{21},x_{22},\ldots,x_{2J},\ldots,x_{KJ}),$$
$$\text{subject to}$$
$$\sum_{i=1}^{K}\sum_{j=1}^{J} p_i x_{ij} = M, \tag{8}$$

where x_{ij} is the consumption of good i by household member j in a household comprised of J members. The key difference between the single-person and multimember household models is not the larger number of (person-specific) goods, but the fact that there are more goods than there are prices. There are K market goods and J household members yielding KJ person-specific goods, but still only prices for K goods.[1] The cost function, conditioning on the consumption of one good by one member of the household, is then

$$c^{11} = c^{11}(\bar{x}_{11},p_1,p_2,\ldots,p_k,U), \tag{9}$$

where indexes are innocuously chosen such that the cost function c^{11} is conditional on the consumption of good 1 by household member 1 (x_{11}). In the multimember household, unlike the single-person household, a reduction in the price of the rationed good p_1 can have substitution effects on demands for all other goods. To see this, note that the household's problem is to allocate KJ - 1 person-specific goods so as to maximize utility subject to consuming $x_{11} = \bar{x}_{11}$ and remaining income $M - p_1\bar{x}_{11}$:

$$\max\ U = U(\bar{x}_{11},x_{12},x_{13},\ldots,x_{1J},x_{21},x_{22},\ldots,x_{2J},\ldots,x_{KJ}),$$

subject to

$$\sum_{j=1}^{J}\sum_{i=2}^{K} p_i x_{ij} + p_1 \sum_{j=2}^{J} x_{1j} = M - p_1\bar{x}_{11}. \tag{10}$$

In equation (10), the price p_1 still appears on the left side of the budget constraint as it reflects the prices for the J-1 "unconstrained" goods x_{12}, x_{13},...,x_{1J}. The derivatives of the conditional (and unconditional) cost function with respect to a price p_1 is not the compensated demand for a person-specific good as it was in the single-person household, because p_1 is the common price of J (or J-1 for p_1) person-specific goods in equation (10). The conditional demand for person-specific good x_{ij} is

$$q_{ij}^{11}(\bar{x}_{11},p_1,p_2,\ldots,p_K,U). \tag{11}$$

In this case, identifying exclusion restrictions are not available to carry out instrumental variable estimation, since the price p_1 of the conditioned good x_{11} is not excluded from the conditional demand equation. All the goods x_{1i} have the same price. Furthermore, one cannot infer the conditional demand equations by estimating the unconditional demand equations. The unconditional demand equations for person-specific goods are themselves not identifiable because person-specific prices do not exist for all goods.

Multimember Household Models with Household Production

The problem of the multimember household can now be generalized to include home-produced goods, such as health, and the time allocation of household members. The household's problem is now

$$
\begin{aligned}
&\max\ U = U(x_{11},x_{12},\ldots,x_{1J},x_{21},x_{22},\ldots,x_{2J},\ldots,x_{KJ},l_1,l_2,\ldots,l_J,h_1,h_2,\ldots,h_J),\\
&\text{subject to}\\
&h_j = h_j(x_{1j},x_{2j},\ldots,x_{kj},l_1,l_2,\ldots,l_j,z_j,\mu_j),\ \ j=1,\ldots,J\\
&\text{and}\\
&\sum_{i=1}^{K}\sum_{j=1}^{J} p_i x_{ij} + \sum_{j=1}^{J} w_j l_j + \sum_{j=1}^{J} p_z z_j = v + \sum_{j=1}^{J} w_j T,
\end{aligned}
\tag{12}
$$

where l_j is the home time of household member j, h_j is the quantity of home-produced good h (health) allocated to person j, w_j is the market wage of member j, z is an input into the production of the home-produced good, p_z is its price, and v is nonearnings (exogenous) income. The term μ_i represents person-specific endowments, such as innate healthiness, which are fixed and not changeable by the household. The health production functions given in equation (12) are general in that they allow for the technology producing h to be different for every household member, and for own-consumption of the market goods x and the home time l of every household member to be inputs into the production of h. But by treating home time as a "household public good"— that is, by not distinguishing among the allocations of person j's time to the production of each household members home good— the treatment of home time in the technology is not perfectly general. If home time were a private good allocatable to each person, there would be J^2 home time allocations and home time demand equations. In which case, even the wage would not be a good-specific price, since the price of the home time of the jth person devoted to the production of each of the household's J members would be identically w_j.[2]

Consider the nature of the conditional demand equations corresponding to equation (12) in the case of single-person household. Even though the h good is not a market good, there is a market good z that does not provide utility directly but only enters into the unconditional demand equations through its effect on h. The price of this good acts as the "price of health." Conditioning on $h = \bar{h}$, the conditional demand equations for the single-person household do not depend on p_z, and, thus, p_z is available as an identifying exclusion restriction for the instrumental variable estimation of the conditional demands. If there is a vector of health inputs like z, then there are over-identifying restrictions.

Unfortunately, as in the case with only market goods available, in a multimember household when the price p_z is not person-specific, p_z does not disappear from demand equations that condition on the level of h provided by any one (or subset of) household member(s).[3] Thus, except in the case of time allocation, the problem of more person-specific goods than prices precludes the estimation of person-specific unconditional demand equations as well as the use of instrumental variable methods to estimate intrahousehold conditional-demand equations.

SOME APPROACHES TO ESTIMATION

In spite of this gloomy theoretical outlook, many studies have indeed estimated intrahousehold conditional demand equations. There are essentially four approaches that have been followed.

One approach is to essentially ignore the problem, treating the conditioned-upon behaviors as exogenous. For example, in some of the literature, the labor supply of women has been regressed on the number of children or their health without regard to the possible effects of unobserved heterogeneity on the estimates.

A second approach is to make exclusion restrictions necessary for the use of instrumental variable methods, even though, as demonstrated above, it is difficult to find such restrictions that are not inconsistent with a general theory of household behavior. For example, Pitt and Rosenzweig (1985) estimate the way the health of male heads of farm households in Indonesia affects their labor supply. They use the prevalence of health programs, such as public health clinics and sanitation facilities, as identifying instruments (corresponding to p_z). A Hausman-Wu test "confirms" the endogeneity of health in this conditional labor-supply equation. But in these multimember households, health programs and facilities must also affect the health of the head's wife and other household members. Interpreting these estimates as the supply of labor conditional on own-health, as Pitt and Rosenzweig do, requires either that health prices affect the household head’s health but not the health of other household members, or that the health status and time allocation of other household members have no effect on the head’s labor supply (as in the single-member household).

Identification Through Cross-Person Restrictions on Demands

A third approach is to put additional structure on the model that, while not necessarily consistent with a general model of household behavior, involves restrictions less onerous than the zero exclusion restrictions of the second approach. Pitt and

Rosenzweig (1990) use this method to estimate the effects of infant morbidity on the allocation of time in Indonesian households. The linearized demand equations for home time of two household members, i and j, conditional on the health of family member k, are:

$$l_i = \alpha_{0i} + \alpha_{1i}p_1 + \alpha_{2i}p_2 + \ldots + \alpha_{Ki}p_K + \beta_{1i}w_1 + \beta_{2i}w_2 + \ldots \beta_{Ji}w_J + \gamma_i p_z + \lambda_i h_k + \epsilon_i \tag{13}$$

and

$$l_j = \alpha_{0j} + \alpha_{1j}p_1 + \alpha_{2j}p_2 + \ldots + \alpha_{Kj}p_K + \beta_{1j}w_1 + \beta_{2j}w_2 + \ldots + \beta_{Jj}w_J + \gamma_j p_z + \lambda_j h_k + \epsilon_j, \tag{14}$$

where ϵ_i is an error term that includes the effects of the J health endowments $\mu_1,\ldots,\mu_J$, and the remaining Greek letters are unknown parameters. Pitt and Rosenzweig impose the restriction that $\gamma_i = \gamma_j$. The plausibility of this restriction depends on the characteristics of the individuals i and j. If behavior is age-dependent, then the restriction is plausible if the individuals i and j are of approximately the same age. If behavior is gender-dependent, then the restriction is more plausible if i and j are of the same gender. Unconditional demand equations that demonstrate differences in price response by gender do not necessarily invalidate this restriction, since gender differences in price response in unconditional demand does not necessarily imply a different price response when conditioned on infant health (or any other behavior). In Pitt and Rosenzweig, this equality restriction is made for three member-types: the mother of the infant whose health status is conditioned on, and the infant's boy and girl teenage siblings.

In their study, time allocation is measured as principal activity in the week prior to the date of the survey, the 1980 National Socioeconomic Survey of Indonesia

(SUSENAS). Four mutually exclusive principal activities are distinguished: work, school, home care, and leisure. The linearized conditional demand equations for family members in a household containing a mother and her teenage son and daughter are:

$$I_{ij}^{*} = (\alpha_{iM} + \delta_{ij}^{A} D_j)' A_j + (\gamma_{iM} + \delta_{ij}^{h} D_j) h^{*} + (\beta_{iM} + \delta_{ij}^{x} D_j)' X + Z\lambda_{ij} + \epsilon_{ij}, \tag{15}$$

where I_{ij}^{*} is the level at which household member j undertakes activity i; D_j has the value of one in the equation for j and zero otherwise in the son and daughter equations; h^{*} is the endogenous health of the mother's infant child; A is a vector of member-specific exogenous variables; X and Z are vectors of household-specific exogenous variables, to be distinguished below; the Greek letters, except ϵ, represent parameters to be estimated; and ϵ_{ij} represents error terms having a multivariate distribution with zero means and covariance matrix Σ. The vector of exogenous variables Z is that for which equality restrictions are imposed:

$$\lambda_{ij} = \lambda_{ik} \quad i \neq k;\ i,k = \text{member type}. \tag{16}$$

The vector Z consists of two subsets of regressors: 26 prices or price indexes for goods, and a set of 16 community characteristics including health facilities and programs, public waste facilities, and drinking water sources.

Neither h^{*} nor the activity variable I_{ij}^{*} is observed in the data, only sets of dichotomous indicators indicating whether an infant had been sick or not and the primary time activity of the household member. The model was estimated using an instrumental variable household fixed-effects method (Chamberlain 1980). The fixed-effects procedure reduces the computational burden greatly, reduces the effects of heterogeneity across households, and eliminates the sample selection problem under suitable assumptions. It does not, however, permit identification of the parameters λ_{ij}, and only the differential effects δ are identified for the regressors X and h*. But the parameters δ are required to test the hypothesis about intrahousehold distribution, and the parameters λ_{ij} are maintained to not differentially affect time allocation, and thus are of little interest.

Table 1 presents estimates of the δ_{ij} parameters, which capture differential activity effects of child illness relative to the household care activity. These parameters were estimated from maximizing a single multinomial logit fixed-effects likelihood containing predicted values of infant health from a first-stage maximum likelihood binary logit regression. In the first stage equation, the sets of regressors upon which identification rests, area-level food prices ($X^2[26] = 80.6$), programs and health facilities ($X^2[7] = 17.6$), and water supplies ($X^2[5] = 18.4$), are statistically significant.

Parameters from two specifications are presented in Table 1: those from a single-stage multinomial fixed-effects logit that treats infant health as exogenous in the differential allocation of time, and one in which instrumental variable methods are applied. The two sets of estimates are not directly comparable since actual health is a dichotomous indicator of morbidity, while predicted morbidity is a continuous estimate of a latent variable. Nonetheless, the signs of the parameters differ in every case. Treating infant health as exogenous results in the (false) inference that there is no statistically different effect of infant health on the activity responses of teenage siblings. In the consistent instrumental variable estimates, infant healthiness does significantly influence differential time allocation. The hypothesis that the responses of sons and daughters to own age, infant illness, and the number and sex composition of household teenagers are identical is strongly rejected ($X^2[15] = 73.6$). However, the

Table 1— Single and two-stage maximum likelihood multinomial fixed-effects logit estimates: Differential effects of infant illness on household activities of daughters, sons, and mothers relative to home care

	Alternate Activity to Home Care					
	Labor Force		School		Leisure	
Household Pair/ Infant Illness	Exog-enous	Endog-enous	Exog-enous	Endog-enous	Exog-enous	Endog-enous
Daughter versus son	2.84 (1.40)[a]	-1.25 (2.23)[b] [3.23][c]	3.21 (1.58)[a]	-1.07 (2.03)[b] [2.85]	3.23 (1.58)[a]	-1.11 (1.88)[b] [2.57]
Daughter versus mother	-.348 (1.87)	.072 (0.49) [0.60]	...	...	...	...
Son versus mother	-3.19 (1.57)	1.32 (2.32) [3.20]				

Source: Pitt and Rosenzweig (1990), p. 981.

[a] Asymptotic t-ratios are in parentheses.

[b] Asymptotic t-ratios corrected for use of stochastic regressors, estimated from first stage, are in parentheses.

[c] Uncorrected t-ratios computed directly from information matrix are in brackets.

hypothesis that the responses of daughters and mothers are not different cannot be rejected ($X^2[6] = 14.5$), while it is rejected in a comparison of sons and mothers ($X^2[6] = 69.3$). The differential responses are thus based more on gender than on age.

Interpretation of the parameters is a bit complicated, since they represent the effect of a change in infant health on the allocation of time to one activity relative to another activity (home care) for one person-type relative to another. The consistently estimated parameter in the second row and first column of Table 1 (-1.25) tells us that increases in the latent illness of an infant reduces the daughter's time in the labor force, as compared to home care, more than the son's. Furthermore, increases in latent infant illness also reduce schooling and leisure time (relative to home care) for daughters more than for sons. These results suggest that reductions on infant morbidity would reduce gender-based inequality among teenagers in Indonesia. Assessing the quantitative importance of the level as opposed to the differential effect of latent infant illness requires at least one additional restriction. In Pitt and Rosenzweig (1990), quantitative estimates of level effects are obtained under the assumption that teenage boys do not alter their time devoted to household activities in response to the illness of an infant sibling.

<u>The "Endowment Method"</u>

The fourth approach to identifying intrahousehold conditional demand equations relies on treating the endowment μ_j as an implicit person-specific "price" for the home-produced good h in equation (12). To see this, note that the shadow price associated with an allocation ("ration") of a home-produced good to household member 1 is

$$p^*_{h1} = \frac{\partial U/\partial h_1}{\partial U/\partial v},$$

(17)

where p^*_{hi} is the shadow price of the allocation $h_1 = \overline{h_1}$, and v is, as before, nonearnings income. The shadow price of h_1 is the reduction in the (minimum) cost of obtaining the prior level of utility as a result of increase of health by one unit. The shadow price associated with an increase in the endowment μ_1 of household member 1 is similarly

$$p^*_{\mu 1} = \frac{\partial U/\partial \mu_1}{\partial U/\partial v} = \frac{\frac{\partial U}{\partial h_1}\frac{\partial h_1}{\partial \mu_1}}{\frac{\partial U}{\partial v}}. \quad (18)$$

If the endowments μ_j are additive in the household technologies (equation (12)), then $\partial h_1/\partial \mu_1 = 1$ and $p^*_{\mu 1} = p^*_{h1}$ Simply put, a unit increase in the endowment μ_j implies a unit increase in h_j when all input allocations are unchanged. The μ_j are exogenous person-specific determinants of the shadow price of h_j, and thus are valid instruments for the estimation of demand equations conditional upon the household allocation of the h good among its members.

In practice, none of the studies that have used the endowment method have estimated conditional demand equations with μ_j as an identifying instrument. Instead they have estimated reduced-form demand equations with the estimated μ_j added to the set of exogenous regressors. These are unconditional demand equations but now with an individual-specific exogenous component to health, μ_j, as an implicit person-specific price. The conditional demand equations can thus be fully recovered from the full set of unconditional demand equations. Regularity conditions of demand theory imply that the sign on μ_j in a reduced-form (unconditional) demand equation for ith input provided person j, x_{ij}, must be the same as the sign on h_j in a conditional (on h_j) demand equation for x_{ij}, with μ_j as an identifying instrument. In particular, regularity requires that

$$\frac{dh_j}{d\mu_j} = \Sigma\left(\frac{\partial h_j}{\partial x_{ij}}\frac{\partial x_{ij}}{\partial \mu j}\right) + \frac{\partial h_j}{\partial \mu_j} \geq 0, \tag{19}$$

which implies that an individual's health is never made worse off by the acquisition of exogenous health. The household may "tax" away some of the exogenous health by reducing health inputs x_{ij}, but will not tax away more than all of it.

The problem with any empirical methodology that estimates cross-person effects using endowments is that the μ_j is not directly observed. However, if the technology is known, it can be calculated as $\mu_j = h_j - h_j(*)$, where $h_j(*)$ is shorthand for the technology found in equation (12), and additive endowments are assumed. In practice, the technology is not known but must be specified and consistently estimated from nonexperimental data, subject to errors of measurement and other sources of stochastic variation.

Consider the (trivial) case of the single-person household and the problem of determining the effect of an exogenous increase in health of an individual on that individual's demand for the single health input z. If the health technology and health input demand equation are linear in the parameters, the result is

$$\begin{aligned} h_j &= \alpha + \beta z_j + \mu_j + \epsilon_j, \\ z_j &= \pi p_{zj} + \gamma p_{xj} + \lambda \mu_j + e_j, \end{aligned} \tag{20}$$

where μ_j is the health endowment of person (household) j, p_x is the price of a market good that does not affect health, and ϵ_j and e_i are random errors for which $E(\epsilon_j,e_j)=0$, $E(\epsilon_j, \mu_j)=0$, and $E(e_j,\mu_j)=0$. These restrictions on the error components imply that only source of error correlation in the two equations in (20) arise from the health endowments. The covariance between the residuals of these equations is thus $cov(\mu+\epsilon,\lambda\mu+e) = \lambda\sigma_\mu^2$ and is straightforward to estimate. Since σ_μ^2 is nonnegative, the sign of this term is the sign of λ. A negative λ implies compensatory behavior on the part of the household—an exogenous increase in health induces a reduction in health input demand, which would not be a very surprising result. Identification of the magnitude of λ requires knowledge of σ_μ^2.

Knowledge of the signs of person-specific λ's is of more interest in the multiperson household framework, where reinforcing behavior is more likely (Pitt, Rosenzweig, and Hassan 1990), but identifying these signs becomes problematic even if strong restrictions are placed on the error covariances (as above) and on the health technologies. Consider the simple case of a two-person household in which both person-types (j and k) have identical health technologies,

$$\begin{aligned} h_j &= \beta z_j + \mu_j + \epsilon_j, \\ h_k &= \beta z_k + \mu_k + \epsilon_k, \end{aligned} \qquad (21)$$

and the demand equations for health input provided persons j and k would be estimated:

$$\begin{aligned} z_j &= \pi_j p_z + \gamma_j p_x + \lambda_j \mu_j + \lambda_{kj} \mu_k + \epsilon_j, \\ z_k &= \pi_k p_z + \gamma_k p_x + \lambda_{jk} \mu_j + \lambda_k \mu_k + \epsilon_k, \end{aligned} \qquad (22)$$

from a sample of households (where the household subscript is dropped for simplicity). The parameter λ_{kj} represents the effect of person k's exogenous health on person j's allocation of good z. If, as before, the only source of residual covariation is through the μ's, and if $E(\mu_j, \mu_k) \neq 0$ as seems likely, identification of the three λ's requires knowledge of the variances and covariances of the μ's.

Assuming that $\sigma_\epsilon^2 = 0$ for both j and k— that is, the residual variance of the health technologies is identical to the variance of the endowment (no measurement error exists)— is sufficient for identifying the λ's. Rosenzweig and Schultz (1983), the first to apply the endowment method to the estimation of intrahousehold demand equations, treated the estimated residuals from the health technology as measured-with-error estimates of the endowments μ_j, and included them as regressors in the demand equations (22).[4] Under the assumption that the "measurement error" ϵ_j was uncorrelated with the error e_j, classical errors-in-variable bias results: parameters are biased towards zero. Thus, Rosenzweig and Schultz interpret their estimates as lower bounds on the true absolute values of the regression coefficients.

The problem is that there is often reason to believe that this measurement error is not orthogonal to the errors e_j of the demand equations. If the source of the error ϵ_j is only measurement error on health or human capital outcome h_j, then the orthogonality condition is not unbelievable. But if the source of the measurement error arises from the input z_j, then a very difficult form of bias arises in the estimation of the conditional demand equation (21). To see this problem, consider the linear health-production function for person j in equation (21) as consisting of measured-with-error output h_j and input z_j. The true (measured-without-error) endowment is

$$\mu_j^* = h_j^* - z_j^* \beta, \tag{23}$$

where h_i^* and z_i^* are the (unobserved) true values of health and the health input, respectively. If both health and the input z have measurement errors η_i and v_i with classical errors-in-variables properties, that is,

$$h_j = h_j^* + \eta_j \tag{24}$$

$$z_j = z_j^* + v_j, \tag{25}$$

where $E(h_i^*, \eta_j) = 0$, $E(z_i^*, v_j) = 0$, and $E(\eta_j, v_j) = 0$, then the estimated endowment, $\hat{\mu}_j$ is

$$\begin{aligned} \hat{\mu}_j &= (h_j^* + \eta_j) - (z_j^* + v_j)\hat{\beta} \\ &= \mu_j^* + \eta_j - v_j\,\hat{\beta}, \end{aligned} \tag{26}$$

where $\hat{\beta}$ is a consistent estimator of β. Thus the health endowment measurement error is $v_j = \eta_j$ - v and the demand for the good z by person j is

$$z_j = \pi_j\, p_z + \lambda_j\, \hat{\mu}_j + \lambda_{kj}\, \hat{\mu}_k + e_j. \tag{27}$$

If the marginal product of the health input is positive ($\hat{\beta} > 0$), then the measurement error v_j is systematically negatively correlated with the error of the person-specific input demand equation (22). Simply put, any error in the measurement of a production function input will impart a proportional measurement error in the estimated endowment. A subsequent regression of this input on estimated endowments will have spurious correlation arising from their common measurement error.

Consistent parameter estimates in the presence of measurement error in the regressors can be obtained by instrumental variable methods. Notice that estimation of the health technology typically requires instrumental variable estimation anyway, as long as there is any endowment heterogeneity ($\sigma_\mu > 0$) and household allocations are influenced by differential endowments ($\lambda_j \neq 0$, $\lambda_{kj} \neq 0$). Prices for health inputs (P_z above), including foods and medical care, are appropriate identifying instruments. However, prices are not valid instruments for $\hat{\mu}_j$ in the estimation of the demand equation (22), since they are by construction uncorrelated with the endowments. The only instruments possible are repeated (noncontemporaneous) measures of health and health inputs, inclusive of noncontemporaneous alternative measures of health. The validity of these instruments requires that the period-specific measurement errors are uncorrelated across periods. This was the approach followed by Pitt, Rosenzweig, and Hassan (1990) in their study of the intrahousehold allocation of food in rural Bangladesh.

In that study, weight-for-height endowment measures were estimated for all members of a sample of Bangladeshi households and used to study the intrahousehold allocation of calories. The study explicitly modeled and estimated the link among food consumption, health, labor-market productivity, occupational choice, and individual heterogeneity. Table 2 presents estimates of weight-for-height production functions, estimated with a sample of 1,737 individuals. Inputs include measured (not reported) calorie consumption over a 24-hour period, measures of the energy intensity of effort, age, age squared, age/sex interaction, dummy variables for pregnancy and lactation, and the quality of drinking water. Calorie consumption, energy intensity of effort, and pregnant/lactating status are considered endogenous in the two-stage least squares estimates. Instruments include household head's age and schooling level, landholdings, and the prices of all foods consumed interacted with individual age and sex variables, land, and head's schooling and age. The first column of Table 3 presents (inconsistent) ordinary least squares (OLS) estimates of the production function. A comparison with the consistent two-stage least squares

Table 2— Effects of calorie consumption, activity level, and pregnancy status on weight-for-height

Variable[a]	Ordinary Least Squares Estimates[b]	Two-Stage Least Squares Estimates[b]
Calorie consumption[c]	0.0295	0.136
	(4.09)	(3.37)
Very active occupation[c]	0.0859	-0.0119
	(5.34)	(0.23)
Exceptionally active occupation[c]	0.0668	-0.0817
	(3.43)	(1.26)
Pregnant[c]	0.262	0.326
	(7.69)	(1.34)
Lactating[c]	0.144	0.513
	(9.28)	(4.65)
Age	0.284	0.0987
	(16.6)	(1.90)
Age squared	-0.00456	0.0174
	(1.44)	(2.37)
Sex (male = 1)	0.00196	-0.0578
	(0.08)	(1.81)
Age x sex	0.0152	0.0687
	(1.74)	(4.04)
Water drawn from tube well	-0.0478	-0.0406
	(3.13)	(2.10)
Water drawn from well	-0.0720	-0.0693
	(4.11)	(3.15)
Water drawn from pond	-0.0460	-0.0649
	(2.30)	(2.55)
Constant	-2.56	-3.12
	(52.4)	(13.9)
N	1,737	1,737
R^2	0.775	...
F	395.1	...
H_0: No influence of calcium, carotene, thiamine, and riboflavin consumption[b] (F)	...	1.23
H_0: No difference in effect of calorie consumption by sex (F)	--	2.16

Source: Pitt, Rosenzweig, and Hassan (1990), p. 1150.

[a] All variables in logs, except sex, water sources, and activity level.

[b] Endogenous variable; instruments include household head's age and schooling level, landholdings, and prices of all foods consumed, used in interaction with individual age and sex variables, land, and head's schooling and age.

[c] Asymptotic t-ratios in parentheses.

(2SLS) estimates of column 2 reveals the importance of heterogeneity bias. Using OLS, the calorie elasticity is seriously underestimated and the effects of the energy-intensity of effort are the opposite sign of the 2SLS estimates. The 2SLS estimates reveal the importance of calorie consumption on weight-for-height and the depleting effect of active occupations.

The individual endowments were estimated based on the technology parameter estimates and the actual resources consumed or expended by each individual. In order to deal with the possibility of systematic measurement error in the measured endowments, a longitudinal subsample of households that were surveyed in four rounds over a 12-month period were used for the estimation of the intrahousehold calorie allocation equations. In addition to repeated measures of weight and height, individuals had measurements of mid-arm circumference and skinfolds taken in every round as well. Production functions for these health outcomes were estimated by two-stage least squares containing the same regressors and instruments as the weight-for-height production function. The instruments for an individual's weight-for-height endowment in any period are the estimated endowments of the three health attributes averaged over all other periods. The estimated effect on an individual's own endowment on his or her calorie allocation was found to be negative without using instruments for measurement error. The effect became positive when instrumental variable methods were applied. These results clearly support the existence of systematic measurement error in the estimated endowments.

A problem that arises in the specification of demand models that include cross-person effects is that households are of different size and demographic composition. If households had only two individuals, each of a different type

Table 3— Fixed effects two-stage generalized least squares: Effects of personal characteristics on individual calorie consumption

	Two-Stage Least Squares Estimates[b]			
	Males		Females	
Variable[a]	Endowment Effects Constant	Endowment Effects Vary With Age	Endowment Effects Constant	Endowment Effects Vary With Age
Own endowment[c]	0.447 (3.58)	...	-0.0278 (0.15)	...
Age < 6[c]	...	-0.435 (1.35)	...	-0.314 (0.46)
6 ≤ age < 12[c]	...	0.923 (2.29)	...	1.86 (2.13)
Age ≥ 12[c]	...	1.21 (2.69)	...	0.0894 (0.13)
Age	1.44 (22.9)	1.31 (14.9)	1.34 (18.1)	1.35 (17.9)
Age squared	-0.201 (16.7)	-0.170 (9.16)	-0.199 (13.4)	-0.206 (13.7)
N	429	429	371	371
X^2 (no individual error components)	46.5	48.35	32.36	26.17
Individual error variance/ total error variance	0.287	0.300	0.258	0.282

Source: Pitt, Rosenzweig, and Hassan (1990), p. 1152.

[a] All variables in logs.

[b] Asymptotic t-ratios in parentheses.

[c] Instrumental variables used are means of individual and family endowments for weight-for-height, skinfold thickness, and arm circumference calculated over all survey rounds, excluding the round from which observation was drawn.

(for example, male and female), then it is easy to specify the demand for goods by person j as a function of the person-specific prices, endowments, and observed exogenous characteristics of family member k as well as own characteristics and prices. But samples of households with differing numbers of individuals by type are "unbalanced" in that the attributes of a second son, for example, can only influence allocations in households that have a second son. In Pitt, Rosenzweig, and Hassan (1990), this problem was handled in two ways. First, the intrahousehold distribution of exogenous characteristics was summarized as moments of distributions. Regressors included the mean weight-for-height endowments, mean age, proportion of male family members, and variance of ages of family members. Higher moments did not significantly improve the fit. Cross-gender effects were estimated by introducing the mean weight-for-height endowment separately for males and females. Only in same-sex households are cross-effects not estimable.[5]

Second, a household fixed-effects two-stage generalized least squares estimation was applied to the sample of individuals, divided by gender. The advantage to the household fixed-effects procedure is that it deals "perfectly" with cross-person effects, since the demographic composition of a household is a fixed-effect to each household member.[6] Table 3 presents the individual calorie consumption equations estimated with this method. Interestingly, the parameter estimates diverge little from those obtained by specifying cross-effects separately for males and females, using moments of the distribution of ages and endowments. Columns 2 and 4 of Table 3 allow the parameters to vary by both age and gender. The pattern of estimated own endowments matches up well with the pattern of activities individuals of these age-gender groups predominately perform. Young children (less than 6 years of age) are not economically productive and thus there is no (current) labor market (productivity) return to additional calorie consumption for them. As a consequence, calorie compensation dominates—part of the better health derived from a higher endowment—is taxed away via a reduction in calorie allocations.

Male and female children 6-12 years of age exhibit calorie reinforcement. During these ages, both genders have the ability to choose among activities of varying levels of energy intensity (and economic return) for which there are apparent returns to health. A 10 percent increase in the health (weight-for-height) endowment increases calorie consumption by 9.2 percent for males and 18.6 percent for females. This is consistent with activities data that show that girls have a greater diversity of activities as characterized by energy intensity of effort than boys. Adult males have the greatest calorie reinforcement of all household member-types, while adult females, with limited choices of activity, have an endowment response that is essentially zero.

Not many data sets have information on person-specific health inputs, food intake in particular. Individual-level food intake data is seldom collected because of the great difficulty and cost of doing so. When it is collected, most often enumerators ask respondents to recall their consumption of a list of common foods during the prior 24 hours (as in the ICRISAT village surveys). Estimation of health production functions, typically using anthropometric measures of health as dependent variables, using data collected in this way have not always been successful. The Nutrition Survey of Rural Bangladesh, used in Pitt, Rosenzweig and Hassan(1990), had trained female enumerator's reside with each household and physically weigh and measure the food consumption of each household member over a 24 hour period. These data seem "better" than recall data in that they were statistically significant determinants of weight-for-height, but still suffer from at least two drawbacks (besides the cost of data collection). First, the typical reference period of 24 hours is rather short, so that even if enumerator's measured the day's consumption without error, food consumption on other days is unlikely to be the same. A single-days observation may be a noisy measure of even short-term level of food intake. This is perhaps less of a problem for investigating the determinants of weight-for-height, an indicator of short-run health, than for other measures of health such as morbidity. The second problem is that inserting an enumerator into a household to weigh and measure each individuals food intake, certainly an intrusive procedure, may cause the household to alter the level and allocation of the household in order to please the enumerator. Recent evidence from the Philippines, where enumerators physically measured the food consumption of each individual in sampled households for a 7 day period, suggests household's consumed much larger levels of high cost status foods such as eggs and milk during the first days of an enumerator's observations but substantially less as the week wore on.

In Pitt and Lavy’s (1995) study of the allocation of preventive medical care in Ghana, endowment measures derived from a morbidity technology, rather than from a nutritional-status (weight-for-height) technology, are more likely to capture those components of innate healthiness associated with medical care decisions. Anthropometric measures, such as weight-for-height, are specified as inputs capturing the effects of individual specific food consumption on morbidity. Weight-for height is thus treated as an aggregator of the food intakes in the health technology, with the assumption that food consumption affects morbidity h only through weight-for-height. Weight-for-height is expected to suffer much less from the measurement error problems associated with individual level food consumption information.[7]

One might ask how it can be claimed that it is, in general, "impossible" to find theoretically justified restrictions that can identify intrahousehold conditional demand relations, and yet the endowment method seems to do just that. While it is true that the endowment method is consistent with the theory of the household presented above, a strong restriction, in the sense of not being very believable, is, nonetheless, required for the statistical consistency of estimates based on the use of estimated endowments. Essentially, the restriction is that the researcher know the correct specification of the technology from which inputs are to be estimated. If the wrong functional form is chosen or if relevant inputs are omitted, the use of estimated endowments will not unbiasedly estimate the intrahousehold demands. Since one seldom can claim to know the "true" functional form for any structural relationship, functional form misspecification is not an issue peculiar to this problem alone. The problem of omitted inputs is a much more difficult one to brush aside. Go back to the simple example of the single-person household having a linear health technology with a single input as in equation (20), but now allow for a second input, q_j, unmeasured or unknown to the researcher:

$$\begin{aligned} h_j &= \alpha + \beta z_j + \gamma q_j + \mu_j + \epsilon_j \ , \\ z_j &= \pi p_{z_j} + \delta p_{q_j} + \gamma_j p_x + \lambda \mu_j + e_j \ , \end{aligned} \tag{28}$$

where p_q is the price of input q. It is clear that the production function (28) estimated without q_j as a regressor will have residuals that include the effects of q_j as well as the bias to the other parameters caused by its omission, since it is likely that z_j and q_j are correlated. Bias will result, since the estimated endowments obtained from those residuals will now be correlated with the prices in the demand equation for z_j, since the demand for q_j, like the demand for z_j, depends on the price of inputs. The endowment method thus relies on a covariance restriction for identification: the errors of the production function are uncorrelated with those of the reduced form demand equation. An omitted variable makes that correlation non-zero and the restriction invalid.[8] Any reasonable application of the endowment method must make a convincing case that it has reasonably complete data on production inputs.

SUMMARY

This paper has set out the problem of specifying and statistically identifying the demand for goods within the household, making use of the concept of conditional demand introduced by Pollak (1969). Essentially the identification problem arises from the absence of prices for most person-specific goods. These prices are required to estimate demand equations having cross-person effects. Two methods for estimating intrahousehold demands were discussed. One method made restrictions on parameters that may be inconsistent with a general model of household behavior. It was suggested that cross-person restrictions on parameters might be less onerous than the usual exclusion restrictions. The second method, known as the endowment method, involved making covariance restrictions, which, while not at odds with a theory of behavior, required rich data on individual-specific inputs. Both methods were illustrated with empirical examples, using data from developing countries.

REFERENCES

Browning, M. J. 1983. Necessary and sufficient conditions for conditional cost functions. *Econometrica* 51 (3): 851-856.

Chamberlain, G. 1980. Analysis of covariance with qualitative data. *Review of Economic Studies* 47 (1): 225-238.

Behrman, Jere R., Mark R. Rosenzweig and Paul Taubman. 1994. Endowments and the Allocation of Schooling in the Family and in the Marriage market: The Twins Experiment. Journal of Political Economy 102 (6): 1131-1174.

Filmer, Deon. 1995. The intrahousehold allocation of health and cognitive skills in developing countries. PhD Thesis. Department of Economics, Brown University.

Lee, L.-F., and M. Pitt. 1986. Microeconometric demand systems with binding nonnegativity constraints: The dual approach. *Econometrica* 54 (5): 1237-1242.

Pagan, A. 1984. Econometric issues in the analysis of regression with generated regressors. *International Economic Review* 25 (February).

Pitt, M. and V. Lavy. 1995. The intrahousehold demand for medical care in low income countries, manuscript.

Pitt, M., and M. Rosenzweig. 1985. Health and nutrient consumption across and within farm households. *Review of Economics and Statistics* 67 (May): 212-222.

________. 1990. Estimating the intrahousehold incidence of illness: Child health and gender inequality in the allocation of time. *International Economic Review* 31 (4): 969-989.

Pitt, M., M. Rosenzweig, and M. D. Hassan. 1990. Productivity, health, and inequality in the intrahousehold distribution of food in low-income countries. *American Economic Review* 80 (5): 1139-1156.

Pollak, R. 1969. Conditional demand functions and consumption theory. *Quarterly Journal of Economics* 83 (1): 70-78.

Rosenzweig, M. 1986a. Program interventions, intrahousehold distribution, and the welfare of individuals: Modeling household behavior. *World Development* 14: 233-243.

Rosenzweig, M., and P. Schultz. 1983. Estimating a household production function: Heterogeneity, the demand for health inputs, and their effects on birth weight. *Journal of Political Economics* 91 (5): 723-746.

Rosenzweig, M., and K. Wolpin. 1991. The scope for policy intervention. *Journal of Econometrics* 50 (1-2): 205-228.

NOTES

1. Of course, restrictions placed on the utility function can lead to the aggregation of goods across persons, reducing the excess number of "goods" relative to prices, but cannot eliminate the excess number altogether, except in the limiting case.

Time is one good that likely has person-specific prices (wages). This fact has been exploited by Rosenzweig (1986a) and others to estimate intrahousehold cross-wage effects. It is difficult to think of important classes of other market goods for which person-specific prices exist, although the shadow prices of goods produced in the household, such as health, are likely to vary across or within a household. These issues are addressed below.

2. In this discussion, the author is assuming interior solutions for time allocation—that is, the opportunity cost of time is the market wage. If no time is spent in the market, the market wage is not the shadow price of time and there is one less exogenous variable in the demand equations. Estimation of demand systems with corner solutions is essentially the estimation of conditional-demand equations with binding rations of zero (Lee and Pitt 1986).

3. It is likely that there are some "Z-goods" that are only inputs into the production of the home-produced good h for certain types of household members. For example, some inputs into the care of infants (diapers, infant formula, certain inoculations) are not also inputs into the care of older household members. There may be gender-specific health inputs reflecting the different

biologies of men and women. In practice, these prices are not often measured and more than one household member is of the same type.

4. Pagan (1984) has shown that ordinary least squares estimates of regression equations having an estimated residual as a regressor provides consistent estimates of the parameter covariance matrix as long as the estimated residuals are orthogonal to the regression residuals.

5. Estimation of a "true" cross-effect would require that the calculated moments of the intrahousehold distributions not include own characteristics. If they do, the estimates conform to the experiment in which a transfer of characteristics (endowment, age, gender) occurs within the household that leaves mean endowment unchanged.

A fully parameterized model would require estimation of demand equations for each demographic mix characterizing households in the sample. If the slopes of demand equations were thought to vary only with gender, then, even in households of four persons, there are five possible demand regimes corresponding to the number of females or males that can be found in households of four persons. In households with differing numbers of members, there would be additional regimes to be estimated, because the effect of a change in an individual's endowment on the resources allocated to them depends on how many other members are available for reallocation. If household size and composition are endogenous, if for no other reason than the response of fertility to endowments,it is a switching regimes model with endogenous regimes.

6. While treating household composition as a fixed effect seems less arbitrary than trying to specify parsimonious functional forms, such as moments of distributions, it is valid only if the underlying parameters of the individual-specific demand equations are themselves not functions of the demographic composition of the household.

7. The endowment method has been applied in new and interesting directions in Behrman et. al. (1994), Rosezweig and Wolpin (1991) and Filmer (1995).

8. The instrumental variable method for dealing with measurement error in the endowments is not applicable here, since the omitted variable is omitted in every period.