C-42

Center for Research on Economic and Social Theory
Research Seminar in Quantitative Economics

Discussion Paper

DEPARTMENT OF ECONOMICS
University of Michigan
Ann Arbor, Michigan 48109

[redacted]

Estimating the Dispersion of Tastes and Willingness to Pay

E. Philip Howrey
Hal R. Varian
The University of Michigan

July, 1983

C-42

Estimating the Dispersion of Tastes and Willingness to Pay

by

E. Philip Howrey and Hal R. Varian

In the standard approach to demand estimation the observed demand behavior is assumed to derive from maximization of a representative consumer's utility function subject to a budget constraint. The assumption of a representative consumer is adopted not for its inherent realism but for its analytical convenience; and in many cases it seems to work well as a tentative hypothesis.

However in some situations we may well desire a more general model that allows for differences in tastes across households. The most general alternative specification would be a model with all consumers having arbitrarily different utility functions. However, such a general model is usually impractical to specify and estimate.

A reasonable intermediate case is one where tastes are allowed to vary across the population according to some parametrically specified distribution. In this case we may well be able to estimate the parameters of the frequency distribution that reflects the variation in tastes across the population.

A circumstance where the estimation of the distribution of tastes is of particular interest is when we are interested in evaluating the distributional impact of some proposed policy change. In many situations we would like to know not only the average "willingness to pay" for a policy change, but also the distribution of the willingness to pay across the population.

King [1982] has described how one might go about estimating the variation in willingness to pay across the population when this variation depends on differences in observed demographic and economic variables. However, there will typically be further differences in willingness to pay that are not directly attributable to observed demographic differences in households. We interpret these differences as differences in tastes, although other interpretations may be possible.

Burtless and Hausman [1978] have estimated a model incorporating variation in tastes in the context of labor supply. They specified that the frequency distribution of an income elasticity was truncated normal and estimated the parameters of this distribution by an iterative maximum likelihood technique. They did not explicitly calculate the welfare distribution implied by their estimated parameter distribution but were well aware that this would be possible. Below we show that much simpler estimation techniques can be used when the distribution of tastes can be assumed to be normal, rather than truncated normal, and we use the parameters derived by our estimation procedure to calculate the distribution of willingness to pay.

The remainder of the paper proceeds as follows. First we examine the concept of willingness to pay and show how this concept can be explicitly calculated as a function of the unknown parameters of the utility function in the one consumer case. We then postulate a model where the parameters of the utility function vary across the population according to some frequency distribution. We can then derive the implied demand equations and estimate the unknown parameters of the distribution of tastes using a variation of the standard error components model. The estimated parameters can then be used to calculate the distribution across the population of the willingness to pay for any particular policy change. Finally we illustrate these methods using some data involving time-of-day pricing of electricity.

1. The Compensation Function

What do we mean by the willingness to pay? In this section we attempt to give a meaningful empirical content to this concept. Further discussion can be found in King [1981] and Varian [1979], [1983].

We begin with the indirect utility function for some specific individual which we denote by $v(p,y)$. The indirect utility function measures the maximum utility the consumer can attain given prices p and income y. Associated with this indirect utility function is its inverse, the expenditure function, denoted by $e(p,u)$. The expenditure function measures the minimum expenditure necessary to achieve a particular utility level u.

Suppose now that we are comparing two possible configurations of prices and income which we denote by (p,y) and (p',y'). We can ask how much money the consumer would need at prices p to be as well off as he would be in the situation described by (p',y'). We denote this number by $\mu(p;p',y')$. From the definition of the indirect utility function and the expenditure function, we have:

$$\mu(p;p',y') \equiv e(p,v(p',y'))$$

Following Hurwicz and Uzawa [1971] we refer to the function $\mu(p;p',y')$ as the "income compensation function" or sometimes just as the "compensation function". King [1981] refers to the same concept as the "equivalent income function."

A reasonable measure of the willingness to pay to avoid a movement from the situation (p,y) to the situation (p',y') is given by:

$$W = \mu(p;p,y) - \mu(p;p',y') = y - \mu(p;p',y')$$

By construction, a consumer who has income $y-W$ at prices p can reach the same level of utility as he could with income y' facing prices p'. Hence this seems like a sensible way to measure the welfare impact of some policy change. Of course, W as we have defined it above, is simply the negative of Hick's notion of the "equivalent variation" -- it is how much income would have to change at prices p so as to make the welfare situation of the consumer at prices p equivalent to that obtained at (p',y').

The compensation function can also be used in ratio form to define various measures of the "change in the cost of living". Consider for example the expression:

$$\pi = \mu(p;p',y')/\mu(p;p,y) = \mu(p;p',y')/y$$

The price index π measures how much income one would need at prices p to be as well off as one would be at (p',y'), expressed as a fraction of actual income at prices p.

It is worthwhile to note that $\mu(p;p',y')$ behaves exactly like an expenditure function with respect to variations in p, holding (p',y') fixed. It also behaves like an indirect utility function with respect to (p',y') holding p fixed. This can easily be seen from the definition: for fixed p, e(p,u) is an increasing function of u -- if you want to get more utility at fixed prices you have to spend more money. Hence e(p,v(p',y')) is simply a monotonic transformation of the indirect utility function v(p',y') and is therefore itself an indirect utility function.

As an example of the above ideas, suppose that the indirect utility function is given by:

$$v(p,y) = G(p) + y^{1-b}/(1-b) \qquad [1.1]$$

where G(p) is some negative monotonic, quasiconvex function of prices. Such a utility function is of special interest because it generates demand functions which exhibit constant income elasticity. By Roy's law the demand for good j is given by:

$$\ln x_j(p,y) = \ln(-\partial G(p)/\partial p_j) + b \ln y \qquad [1.2]$$

If b = 1 (the case of homothetic demand) then the indirect utility function in [1.1] takes the form

$$v(p,y) = G(p) + \ln y$$

so that the demand functions have the form:

$$\ln x_j(p,y) = \ln(-\partial G(p)/\partial p_j) + \ln y.$$

The expenditure function for an indirect utility function of form [1.1] can be found by solving for income as a function of utility:

$$e(p,u) = [(1-b)(u - G(p))]^{1/(1-b)}$$

Substituting $v(p',y') = u$ we have the income compensation function:

$$\mu(p;p',y') = [(1-b)(G(p') - G(p)) + y'^{1-b}]^{1/(1-b)} \quad [1.3]$$

In the homothetic case, similar calculations show that:

$$\mu(p;p',y') = \exp[G(p')-G(p)]y'$$

Thus parametric specification and estimation of the function $G(p)$ is sufficient to identify and calculate the compensation function $\mu(p;p',y')$. For a specific example, which we will refer to later, consider the Cobb-Douglas specification in which $G(p) = \Sigma_j \beta_j \ln p_j$. Then:

$$\ln\mu(p;p',y') = \ln y + \Sigma_j \beta_j (\ln p_j - \ln p_j')$$

Or:

$$\mu(p;p',y') = y\ \Pi_j (p_j/p_j')^{\beta_j}$$

Of course when estimating a system of equations it is essential to verify (or impose) the relevant cross-equation restrictions involving the appropriate Slutsky conditions. The concept of willingness to pay only makes sense if the demand behavior can be taken to derive from utility maximization in the first place.

2. Variations in Tastes

We turn now to the specification of taste variation. Suppose that household i has an indirect utility function $v(p,y,\delta,\varepsilon)$ where δ is a vector of parameters specific to the household, and ε is a non-household specific error term. We suppose that δ is distributed across households according to the frequency function $h(\delta,\Delta)$ where Δ is a vector of unobserved parameters, and that ε has the usual properties of an error term.

The demand function for the good j by the household with characteristics δ and error term ε is given by:

$$x_j = -\partial v(p,y,\delta,\varepsilon)/\partial p_j \;/\; \partial v(p,y,\delta,\varepsilon)/\partial y$$

Given observations on (p,x,y) for a number of households, it will typically be possible to estimate the parameters in Δ and thereby construct an estimate of the variation in tastes across the population.

Suppose for example that we observe several choices made by household i over time and that the indirect utility function for household i takes the Cobb-Douglas form:

$$v(p,y,\delta_i,\varepsilon_i) = \ln y_{it} + \Sigma(\beta_{ij} + \delta_{ij} + \varepsilon_{ijt})\ln p_{ijt}$$

In this case, the share equations for household i for good j at time t take the form:

$$w_{ijt} = \beta_j + \delta_{ij} + \varepsilon_{ijt}$$

The random variable δ_{ij} is specific to household i and remains fixed over time. The random variable ε_{ijt} is an additive disturbance term that varies over both households and time. The variation of δ_{ij} over households is what we refer to as variation in tastes. The share equations thus form a system of regression equations with an additive components disturbance term.

3. Estimation of Systems of Equations with Error Components

We now consider how to estimate the parameters of a system of equations with error components using panel data. The general form that we consider derives from the expenditure share equations shown above; namely,

$$y_{ijt} = x_{ijt}\,\beta_j + u_{ijt}$$

where y_{ijt} is the observed value of the dependent variable in equation j at time t for household i, x_{ijt} is a vector of k_j explanatory variables, and β_j is a vector of k_j regression coefficients. (We have changed our notation a bit to conform with econometric practice.) The disturbance term u_{ijt} is assumed to be of the form

$$u_{ijt} = \delta_{ij} + \varepsilon_{ijt}$$

where δ_{ij} is that part of the disturbance term specific to equation j of household i.

The T time-series observations for equation j of household i can be written in matrix form as

$$y_{ij.} = x_{ij.}\,\beta_j + u_{ij.}$$

where $y_{ij.}$ is a column vector with elements (y_{ijt}; t=1,2,...,T), $x_{ij.}$ is a T x k_j matrix with x_{ijt} in row t, and $u_{ij.}$ is a column vector with elements (u_{ijt}; t=1,2,...,T). The vector $u_{ij.}$ can similarly be written as

$$u_{ij.} = \delta_{ij}\, e_T + \varepsilon_{ij.}$$

where e_T is a column vector of ones and $\varepsilon_{ij.}$ is a column vector with elements (ε_{ijt}; t=1,2,...,T). The mT observations for individual i can now be written as

$$Y_i = Z_i\, \beta + U_i$$

where

$$Y_i = \begin{bmatrix} y_{i1.} \\ y_{i2.} \\ \vdots \\ y_{im.} \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_m \end{bmatrix}, \quad U_i = \begin{bmatrix} u_{i1.} \\ u_{i2.} \\ \vdots \\ u_{im.} \end{bmatrix}$$

and

$$Z_i = \begin{bmatrix} x_{i1.} & 0 & \cdots & 0 \\ 0 & x_{i2.} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & x_{im.} \end{bmatrix}.$$

The disturbance vector can be expressed as

$$U_i = \begin{bmatrix} \delta_{i1}\, e_T \\ \delta_{i2}\, e_T \\ \vdots \\ \delta_{im}\, e_T \end{bmatrix} + \begin{bmatrix} \varepsilon_{i1.} \\ \varepsilon_{i2.} \\ \vdots \\ \varepsilon_{im.} \end{bmatrix}$$

$$= \delta_i \otimes e_T + \varepsilon_i$$

where $\delta_i \otimes e_T$ denotes the Kronecker product of the vectors

δ_i and e_T.

Linear restrictions on the vector of regression coefficients,

$$R\beta = r,$$

are easily handled by transforming the variables. If R is a g x k matrix of rank g, it is possible to express g of the elements of β in terms of the remaining k - g elements. Thus, by reordering the variables in x_{ijt} if necessary, we can partition β in such a way that

$$R_1\beta^* + R_2\beta^{**} = r$$

were R_2 is a nonsingular g x g matrix. Solving for β^{**} yields

$$\beta^{**} = R_2^{-1}r - R_2^{-1}R_1\beta^*$$

so that

$$\beta = \begin{bmatrix} I \\ -R_2^{-1}R_1 \end{bmatrix} \beta^* + \begin{bmatrix} 0 \\ R_2^{-1}r \end{bmatrix} .$$

$$= R^*\beta^* + r^*$$

Substituting this expression into the regression model yields

$$Y_i^* = Z_i^*\beta^* + U_i$$

where

$$Y_i^* = Y_i - Z_i r^*$$

and

$$Z_i^* = Z_i R^*.$$

Thus homogeneity and symmetry restrictions are easily handled for linear systems of demand equations.

The standard assumptions about the error components; namely,

$$E(\delta_{ij}) = 0,$$
$$E(\delta_{ij}\delta_{ik}) = \Delta_{jk},$$
$$E(\varepsilon_{ijt}) = 0,$$
$$E(\varepsilon_{ijt}\varepsilon_{iks}) = \begin{matrix} \Omega_{jk} & t = s \\ 0 & t \neq s, \end{matrix}$$

and

$$E(\delta_{ij}\varepsilon_{ikt}) = 0,$$

imply that

$$E(U_i) = 0$$

and

$$E(U_i U_i') = \Psi = \Omega \otimes I_T + \Delta \otimes J_T$$

where I_T is a T x T identity matrix and $J_T = e_T e_T'$ is a T x T matrix with each element equal to one. These assumptions allow for correlation among the individual specific effects δ_{ij} as well as contemporaneous correlation of the disturbances ε_{ijt} across equations. The assumption that the error components are normally distributed and independent across individuals completes the specification of the model.

It is shown in the Appendix that maximum likelihood estimates of the parameters β (or β^*), $\Delta = (\Delta_{jk})$, and $\Omega = (\Omega_{jk})$ can be obtained by iterating the usual generalized least squares estimation procedure until convergence is achieved. In addition the structure of the covariance matrix Ψ can be used to simplify estimation of the elements in Δ and Ω. A relatively simple expression for Ψ^{-1} which is needed to calculate the generalized least squares estimates of β is also given in the Appendix.

4. Empirical Results

We now turn to an investigation of the demand for electricity by time-of-day. We consider a two-stage budgeting process of the form:

$$q_i = q_i(p_1, p_2, p_3, e)$$
$$e = e(\phi_1, \phi_2, y)$$

where

q_i = demand for electricity during period i (i=1,2,3),

p_i = price of electricity during period i,

e = expenditure on electricity,

ϕ_1 = price index of electricity,

ϕ_2 = price index of all other goods, and

y = household income.

Our empirical results are based on data collected in 1976 by the Arizona Public Service Company. A random sample of 80 households in the Phoenix and Yuma service areas were assigned at random to the 16 time-of-day rates shown in Table 1. Electricity usage of these households was recorded for a six-month interval. The following results are based on the records of 60 of these households for the last five months of the experiment.[1]

The subutility function for electricity consumption on which our empirical results are based is

$$v(p,e) = \ln e - \Sigma_j \beta_j \ln p_j$$

This leads to the expenditure share equations

$$\begin{aligned} w_{ijt} &= \beta_{ijt} \\ &= \beta_j + \delta_{ij} + \varepsilon_{ijt} \end{aligned}$$

for individual i at time t. Since these expenditure shares sum to one, it follows that $\beta_1 + \beta_2 + \beta_3 = 1$, and that

$$v(p,e) = \ln e - \sum_{j \neq k} \beta_j \ln (p_j/p_k) ,$$

so that only two of the equations need to be estimated. The maximum likelihood estimates of the parameters are shown in Table 2.

These parameter estimates exhibit two interesting features. First, the variation in the δ_{ij} across households is substantial. The estimated standard deviation of δ_{i1}, for example, is $\sqrt{.0070} = 0.084$ which is more than twice the estimated standard deviation of ε_{i1t} ($\sqrt{.0016} = .04$). Second, the estimated covariance between δ_{i1} and δ_{i2} is negative. Thus households that spend relatively more for

---

[1]A detailed description of the experimental design is given in Hill et al. (1979). Incomplete data prevented the use of all 80 households in our analysis.

Table 1

ARIZONA TOD EXPERIMENTAL RATE SCHEDULES
(¢/kWh)

Rate Group	Periods		
	Peak 2 pm - 5 pm	Shoulder 9 am - 2 pm 5 pm - 10 pm	Base 10 pm - 9 am
1	16	5	3
2	15	4	2
3	15	7	4
4	14	4	2
5	14	6	4
6	13	3	3
7	13	4	2
8	13	7	3
9	12	5	1
10	12	6	3
11	11	4	2
12	11	7	4
13	10	4	1
14	10	6	3
15	9	5	2
16	8	4	1

Source: Taylor (1979).

Table 2

Maximum Likelihood Estimates of
the Expenditure Share Equations
and Error Variances

Coefficient	Estimate
β_1	.4124
β_2	.4481
Δ_{11}	.0070
Δ_{22}	.0049
Δ_{12}	-.0048
Ω_{11}	.0016
Ω_{22}	.0009
Ω_{12}	-.0007

electricity during one of these periods would generally spend less during the other period.

We close this section with the remark that we are well aware of the restrictive nature of the Cobb-Douglas specification of the subutility function. It would be preferable to use a more flexible functional form such as the translog. Indeed, we initially examined the translog form but found that the quasi-convexity condition required of indirect utility functions was not satisfied by the parameter estimates at any of the rates shown in Table 1. Rather than estimate the parameters subject to the restriction of quasiconcavity, which is a relatively difficult procedure in our problem, we decided to impose the Cobb-Douglas form which guarantees the appropriate curvature in order to illustrate the method we have derived. In future work we intend to examine methods of estimation in which we can impose curvature restrictions on the estimated indirect utility function.

5. Estimating the Distribution of Willingness to Pay

The Arizona time of day pricing experiments are an interesting example for our purposes since they were specifically conducted in order to determine the feasibility of time-of-day pricing of electricity. The feasibility of time of day pricing depends, at least in part, on the willingness of households to accept those rates. Any initial resistance to time of day rates would presumably evaporate if households were to find that they were better off with TOD rates. Furthermore, the optimal design of time-of-day prices depends on consumers' utility functions for electricity consumption. Thus the estimation of the utility impact of TOD pricing seems of considerable interest.

Hence we consider now how to measure the expenditure necessary at some time varying prices p to achieve the same level of subutility for electricity expenditure achieved at a flat rate schedule $\bar{p}$ and initial expenditure $\bar{e}$; that is, we wish to calculate $\mu(p;\bar{p},\bar{e})$.

Using the expression for the Cobb-Douglas compensation function derived in Section 1:

$$\mu(p;\bar{p},\bar{e}) = \bar{e}\Pi_j(p_j/\bar{p}_j)^{\beta_j}$$

or:

$$\mu(p;\bar{p},\bar{e}) = \nu(p,\bar{p})\bar{e}$$

where $\nu(p,\bar{p}) = \Pi_j(p_j/\bar{p}_j)^{\beta_j}$.

The fraction:

$$\pi = \mu(p;\bar{p},\bar{e})/\mu(\bar{p};\bar{p},\bar{e}) = \nu(p,\bar{p})$$

measures the relative change in the compensation function when moving from flat rates to time of day rates. It measures how much money one would need to have at the TOD rate schedule to have the same subutility one had at the flat rate schedule, expressed as a fraction of the expenditure at the flat rate schedule.

Since electricity consumption is only part of the entire consumption bundle, we cannot interpret π as a measure of the change in overall welfare. However, π can be interpreted as a kind of price index for electricity consumption. We will briefly describe this interpretation below.

Recall the two stage budgeting process mentioned earlier. The utility maximization problem involved can be written as:

$$\max_{x,q} \; u(x,w(q)) \quad \text{s.t.} \quad rx + pq = y$$

where (r,x) are the vectors of prices and quantities of nonelectricity consumption, and (p,q) are the analogous vectors for electricity consumption. The subutility function for electricity consumption, w(q), is assumed to be homothetic. It follows that the compensation function will be of the form $\mu(p;\bar{p},\bar{e}) = \nu(p,\bar{p})\bar{e}$.

Using the compensation function as an indirect subutility function for electricity consumption, we can rewrite the consumer's maximization problem as:

$$\max_{x,\bar{e}} \; u(x, \nu(p,\bar{p})\bar{e}) \quad \text{s.t.} \quad rx - \bar{e} = y$$

Letting $Q = \nu(p,\bar{p})\bar{e}$ be a "quantity index" for electricity consumption, we can write this problem as:

$$\max_{x,Q} \; u(x,Q) \quad \text{s.t.} \quad rx + Q/\nu(p,\bar{p}) = y$$

Thus, $1/\nu(p,\bar{p})$ serves as a price index for electricity consumption. The ratio:

$$\pi = \nu(p,p)/\nu(p,\bar{p})$$

can be referred to as an index of the change in the cost of electricity. If π is greater than 1 then the price of electricity consumption has risen in the move from flat to TOD rates, and if it is less than 1, the price of electricity has fallen. Note that this price index can be given a welfare interpretation: if π is greater than 1, and no other prices change, the consumer is definitely worse off at time-of-day prices than at flat rate prices.

The above discussion is true for an arbitrary homothetic subutility function. For the Cobb-Douglas case used in our empirical study, π is given by the explicit formula:

$$\pi = \Pi_j(\bar{p}_j/p_j)^{\beta_j} .$$

The value of π is shown in Table 3 for each of the 16 rate schedules of Table 1 and for 4 different flat rates: $\bar{p}$ = 4, 6, 8, and 10. The values in this table are computed for the average, or "representative" household in the sample.

Prior to the experiment, households faced a declining block rate schedule and paid an average of approximately 4¢ per kwh. It is clear from the entries in Table 3 that from the point of view of households none of the TOD rate schedules is superior to the 4¢ flat rate. Discounts of up to 50% are required to make households indifferent between the TOD rates and the 4¢ flat rate. As the flat rate increases, several of the TOD schedules become quite attractive, as we would expect. All of the TOD schedules

Table 3

Estimate of the Change in the Cost of Electricity

Rate Schedule	Flat Rates 4	6	8	10	PEQ
1	1.8806	1.2537	0.9403	0.7522	7.52
2	1.5658	1.0439	0.7829	0.6263	6.26
3	2.2163	1.4776	1.1082	0.8865	8.87
4	1.5219	1.0146	0.7609	0.6088	6.09
5	2.0104	1.3403	1.0052	0.8042	8.04
6	1.3731	0.9154	0.6865	0.5492	5.49
7	1.4761	0.9840	0.7380	0.5904	5.90
8	2.0071	1.3381	1.0036	0.8029	8.03
9	1.4329	0.9552	0.7164	0.5731	5.73
10	1.8124	1.2082	0.9062	0.7249	7.25
11	1.3778	0.9185	0.6889	0.5511	5.51
12	1.9502	1.3002	0.9751	0.7801	7.80
13	1.2026	0.8017	0.6013	0.4810	4.81
14	1.6811	1.1207	0.8405	0.6724	6.72
15	1.4018	0.9345	0.7009	0.5607	5.61
16	1.0969	0.7313	0.5484	0.4388	4.39

are superior to a 10¢ flat rate.

The last column in Table 3 gives the flat rate that is equivalent to the corresponding TOD schedule in the sense that the same expenditure on electricity gives the same level of subutility with the TOD rates and the equivalent flat rate. For example, the equivalent flat rate for rate schedule 1 is 7.52¢ per kwh. Flat rates below 7.52¢/kwh are preferable to TOD schedule 1 whereas for flat rates above 7.52¢/kwh, the TOD schedule is preferable.

The above figures are presented for the representative household. Since our econometric results indicated significant dispersion of tastes across households we also examine the variation in willingness to pay for TOD rates across households. In particular, the value of the compensation function for household i is given by:

$$\mu_i = (\nu(p,\bar{p}) + h_i)\bar{e}_i.$$

where

$$h_i = \exp(\Sigma_j \delta_{ij} \ln p_j)$$

The variability of h_i can be characterized in several ways. One way is to note that since δ_{ij} is a Normal random variable, h_i is also Normally distributed, and we can easily compute its variance. However, that number itself does not have much intuitive content. A more interesting way to characterize the variability of h_i is to calculate the probability that $v(P,\bar{P}) + h_i \leq 1$. This is the probability that a household selected at random would benefit from a switch from flat to TOD rates. The results of this calculation are shown in Table 4.

For rate schedule 1, for example, virtually none of the households would benefit from a switch to TOD rates from a 4¢ flat rate. If the flat rate were 8¢/kwh, 68% of the households would benefit from a switch to the TOD schedule. The corresponding entry in Table 3 indicates that the representative household would benefit from this change. Thus while the representative household would be better off with the TOD rates, 32% of the households would lose in the sense that in order to attain the same level of utility they would have to pay more for electricity.

Another way to characterize the variability across households is to obtain a value for the equivalent price such that the probability is at least α that a household selected at random would be better off with the TOD rate than with the equivalent flat rate. This is the value of $\bar{p}$ such that the probability that $\ln v(P,\bar{P}) < 0$ is α. The 90% certainty equivalent flat rate is shown as PEQ^* in Table 4. For TOD schedule 1, the 90% certainty equivalent rate is 8.88¢. This means that at least 90% of the households would prefer TOD schedule 1 to any flat rate in excess of 8.88¢/kwh.

6. **Summary**

We have shown how one can estimate the distribution of willingness to pay across the population using panel data. In our example, there seems to be a significant dispersion of willingness to pay for time-of-day pricing of electricity. Such dispersion should be taken into account in examining the welfare implications of policy choices.

Table 4

Probability of Benefit from a Switch from Flat to TOD Rates

Rate Schedule	Flat Rate				PEQ*
	4	6	8	10	
1	.00	.04	.68	.99	8.88
2	.00	.39	.94	1.00	7.64
3	.00	.00	.16	.88	10.09
4	.00	.46	.97	1.00	7.38
5	.00	.00	.48	.99	9.10
6	.00	.76	1.00	1.00	6.43
7	.00	.54	.98	1.00	7.10
8	.00	.00	.49	.97	9.27
9	.03	.59	.96	1.00	7.33
10	.00	.04	.82	1.00	8.30
11	.01	.74	1.00	1.00	6.52
12	.00	.00	.63	1.00	8.62
13	.15	.89	1.00	1.00	6.04
14	.00	.11	.97	1.00	7.57
15	.00	.72	1.00	1.00	6.50
16	.28	.97	1.00	1.00	5.39

Appendix

Maximum Likelihood Estimation of Seemingly Unrelated Regressions with Error Components

In this appendix maximum likelihood estimates of the parameters of a set of regression equations with additive error components are obtained. This provides a generalization of the single-equation results given by Graybill [1961] for the model

$$y_{it} = \mu + \delta_i + \varepsilon_{it}.$$

The recent work by Avery [1977] and Baltagi [1980], building on earlier results obtained by Wallace and Hussain [1969], Amemiya [1971], Nerlove [1971], and Maddala [1971], deals with systems of equations of the form

$$y_{it} = x_{it}\beta + \delta_i + \eta_t + \varepsilon_{it} .$$

In the model considered here, the η_t term is missing. The essential feature of this simpler model is that for a given value of β, maximum likelihood estimates of the variances of the error components can be calculated recursively. The operational result is that maximum likelihood estimates can be obtained by iterating the usual generalized least squares estimation procedure with analysis of variance estimates of the covariance matrices of the error components.

1. The Error Components Model

The model consists of m seemingly unrelated regression equations for each of n individuals. Thus, at time t, we have

$$y_{ijt} = x_{ijt}\beta_j + u_{ijt} \qquad \begin{array}{l} i = 1,2, \ldots, n \\ j = 1,2, \ldots, m \\ t = 1,2, \ldots, T \end{array}$$

where y_{ijt} is the observed value of the dependent variable in equation j at time t for individual i, x_{ijt} is a vector of k_j explanatory variables, and β_j is a vector of k_j regression coefficients. The disturbance term u_{ijt} is assumed to be of the form

$$u_{ijt} = \delta_{ij} + \varepsilon_{ijt}$$

where δ_{ij} is that part of the disturbance term specific to equation j of individual i.

The observations for equation j of individual i can be written in matrix notation as

$$y_{ij.} = x_{ij.}\beta_j + u_{ij.}$$

where $y_{ij.}$ denotes the column vector with elements (y_{ijt}; t = 1,2, ..., T), $x_{ij.}$ is the T x k_j matrix with x_{ijt} in row t, and $u_{ij.}$ is the column vector with elements (u_{ijt}; t = 1,2, ..., T). The vector $u_{ij.}$ can similarly be written as

$$u_{ij.} = \delta_{ij} e_T + \varepsilon_{ij.}$$

where e_T is a column vector of ones and $\varepsilon_{ij.}$ is the column vector with elements (ε_{ijt}; t=1,2, ..., T). The mT observations for individual i can now be written as

$$Y_i = Z_i \beta + U_i$$

where

$$Y_i = \begin{bmatrix} y_{i1.} \\ y_{i2.} \\ \cdot \\ \cdot \\ \cdot \\ y_{im.} \end{bmatrix}, \qquad \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \cdot \\ \cdot \\ \cdot \\ \beta_m \end{bmatrix}, \qquad u_i = \begin{bmatrix} u_{i1.} \\ u_{i2.} \\ \cdot \\ \cdot \\ \cdot \\ u_{im.} \end{bmatrix},$$

and

$$Z_i = \begin{bmatrix} x_{i1.} & 0 & \cdots & 0 \\ 0 & x_{i2.} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & x_{im.} \end{bmatrix}.$$

The disturbance vector can be expressed as

$$U_i = \begin{bmatrix} \delta_{i1}e_T \\ \delta_{i2}e_T \\ \cdot \\ \cdot \\ \cdot \\ \delta_{im}e_T \end{bmatrix} + \begin{bmatrix} \varepsilon_{i1.} \\ \varepsilon_{i2.} \\ \cdot \\ \cdot \\ \cdot \\ \varepsilon_{im.} \end{bmatrix}$$

$$= \delta_i \otimes e_T + \varepsilon_i$$

where $\delta_i \otimes e_T$ denotes the Kronecker product of the vectors δ_i and e_T. The standard assumptions about the error components; namely,

$$E(\delta_{ij}) = 0,$$

$$E(\delta_{ij}\delta_{ik}) = \Delta_{jk},$$

$$E(\varepsilon_{ijt}) = 0,$$

$$E(\varepsilon_{ijt}\varepsilon_{iks}) = \begin{cases} \Omega_{jk} & t=s \\ 0 & t\neq s, \end{cases}$$

and

$$E(\delta_{ij}\varepsilon_{ikt}) = 0,$$

imply that

$$E(U_i) = 0$$

and

$$E(U_iU_i') = \Omega \otimes I_T + \Delta \otimes J_T$$

where I_T is a TxT identity matrix and $J_T = e_Te_T'$ is a TxT matrix with each element equal to one. The assumption that the error components are normally distributed and independent across individuals completes the specification of the model.

2. The Likelihood Equations

We now seek the equations that the parameter estimates must satisfy if they are to maximize the likelihood function. The log-likelihood function for y_{ijt} is

$$L(Y;\beta,\Delta,\Omega) = \sum_{i=1}^{n} \log f(Y_i).$$

Since each Y_i is a multivariate normal vector with mean $\mu_i = Z_i\beta$ and covariance matrix $\Psi = \Omega \otimes I + \Delta \otimes J$, it follows that

$$f(Y_i) = (2\pi)^{-mT/2}|\Psi|^{-1/2}\exp(-(Y_i-Z_i\beta)'\Psi^{-1}(Y_i-Z_i\beta)/2).$$

Omitting the inessential constant term and multiplying by two, the (modified) log-likelihood for Y_i is

$$L_i^* = -\log|\Psi| - (Y_i - Z_i\beta)'\Psi^{-1}(Y_i - Z_i\beta)$$

and the corresponding log-likelihood for Y is

$$L_\cdot^* = \sum_{i=1}^{n} L_i^*$$

$$= -n \log|\Psi| - \sum_{i=1}^{n} (Y_i - Z_i\beta)'\Psi^{-1}(Y_i - Z_i\beta).$$

In principle, maximum likelihood estimates could be obtained by maximizing the likelihood function directly using numerical methods. However, it is possible in this case to take advantage of the structure of the covariance matrix Ψ to obtain the likelihood equations explicitly. An iterative procedure can then be used to solve the likelihood equations for the maximum likelihood estimates.

Setting the derivative of L^* with respect to β equal to zero, it follows that the likelihood equation for β is

$$\beta = \left(\sum_{i=1}^{n} Z_i'\Psi^{-1}Z_i\right)^{-1}\left(\sum_{i=1}^{n} Z_i'\Psi^{-1}Y_i\right);$$

that is, the maximum likelihood estimates $\hat{\beta}$ and $\hat{\Psi}$ must satisfy this equation. Thus $\hat{\beta}$ is simply the generalized least squares estimate of β based on the maximum likelihood estimate $\hat{\Psi}$ of Ψ.

In order to obtain the likelihood equations for Δ and Ω, the likelihood function is written as

$$L^* = - n \log |\Psi| - \sum_{i=1}^{n} U_i' \Psi^{-1} U_i$$

where $U_i = Y_i - Z_i\beta$ is obviously a function of β but not of either Δ or Ω. The following results will be used to rewrite the likelihood function in a more convenient form and then obtain the maximum.

<u>Proposition 1</u>. The determinant of $\Psi = \Omega \otimes I_T + \Delta \otimes J_T$ is:

$$|\Psi| = |\Omega|^{T-1} |\Omega + T\Delta|$$

<u>Proof</u>. Since Ψ may be factored according to

$$\Psi = (\Omega \otimes I)(I \otimes I + \Omega^{-1} \Delta \otimes J),$$

it follows that

$$|\Psi| = |\Omega \otimes I| \; |I + \Omega^{-1} \Delta \otimes J|$$

provided, of course, that Ω is non-singular. Recall that the determinant of a matrix is the product of its characteristic roots which, in the case of $A \otimes B$ are the products of the characteristic roots of A and B (Theil [1971, 305]). The characteristic roots of $\Omega \otimes I$ are the characteristic roots ω_1 ω_2, ... ω_m of Ω each repeated T times so that

$$|\Omega \otimes I| = \omega_1^T \omega_2^T \ldots \omega_m^T$$

$$= (\omega_1 \omega_2 \ldots \omega_m)^T$$

$$= |\Omega|^T.$$

The characteristic roots of $I + \Omega^{-1}\Delta \otimes J$ are equal to one plus the characteristic roots of $\Omega^{-1}\Delta \otimes J$. The symmetric matrix J has unit rank and therefore only one non-zero characteristic root which is equal to T. The non-zero characteristic roots of $\Omega^{-1}\Delta \otimes J$ are therefore T times the characteristic roots λ_1, λ_2, ..., λ_m of $\Omega^{-1}\Delta$. Thus

$$|I + \Omega^{-1}\Delta \otimes J| = \Pi(1 + \lambda_j T)$$

$$= |I + T\Omega^{-1}\Delta|$$

$$= |\Omega^{-1} (\Omega + T\Delta)|$$

$$= |\Omega|^{-1}|\Omega + T\Delta|.$$

and the desired result follows.

<u>Proposition 2</u>. Provided Ω is nonsingular, the inverse of $\Psi = \Omega \otimes I_T + \Delta \otimes J_T$ is

$$\Psi^{-1} = \Omega^{-1}\otimes I - (\Omega+T\Delta)^{-1}\Delta\Omega^{-1}\otimes J.$$

Proof. Using the property that

$$(A\otimes B)(C\otimes D) = AC\otimes BD,$$

it is easy to verify that $\Psi^{-1}\Psi = \Psi\Psi^{-1} = I$.

Proposition 3. Let

$$\hat{\delta}_{ij} = T^{-1} \sum_{t=1}^{T} u_{ijt}$$

and

$$\hat{\varepsilon}_{ijt} = u_{ijt} - \hat{\delta}_{ij}$$

so that

$$U_i = \hat{\varepsilon}_i - \hat{\delta}_i \otimes e_T.$$

Then

$$U_i'\Psi^{-1}U_i = \hat{\varepsilon}_i'(\Omega^{-1}\otimes I)\hat{\varepsilon}_i + T\hat{\delta}_i'(\Omega+T\Delta)^{-1}\hat{\delta}_i.$$

Proof. As a preliminary matter, notice that

$$e_T'\hat{\varepsilon}_{ij.} = \sum_{t=1}^{T} \hat{\varepsilon}_{ijt} = 0$$

so that for any matrix A with m columns

$$(A \otimes e_T') \hat{\varepsilon}_i = 0$$

and

$$(A \otimes J) \hat{\varepsilon}_i = 0.$$

Now using the result of Proposition 2 for Ψ^{-1},

$$U_i' \Psi^{-1} U_i = U_i'(Q \otimes I - R \otimes J) U_i$$

$$= U_i'(Q \otimes I) U_i - U_i'(R \otimes J) U_i$$

where $Q = \Omega^{-1}$ and $R = (\Omega + T\Delta)^{-1} \Delta \Omega^{-1}$. The first term $U_i'(Q \otimes I) U_i$ can be expressed as

$$U_i'(Q \otimes I) U_i = (\hat{\varepsilon}_i + \hat{\delta}_i \otimes e_T)'(Q \otimes I)(\hat{\varepsilon}_i + \hat{\delta}_i \otimes e_T)$$

$$= \hat{\varepsilon}_i'(Q \otimes I)\hat{\varepsilon}_i + \hat{\varepsilon}_i'(Q\hat{\delta}_i \otimes e_T)$$

$$+ (\hat{\delta}_i' Q \otimes e_T')\hat{\varepsilon}_i + T\hat{\delta}_i' Q \hat{\delta}_i$$

$$= \hat{\varepsilon}_i'(Q \otimes I)\hat{\varepsilon}_i + T \hat{\delta}_i Q \hat{\delta}_i$$

in view of the orthogonality of $\hat{\delta}_i' Q \otimes e_T'$ and $\hat{\varepsilon}_i$. Similarly the second term becomes

$$U_i'(R\otimes J)U_i = (\hat{\varepsilon}_i+\hat{\delta}_i\otimes e_T)'(R\otimes J)(\hat{\varepsilon}_i+\hat{\delta}_i\otimes e_T)$$

$$= \hat{\varepsilon}_i'(R\otimes J)\hat{\varepsilon}_i + T\hat{\varepsilon}_i'(R\hat{\delta}_i\otimes e_T)$$

$$+ T(\hat{\delta}_i'R\otimes e_T')\hat{\varepsilon}_i + T^2\hat{\delta}_i'R\hat{\delta}_i$$

$$= T^2\hat{\delta}_i'R\hat{\delta}_i.$$

Combining these two results, we have

$$U_i'(Q\otimes I)U_i = \hat{\varepsilon}_i'(Q\otimes I)\hat{\varepsilon}_i + T\hat{\delta}_i'(Q-TR)\hat{\delta}_i.$$

Finally,

$$Q-TR = \Omega^{-1} - T(\Omega+T\Delta)^{-1}\Delta\Omega^{-1}$$

$$= (\Omega+T\Delta)^{-1}[(\Omega+T\Delta)\Omega^{-1} - T\Delta\Omega^{-1}]$$

$$= (\Omega+T\Delta)^{-1}$$

which yields the desired result.

<u>Proposition 4</u>. If

$$f(A) = -N\log|A| - \mathrm{tr}(A^{-1}B)$$

where $A = (a_{ij})$ and $B = (b_{ij})$ are positive definite matrices of order p, f(A) assumes the maximum value

$$-N\log\left|N^{-1}B\right| - pN$$

when $A = N^{-1}B$.

Proof. See Anderson (1958), Lemma 3.2.2 and Lemma 3.2.3.

Armed with these results, we see that L^* can now be written as

$$L^* = -n(T-1)\log|\Omega| - n\log|\Omega+T\Delta| - \sum_{i=1}^{n} \hat{\varepsilon}_i'(\Omega^{-1}\otimes I)\hat{\varepsilon}_i$$

$$- T\sum_{i=1}^{n} \hat{\delta}_i'(\Omega+T\Delta)^{-1}\hat{\delta}_i.$$

It follows from Proposition 4 that the maximum likelihood estimate of $\Lambda = \Omega + T\Delta$ is

$$\hat{\Lambda} = Tn^{-1}\sum_{i=1}^{n} \hat{\delta}_i\hat{\delta}_i'.$$

Thus the likelihood equation for Δ is

$$\Delta = n^{-1}\sum_{i=1}^{n} \hat{\delta}_i\hat{\delta}_i' - T^{-1}\Omega.$$

When Δ is determined according to this equation, the likelihood function is

$$L^* = k - n(T-1)\log|\Omega| - \sum_{i=1}^{n} \hat{\varepsilon}_i'(\Omega^{-1}\otimes I)\hat{\varepsilon}_i$$

where k is a function of $\hat{\delta}_i$ and hence β but not a function of Ω. A direct calculation reveals that

$$\sum_{i=1}^{n} \hat{\varepsilon}_i'(\Omega^{-1}\otimes I)\hat{\varepsilon}_i = tr(\Omega^{-1}E)$$

where the typical element of E is

$$E_{jk} = \sum_{i=1}^{n}\sum_{t=1}^{T} \hat{\varepsilon}_{ijt}\hat{\varepsilon}_{ikt}.$$

Hence the likelihood equation for Ω is

$$\Omega = n^{-1}(T-1)^{-1}E.$$

3. **Maximum Likelihood Estimates.**

Collecting the results of the previous section, the maximum likelihood estimates must satisfy three sets of equations:

$$\beta = (\sum_{i=1}^{n} Z_i'\Psi^{-1}Z_i)^{-1}(\sum_{i=1}^{n} Z_i'\Psi^{-1}Y_i)$$

$$\Delta = n^{-1}\sum_{i=1}^{n} \hat{\delta}_i\hat{\delta}_i' - T^{-1}\Omega$$

$$\Omega = n^{-1}(T-1)^{-1}E.$$

The usual generalized least squares estimator provides the first step of an iterative process which, if it converges, will produce maximum likelihood estimates of the parameters. In particular, let $\hat{\beta}(1)$ be the least squares estimator of β. Define $\hat{U}_i(1)$, $\hat{\delta}_i(1)$ and $\hat{\varepsilon}_i(1)$ corresponding to this value of β. Initial estimates of $\hat{\Omega}(1)$ and $\hat{\Delta}(1)$ can now be obtained from the likelihood equations. This yields an initial estimate $\hat{\Psi}(1)$ of Ψ which can be used to obtain $\hat{\beta}(2)$, the generalized least squares estimate of β. This estimate can then be used to define $\hat{U}_i(2)$, $\hat{\delta}_i(2)$, and $\hat{\varepsilon}_i(2)$. Revised estimates $\hat{\Omega}(2)$, $\hat{\Delta}(2)$, and $\hat{\Psi}(2)$ would then be obtained. Continuing in this way until convergence is achieved yields estimates that satisfy the likelihood equations.

References

Amemiya, T.: "The Estimation of the Variances in A Variance Components Model," International Economic Review, 12 (1971), 1-13.

Anderson, T. W.: An Introduction to Multivariate Statistical Analysis, New York: Wiley, 1958.

Avery, R. B.: "Error Components and Seemingly Unrelated Regressions," Econometrica, 45 (1977), 199-209.

Baltagi, B. H.: "On Seemingly Unrelated Regressions with Error Components," Econometrica, 48 (1980), 1547-1551.

Burtless, G. and J. Hausman: "The Effect of Taxation on Labor Supply: Evaluating the Gary Negative Income Tax Experiment", Journal of Political Economy, 86 (1978), 1103-30.

Graybill, F. A.: An Introduction to Linear Statistical Models, Volume I, New York: McGraw-Hill, 1961.

Hill, D. H., R. M. Groves, E. P. Howrey, A. C. Kline, D. F. Kohler, J.S. Lepkowski, and M. A. Smith, "Evaluation of the Federal Energy Administration Load Management and Rate Design Demonstration Projects", Palo Alto, CA: Electric Power Research Institute, 1979.

King, M., "Welfare Analysis of Tax Reforms Using Household Data", NBER Technical Working Paper No. 16, 1981.

Hurwicz, L. and H. Uzawa: "On the Integrability of Demand Functions" in Chipman etal., Preferences, Utility, and Demand, Harcourt Brace Jovanovich, 1971.

Maddala, G. S.: "The Use of Variance Components in Pooling Cross Section and Time Series Data," Econometrica, 39 (1971), 341-358.

Nerlove, M.: "A Note on Error Components Models," Econometrica, 39 (1971), 383-396.

Taylor, L. D., "On Modeling the Residential Demand for Electricity by Time of Day", Journal of Econometrics, 9 (1979), 97-115.

Theil, H.: Principles of Econometrics, New York: Wiley, 1971.

Varian, H., "Notes on Cost-Benefit Analysis", University of Michigan Working Paper, 1979.

Varian, H., Microeconomic Analysis, W. W. Norton and Company, New York, Second edition, 1983.

Wallace, T. D. and A. Hussain, "The Use of Error Components Models in Combining Cross Section With Time Series Data," Econometrica, 37 (1969), 55-72.