Empirical Demand Analysis:
Part-II

Manoranjan Pal

System of Demand Equations

- **System of Demand Equations:**
- Suppose there are n commodities in the market. Then the set of n demand functions

$$y_j = f_j(p_1, p_2, \ldots, p_n, x, u_j); j = 1,2, \ldots, n,$$

- together with the budget equation, which the n demand equations are assumed to satisfy, forms a complete system.
- In such a system, the variables $y_1, y_2, \ldots, y_n$, the quantities consumed of each of the commodities are treated as endogenous variables. The exogenous variables are $p_1, p_2, \ldots, p_n$ and x.
- The estimation of the complete system is important in identifying the interdependence among the goods, especially the effects of changes in price of certain goods on the demand for other goods.

Functional Forms

- Various functional forms are employed in estimating the system. One such functional form is the linear system.

$$\mathrm{y_j = a_j + \sum_k b_{jk} p_k + c_j x + e_j; j = 1,2, \ldots, n,}$$

- and another is the log-linear or constant elasticity system for which

$$\mathrm{Ln(y_j) = a_j + \sum_k b_{jk} Ln(p_k) + c_j Ln(x) + e_j; j = 1,2, \ldots, n.}$$

- Each of these is a straightforward generalization of the corresponding single equation demand function.

Linear Expenditure System

- **Linear Expenditure System:**
- One of the most widely used functional forms is the Linear Expenditure System (LES), which can be written as

$$p_j y_j = p_j y_j^0 + \beta_j \left(x - \sum_k p_k y_k^0 \right) + e_j; j = 1,2, \dots, n,$$

- where $y_j - y_j^0 > 0, 0 < \beta_j < 1, \sum \beta_j = 1.$
- This system can be interpreted as follows:
- Expenditure on good j, given as $p_j y_j$, can be decomposed into two components. The first is the expenditure on a certain "base amount" y_j^0 of good j, which is the minimum expenditure to which the consumer is committed. The second is a fraction β_j of the so called "supernumerary income", defined as the income above the "subsistence income" $\sum_k p_k y_k^0$ needed to purchase base amount of all goods. These two components correspond, respectively, to committed and discretionary expenditures on good j.

Properties of LES

- Dividing through the LES by the price p_j gives the corresponding system of demand equations.

$$y_j = y_j^0 + \frac{\beta_j}{p_j}\left(x - \sum_k p_k y_k^0\right) + \frac{e_j}{p_j};$$

- Which is hyperbolic in own price and linear in income.

- The LES is widely used for three reasons. First, it has a straightforward and reasonable interpretation as given above. Second, it is one of the systems that automatically satisfies all the theoretical restrictions discussed before on system of demand equations, namely budget constraint, homogeneity, AUC and Cournot aggregation conditions. This also satisfies Slutsky negativity and Slutsky symmetry conditions. Third, it can be derived from a specific utility function known as "Stone-Geary utility function" or "Klein-Rubin utility function" defined as

$$U = \sum \beta_j \mathrm{Ln}\left(y_j - y_j^0\right), \qquad y_j > y_j^0, 0 < \beta_j < 1, \sum \beta_j = 1.$$

- **Home Task:** Prove it.

Estimation of Parameters of LES

- **Estimation of Parameters of LES:**
- The system is estimated from data on quantities y_j and prices p_j of the n goods and data on income x (or total expenditure). The parameters that are estimated are the n base quantities $y_1^0, y_2^0, \dots, y_n^0$ and the n marginal shares $\beta_1, \beta_2, \dots, \beta_n$.
- The estimation of the LES presents certain complications because, while it is linear in variables, it is non-linear in the parameters, involving the product of β_j and each y_k^0. There are in fact several approaches to the estimation of the system.
- One approach determines the base quantities y_k^0 on the basis of extraneous information or prior judgements. LES then implies that expenditure on each good in excess of the base expenditure $(p_j y_j - p_j y_j^0)$ is a linear function of supernumerary income, so each of the marginal budget shares β_j can be estimated using the usual single equation simple linear regression methods (without constant term).

Estimation of Parameters of LES (Continued)

- A second approach reverses this procedure by first determining the marginal budget shares β_j on the basis of extraneous information or prior judgements. It then estimates the base quantities y_k^0 by estimating the system in which the expenditure less the marginal budget share times income $(p_j y_j - \beta_j x)$ is a linear function of all prices.
- A third approach is an iterative one, using the estimates of the β_j conditional on y_k^0 (as in the first approach) and the estimates of the y_k^0 conditional on β_j (as in the second approach) iteratively so as to minimize the total sum of squares. The process would continue until convergence to the desirable level.
- A fourth approach selects β_j and y_k^0 simultaneously by setting up a grid of possible values for the (2n – 1) parameters (– 1, because the β_js sum to unity) and obtaining the point on the grid when the total sum of squares over all goods and all observations is minimized.

Difficulties in Estimation

- There are several difficulties in actually estimating system of demand equations, such as the LES.
- One such difficulty is the multicollinearity among the prices, which all tend to move together. This difficulty is partly counter balanced by the constraints imposed on the system theory.
- Three additional difficulties, which apply both to individual demand equations and systems of demand equations, are identification, aggregation and dynamic factors.

Effect of Household Size and Composition

- **Effect of Household Size and Composition on Household Consumption Pattern:**
- We have already seen that per-capita formulation of Engel Curve is written as

$$\frac{y}{n} = f\left(\frac{x}{n}\right),$$

- where
  - y: household consumption of a specific item
  - x: household total consumption/income
  - n: household size
- Instead of the above formulation we can take

$$y = ax^{b}n^{\gamma}. \qquad \ldots \text{(I)}$$

- or

$$\frac{y}{n} = a\left(\frac{x}{n}\right)^{b} n^{\delta}. \qquad \ldots \text{(II)}$$

Economies and Diseconomies of Scale

- By comparing (I) and (II) one gets $\delta = \gamma + b - 1$. Hence,
  - $\gamma + \mathrm{b} > 1 \Rightarrow$ diseconomy of scale
  - $\gamma + \mathrm{b} < 1 \Rightarrow$ economy of scale
  - $\gamma + \mathrm{b} = 1 \Rightarrow$ independent of scale
- or equivalently
  - $\delta > 0 \Rightarrow$ diseconomy of scale
  - $\delta < 0 \Rightarrow$ economy of scale
  - $\delta = 0 \Rightarrow$ independent of scale
- Estimation of the parameters can be made by taking logarithmic transformation of equation (I).

$$\mathrm{Ln}(\mathrm{y}) = \mathrm{a}' + \gamma \mathrm{Ln}(\mathrm{n}) + \mathrm{b}\mathrm{Ln}(\mathrm{x}),$$

- where $\mathrm{a}' = \mathrm{Ln}(\mathrm{a})$.
- If consumption increases less/more than proportionately as n increases, then we say that there is economy/diseconomy of scale in the consumption.

Estimation when the Parameters Vary with Household Size

- In the above formulation we have assumed b to be same for all n. This may be relaxed given sufficient number of observations.
- We write the equation

$$\mathrm{Ln(y)} = \left(\mathrm{a}' + \gamma \mathrm{Ln(n)}\right) + \mathrm{bLn(x)}$$

- as

$$\mathrm{Ln(y)} = \mathrm{a_n} + \mathrm{b_n Ln(x)}.$$

- $\mathrm{b_n}$ is the income elasticity of the household with size n. Given sufficient sample size we can run the regression with data for each household size and estimate $\mathrm{a_n}$ and $\mathrm{b_n}$. In fact, it is also possible to test whether $\mathrm{a_n}$ values and/or $\mathrm{b_n}$ values are same for all household size (Use dummy variables for both intercepts and slopes).

Reasons for economy of scale

- Reasons for economy of scale:
- (i) Bulk purchasing facilities.

  (The consumer may be offered the good at a lower price if a large quantity is purchased.)
- (ii) Economies in consumption itself.

  (Electricity consumption may not increase much if a new member enters in the household.)
- We have the concept of economy/diseconomy of scale for a single commodity as well as for the total household consumption.

Prais-Houthakker Formulation

- Assume that a household with size n has an effective size n^{θ_i} with respect to the specific commodity i and n^{θ_0} with respect to the total consumption due to the economy/diseconomy of scale of that commodity or total consumption. $\theta_i \lesseqgtr 1$ according as economy/diseconomy of scale. Similarly, $\theta_0 \lesseqgtr 1$ according as economy/diseconomy of scale for total consumption.
- In the study of Cross Section analysis Prais and Houthakker formulated

$$\frac{y_i}{n^{\theta_i}} = f\left(\frac{x}{n^{\theta_0}}\right),$$

- where

  y_i = household consumption of the ith commodity

  x = household total consumption

  n = household size

  and θ_i and θ_0 are parameters.

PH Formulation: Interpretations

- $\frac{\theta_i}{\theta_0}$ reflects the economy or diseconomy of scale for the ith commodity compared to that of the total consumption. Initially, we see that θ_0 and $\theta_1, \theta_2,$ …, θ_K are related, because $\sum y_i = x$ (budget constraint).
- We make the following two assumptions:

$$(i) \sum_{1}^{K} \frac{\partial y_i}{\partial x} = 1,$$

- i.e., increase in income is distributed through consumption of commodities if n is fixed.

$$(ii) \sum_{1}^{K} \frac{\partial y_i}{\partial n} = 0,$$

- i.e., if the total consumption is fixed, then the increase in n will imply only redistribution of consumption over commodities. For increase in n, consumptions of necessary commodities will increase, and luxury commodities will decrease, keeping the sum of all the changes to zero.

PH Formulation: Restrictions

- We can rewrite the two restrictions as

$$\sum \mathrm{w_i}\eta_{\mathrm{ix}} = 1 \text{ and } \sum \mathrm{w_i}\eta_{\mathrm{in}} = 0,$$

- where $\mathrm{w_i} = \mathrm{y_i}/\mathrm{x}$ and η_{ix} and η_{in} are the elasticities of household's expenditure on commodity i with respect to x and n respectively. Recall that

$$\sum_{1}^{\mathrm{K}} \frac{\partial y_i}{\partial n} = 0 \Rightarrow \sum_{1}^{K} \frac{\partial y_i}{\partial n} \frac{n}{y_i} \frac{y_i}{n} \frac{x}{x} = 0 \Rightarrow \frac{x}{n} \sum w_i \eta_{in} = 0, and$$

$$\sum \frac{\partial y_i}{\partial x} = 1 \Rightarrow \sum \frac{\partial y_i}{\partial x} . \frac{x}{y_i} . \frac{y_i}{x} = 1 \Rightarrow \sum \frac{y_i}{x} \eta_{ix} = 1 \Rightarrow \sum w_i \, \eta_{ix} = 1.$$

Restriction on Parameters

- Now, since

$$y_i = n^{\theta_i} f\left(\frac{x}{n^{\theta_0}}\right),$$

$$\eta_{ix} = \frac{x}{n^{\theta_i} f\left(\frac{x}{n^{\theta_0}}\right)} . n^{\theta_i} f'\left(\frac{x}{n^{\theta_0}}\right) \frac{1}{n^{\theta_0}} = \frac{x}{n^{\theta_0}} \frac{f'\left(\frac{x}{n^{\theta_0}}\right)}{f\left(\frac{x}{n^{\theta_0}}\right)}, \text{and}$$

$$\eta_{in} = \frac{n}{n^{\theta_i} f\left(\frac{x}{n^{\theta_0}}\right)} . \left[\theta_i n^{\theta_i - 1} f\left(\frac{x}{n^{\theta_0}}\right) + n^{\theta_i} f'\left(\frac{x}{n^{\theta_0}}\right) x(-\theta_0) n^{-\theta_0 - 1}\right]$$

$$= \theta_i - \frac{x}{n^{\theta_0}} \frac{f'\left(\frac{x}{n^{\theta_0}}\right)}{f\left(\frac{x}{n^{\theta_0}}\right)} \theta_0.$$

- Thus, $\boxed{\boldsymbol{\eta_{in} = \theta_i - \eta_{ix}\theta_0}.}$

Restriction on Parameters (Continued)

$$\sum w_i \eta_{\mathrm{in}} = \sum w_i \theta_{\mathrm{i}} - \sum w_i \eta_{\mathrm{ix}} \theta_0$$

$$= \sum w_i \theta_{\mathrm{i}} - \left(\sum w_i \eta_{ix} \right) \theta_0.$$

$$= \sum w_i \theta_{\mathrm{i}} - \theta_0, \text{since} \sum w_i \eta_{ix} = 1.$$

- Again, since $\sum w_i \eta_{in} = 0$, we have

$$\theta_0 = \sum w_i \theta_{\mathrm{i}} \, .$$

- The restriction can also be written as

$$1 - \theta_0 = \sum w_i \,(1 - \theta_i), \text{since} \sum w_i = 1.$$

Estimation

- The PH formulation is

$$\frac{y_i}{n^{\theta_i}} = f\left(\frac{x}{n^{\theta_0}}\right).$$

- We need to take specific form and the form taken by them is

$$\frac{y_i}{n^{\theta_i}} = a_i\left(\frac{x}{n^{\theta_0}}\right)^{b_i},$$

$$or\ y_i = a_i x^{b_i} n^{\theta_i - b_i\theta_0},$$

$$or\ y_i = a_i x^{b_i} n^{c_i},$$

$$where\ \ c_i = \theta_i - b_i\,\theta_0.$$

- Once we estimate a_i, b_i and c_i, there is no way that we can separate out θ_i and θ_0 from $\theta_i - b_i\,\theta_0$ unless θ_0 or any of θ_i's is known. Thus, this form seems to be inconvenient for estimation purpose.

PH Method of Estimation

- **Prais-Houthakker Method of Estimation:**
- Prais-Houthakker devised an ingenious method of estimating the parameters of the model. They assumed that quality of the commodity consumed by the household depends on effective per-capita total expenditure of the household, i.e., $\frac{x}{n^{\theta_0}}$. They defined P_n^j, the amount that the jth household pays with respect to an average household, as

$$P_n^j = \frac{\sum_i P_{ji}^{(n)} Q_{0i}}{\sum_i P_{0i} Q_{0i}},$$

- where i is the suffix for commodity, j represents the household and
- P_{0i} is the price of the commodity i paid by an average household,
- Q_{0i} is the quantity of the commodity i consumed by an average household, and
- $P_{ji}^{(n)}$ is the price of commodity i paid by jth household of size n.
- P_n^j is a kind of index representing quality of the commodities purchased by the jth household.

PH Method of Estimation (Continued)

- PH related P_n (j is suppressed for convenience) to $\frac{x}{n^{\theta_0}}$ and formulated a quality equation as

$$P_n = \alpha + \beta Ln\left(\frac{x}{n^{\theta_0}}\right).$$

- or

$$P_n = \alpha + \beta Ln(x) - \beta\theta_0 Ln(n).$$

- The values of the regressors and the regressand vary over the observations j (household) and n (household size). However, the observational units are the households. Once household is selected, n is automatically determined. P_n is calculated for each household as defined above. Thus, one can regress P_n on $Ln(x)$ and $Ln(n)$ to get the estimates $\hat{\alpha}$, $\hat{\beta}$ and $\widehat{\beta\theta_0}$ and then get the estimate of θ_0 as

$$\widehat{\theta_0} = \frac{\widehat{\beta\theta_0}}{\hat{\beta}}.$$

PH Method of Estimation (Continued)

- Let us now recall that we need to estimate a_i, b_i and θ_i from a_i, b_i and c_i using the equation

$$y_i = a_i x^{b_i} n^{c_i}, \qquad \text{where } c_i = \theta_i - b_i\,\theta_0.$$

- Taking logarithmic transformation, one gets

$$Ln(y_i) = Ln(a_i) + b_i Ln(x) + c_i Ln(n).$$

- This is a single equation method, in which we are interested in the estimation of the parameters for the ith commodity only and the observational units are again the households. Thus, we regress $Ln(y_i)$ on $Ln(x)$ and $Ln(n)$ to get

$$\widehat{Ln(a_i)}, \hat{b}_i \text{ and } \hat{c}_i,$$

- From which $\hat{a}_i, \hat{b}_i$ and $\hat{\theta}_i$ can easily be found since

$$\hat{c}_i = \hat{\theta}_i - \hat{b}_i\widehat{\theta_0},$$

- and $\widehat{\theta_0}$ is already estimated.

Thank you