Cowles Foundation Paper 26b
Reprinted from Review of Economic Studies, 1947

Some Implications of "Linearity."[1]

Statisticians often make the assumption that various economic relations can be approximated, within a given range of observable data, by linear relationships. This is a perfectly proper procedure. As applied to the measurement of demand for good X_i in terms of prices $P_1, \ldots, P_n$ and income, r, this involves the customary Taylor's series approximation around any given point $(X_1^0, X_2^0, \ldots, X_n^0 \, ; \, P_1^0, P_2^0, \ldots, P_n^0, r^0)$; or $X_i = X_i^0 + \left(\frac{\partial X_i}{\partial P_1}\right)^0 (P_1 - P_1^0) + \ldots + \left(\frac{\partial X_i}{\partial P_n}\right)^0 (P_n - P_n^0) + \left(\frac{\partial X_i}{\partial r}\right)^0 (r - r^0) +$ remainder. So long as the price and income changes are small, the remainder error term can often be neglected. The change in the cost of living, r, along an initial level of satisfaction, U^0, can be written down to an even higher order of approximation as :

$$r = r^0 + \sum_i X_i \; (P_i - P_i^0) + \frac{1}{2} \sum_j \sum_k \left(\frac{\partial X_k}{\partial P_j}\right)_{U^0} (P_k - P_k^0) \; (P_j - P_j^0) + \text{remainder}$$

where $\left(\frac{\partial X_k}{\partial P_j}\right)_{U^0} = \left(\frac{\partial X_j}{\partial P_k}\right)_{U^0} = \frac{\partial X_j}{\partial P_k} + X_k \, \frac{\partial X_j}{\partial r}$ are the Slutsky-Hicks compensated substitution terms. Without making stronger assumptions no better result can be stated.

No empirical limitations are implied by the above kind of " local " linearity assumption. The parameters $(\partial X_i / \partial P_j)^0$ and $(\partial X_i / \partial r)^0$ are actually functions of the initial point (X^0, P^0, r^0).

But the assumption[2] that :

$$X_i = \beta_i \; \gamma_1 \, \frac{P_1}{P_i} + \ldots + \beta_i \; \gamma_n \, \frac{P_n}{P_i} - \gamma_i + \beta_i \, \frac{r}{P_i}$$

is not a local linearity assumption. It has numerous restrictive empirical implications ; indeed, it is only because of these implications that the particular formula for measuring the cost of living emerges.

Some of the empirical implications may be briefly indicated. Suppose that at some point $\partial X_1 / \partial P_2$ and $\partial X_1 / \partial P_3$ were evaluated. Then instead of the ratio $(\partial X_4 / \partial P_2) \div (\partial X_4 / \partial P_3)$ being a variable to be determined by the facts, it and many other ratios would be frozen. Similarly, a person's cross-elasticity of demand, when poor, would have to be the same as when he is rich. Numerous other implied strait-jackets on the facts could be pointed out.

However, in this case, the full empirical implications can be neatly summarized in the following stringent empirical assumption :

(*a*) The consumer is assumed always to buy a *necessary set* of goods, $(\overline{X}_1, \overline{X}_2, \ldots, \overline{X}_n) = (-\gamma_1, -\gamma_2, \ldots, -\gamma_n)$;

(*b*) Any algebraic income left over after buying this necessary set is spent in constant proportions $(\beta_1, \beta_2, \ldots, \beta_n = 1 - \beta_2 - \ldots)$ on the various goods.

[1] Included in Cowles Commission Paper, New Series, No. 26.

[2] This is the fundamental equation of the Klein-Rubin paper in this issue. Together with the familiar Slutsky relations and the requirement that the demand functions be homogeneous of order zero in all prices and income, their strong linearity assumption—that each good is a strictly linear function of the ratio of other prices to its price and of income to its price—implies this fundamental equation.

This follows from rewriting our equation:

$$X_i = \Sigma\ \beta_i\ \gamma_j\ \frac{P_j}{P_i} - \gamma_i - \beta_i\ \frac{r}{P_i}$$

as

$$X_i = \bar{X}_i + \frac{\beta_i\ (r - \Sigma\ P_k\ \bar{X}_k)}{P_i}$$

where $-\gamma_i$ has been replaced by $\bar{X}_i$ and where the expression in parenthesis is "supernumerary" income remaining after buying the necessary batch $(\bar{X}_1, \ldots, \bar{X}_n)$.

Immediately this explains the intuitive meaning of the cost-of-living formula. By itself, the necessary batch would imply a simple summation cost-of-living formula: $\Sigma\ P_i\bar{X}_i = -\Sigma\ P_i\gamma_i$. The proportionality of expenditure assumption, by itself, is known to lead to a weighted geometric mean formula for the cost of living: $\Pi P_i{}^{\beta_i}$. Therefore, we get in this case a linear combination of the two, with the "constant of integration," C, being determined by the relative importance of necessary expenditure and supernumerary income.

By working with new variables $X_i' = X_i - \bar{X}_i$, we also quickly see that the ordinal utility or preference field must be of the form:

$$U = F\ [\beta_1 \log X_1' + \ldots + \beta_n \log X_n']$$

$$= F\ [\beta_1 \log (X_1 - \bar{X}_1) + \ldots + \beta_n \log (X_n - \bar{X}_n)]$$

where F is any function with $F' > 0$.

This is a highly restrictive empirical assumption, indeed. Such facts as are available concerning budgetary family data do not appear to be in accord with this hypothesis. Thus, on an indifference curve diagram, we do not find income expansion paths that are straight lines converging on a single focal point, as this theory requires.

In conclusion, the question may be asked as to the empirical observations needed to refute or verify the non-local linearity hypothesis under discussion. What kind and how many observations are needed to determine the n constants $\bar{X}_1, \ldots, \bar{X}_n$ and the $(n - 1)$ constants $(\beta_1, \ldots, \beta_{n-1})$? Each single price-quantity observation seems to imply $(n - 1)$ independent implicit equations for these parameters. One might hope, therefore, that just more than two independent price and income situations would be sufficient.

However, the problem is not this simple. The P's and γ's, taken together, are not involved in a linear fashion, since cross-products appear. A full answer would involve mathematical reasoning of more complexity. But whatever the requisite number of observations, be it 3, 4, or more, one single additional observation could be used to check the hypothesis.

As a matter of fact, without solving the mathematical question of the minimum number of observations needed to determine the parameters, we can still state various special ways of refuting the hypothesis in question. Thus, along any constant-price budget path, the β's are simply the marginal propensities to consume the different goods, and we can easily check on their constancy by means of three or more observations. With two points on each of a number of different budget lines, we can also check upon the constancy and uniformity of the β's. With pairs of points from three or more budget lines, we can check on the existence of a focal point of convergency of budgetary expansion paths.

If two points on the same budget path are available, it is obvious that the β's are all determined, and the $\bar{X}$'s are left indeterminate only upon a straight line, with only one degree of freedom left. A third point along the same straight line will not add to our information. But a single observation on a different budget line will, in a simple two commodity world, be just sufficient to determine the remaining degree of freedom of the X's. If there are more than two commodities, the problem is over-determined; hence, with two points on one budget line and with a third different point, we could conveniently refute the assumed hypothesis of "linearity."

An important word of qualification should be added. All the above disregards the inevitable stochastical "errors" present in any empirical situation. Only if each observation were perfectly "exact" could the minimum number of observations described above be sufficient to test our hypothesis. No one would be so foolish as to think that three observations could tell us much about anything. In practice we should have to have numerous repeated independent observations in order to have confidence in our empirical inferences."

Cambridge, Mass. P. A. SAMUELSON.