CHAPTER 5

Restrictions on preferences

In Part I, we saw how the theory of choice, initially developed for application to the allocation of a fixed total expenditure over a number of goods, could be extended to deal with labor supply, the allocation of income between saving and consumption, and the purchase of durable goods. So far, we have looked at these extensions as separate problems, but in principle, each consumer has to deal with all of them simultaneously. At any given time, current assets and current and future income must be allocated over nondurable and durable goods for current and future periods, while for consumers who are free to do so, plans must be made for allocating time between work and leisure in the present and the future. There are also the complications of choosing types of real and financial assets as well as the problems of dealing with uncertainty. All of the parts of this allocation problem may interact: changes in future wage rates may alter retirement plans with important consequences for durable good purchases now; an anticipated increase in asset values three years hence may cause an elated investor to buy more of his favorite dinner, and so on. Such interactions pose formidable problems, not only for the consumer but also for the economist who attempts to describe consumer behavior. If making today's purchases requires knowledge of the price of bootlaces 30 years from now, daily living is close to impossible. Nor is there much hope of predicting behavior if all such possible interactions must be allowed for. It is thus important to find ways in which the problem can be simplified, either by aggregation, so that whole categories can be dealt with as single units, or by separation, so that the problem can be dealt with in smaller, more manageable units. For example, by making total consumption expenditure rather than its components a function of wealth, prices, and interest rates, we are aggregating; similarly in Chapter 3, expenditures on individual goods were aggregated into broad groups and these were related to group price indices and total expenditure. This latter also involves a separation of decision making since the group expenditures are related, not to their ultimate determinants (assets, wage rates, prices, interest rates) but to total expenditures within the current period. Hence, it is being assumed that the decision of how to allocate total current expenditure into various broad categories of goods can be made separately from the decision of how to arrange the intertemporal flow of expenditure. Similar assumptions about decision making are entailed when labor supply is analyzed independently of the structure of commodity demands.

There is a variety of ways in which assumptions of this type can be justified. In particular, we can place restrictions on preferences that allow separable decision making or we can restrict the allowable range of exogenous variables, particularly prices. The various possibilities and combinations are discussed in §5.1 and §5.2. The first of these is informal and presents an overview of the more detailed and more difficult material in §5.2. At first reading, the account in §5.1 of the composite commodity theorem, of two-stage budgeting, of utility trees, and of the types of preference separability will give sufficient knowledge of the more important results. Section 5.2, although still nonrigorous, goes further as well as outlining some of the derivations.

Sections 5.3 and 5.4 explore further types of restrictions on preferences. It is frequently important to know what structure of preferences will allow some property of interest to hold. Similarly, we must be aware of the empirical consequences of particular assumptions about preferences so that plausible restrictions can be used to generate degrees of freedom in econometric work while implausible ones are avoided. Section 5.3 examines strong separability or want independence; this has been heavily used in econometric work, often without a full appreciation of its consequences. Section 5.4 discusses the case of homothetic preferences which have the consequence that expenditure proportions are independent of outlay. This case is extremely easy to handle from a theoretical point of view and, as we shall see later in the book, will repeatedly appear as a necessary condition for several sorts of seemingly attractive results. However, homotheticity is of very limited applicability in econometric work, and we shall examine the important generalization that leads to linear Engel curves.

5.1 Commodity groups and separability: an elementary overview

In this and the next section, we shall revert to the model of Chapter 2, the allocation of a predetermined total over n goods, for most of the formal discussion. However, if the model is extended as in Chapter 4 to include n goods in each of N periods plus durable goods and labor supply, the analysis carries through. We shall indicate the important applications as we proceed, leaving proofs to be completed by analogy or through exercises.

Commodity groups: the composite commodity theorem

The first set of conditions for the existence of commodity aggregates we owe to Hicks (1936) although an essentially similar result was proved by Leontief

(1936). This is the *composite commodity theorem,* which asserts that if a group of prices move in parallel, then the corresponding group of commodities can be treated as a single good. We illustrate with a three-good model in which two prices always move in proportion. Write the three prices p_1, p_2, p_3, and assume that p_2 and p_3 bear some fixed ratio θ to some base period prices p_2^0 and p_3^0; that is,

$$p_2 = \theta p_2^0, \quad p_3 = \theta p_3^0 \tag{1.1}$$

where θ varies with time but is common to both prices so that the ratio p_2/p_3 remains fixed at p_2^0/p_3^0. The obvious possibility is that θ can act as a "price" for a new combined group with a "quantity" defined by weighting the individual quantities using the base-period prices p_2^0 and p_3^0. To see that this is so, define the composite quantity $p_2^0q_2 + p_3^0q_3$. The cost function $c(u, p_1, p_2, p_3)$ can be written $c(u, p_1, \theta p_2^0, \theta p_3^0)$ which, since p_2^0 and p_3^0 are fixed can be thought of as a function of u, p_1, and θ alone and this, say, $c^*(u, p_1, \theta)$ is the grouped cost function, that is,

$$c^*(u, p_1, \theta) = c(u, p_1, \theta p_2^0, \theta p_3^0) \tag{1.2}$$

It is left as an exercise to show that $c^*(u, p_1, \theta)$ satisfies all the properties of a "proper" cost function; it is increasing in u, p_1, and θ, homogeneous of the first degree and concave in p_1 and θ. Further, if we differentiate $c^*(u, p_1, \theta)$ with respect to θ, we have

$$\frac{\partial c^*}{\partial \theta} = \frac{\partial c}{\partial p_2} \cdot \frac{\partial p_2}{\partial \theta} + \frac{\partial c}{\partial p_3} \cdot \frac{\partial p_3}{\partial \theta} = p_2^0q_2 + p_3^0q_3 \tag{1.3}$$

Hence, $p_2^0q_2 + p_3^0q_3$ is confirmed as the quantity of the composite commodity corresponding to the price θ. Since the cost function provides a complete picture of preferences, this demonstration shows that, provided (1.1) holds, new preferences can be defined over q_1 and $p_2^0q_2 + p_3^0q_3$ and that these preferences lead to the same choices as the original ones. The extension to groupings of any size within a total of n commodities is straightforward, see Exercise 5.2.

The usefulness of this theorem in constructing commodity groupings for empirical analysis is likely to be somewhat limited. If we take the view that relative prices are largely independent of the pattern of demand, at least in the long run, then commodity groups should be chosen so that close substitutes in production are grouped together. For example, if two food crops require similar ratios of land, labor, and fertilizer and if progress in developing improved seed strains occurs at a similar rate, we should not expect sharp fluctuations in their relative prices. A fuller treatment of relative price constancy is contained in Exercises 5.7 and 5.8. On this basis, it might be possible to construct a relatively small number of aggregates. However, in an open economy with a floating exchange rate, considerable fluctuation in relative prices can be expected and even without this, it is not clear that we could justify the type of aggregates that are usually available. For example, the fact that the price of fish is relatively volatile would prevent its classification with other foods, and we should have to recognize that the definition of aggregates would shift with institutional changes such as alterations in tariffs or internal government policies. Fewer difficulties occur when we consider the aggregation of all purchases within each period to give total consumption as considered in §4.2. When making intertemporal choices, consumers are unlikely to have detailed information or expectations about changes in relative prices in future periods. There is thus some merit in the simplistic view that in the future, relative prices will remain unchanged, so that only variations in the expected absolute price level are accounted for. This would then allow one price and one composite commodity for each future period so that the analysis of the consumption function can proceed as in §4.2. For our purposes, this is perhaps the most useful application of the theorem.

The utility tree and two-stage budgeting

If we cannot rely on an external factor, the constancy of relative prices, to define commodity groups, we must instead ask whether or not preferences themselves might not provide some natural structuring of commodities. The first important idea in this context is that of *separability* of preferences. If this holds, the commodities can be partitioned into groups so that preferences within groups can be described independently of the quantities in other groups. For example, if food is a group, the consumer can rank different food bundles in a well-defined ordering which is independent of his consumption of housing, fuel, entertainment, and everything else outside the group. By the arguments of §2.1, this implies that we can have subutility function for each group and that the values of each of these subutilities combine to give total utility. For example, suppose there are six goods, q_1 and q_2 are foods, q_3 and q_4 are housing and fuel, and q_5 and q_6 are TV and watching sports. Then if separable groups, food, shelter, and entertainment are formed, the utility function can be written

$$u = v(q_1, q_2, q_3, q_4, q_5, q_6) = f[v_F(q_1, q_2), v_S(q_3, q_4), v_E(q_5, q_6)] \qquad (1.4)$$

where $f(\ )$ is some increasing function and v_F, v_S, and v_E are the subutility functions associated with food, shelter, and entertainment, respectively. There is no reason why each subutility function could not have one or more deeper subgroupings within it, nor should we rule out the possibility that some group may only contain one good. If we put all this together, we get the *utility tree;* the individual commodities are the outermost twigs that join together to form branches that, in turn, join up to form the tree. One possibility is illustrated in Figure 5.1.

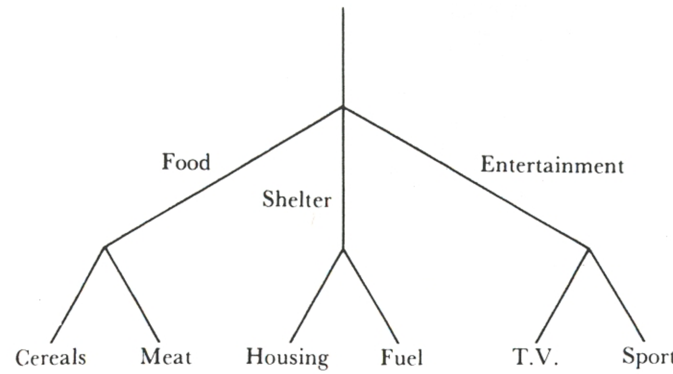

Figure 5.1. A possible utility tree.

The diagram suggests the second important idea, that of *two-stage budgeting*. This occurs when the consumer can allocate total expenditure in two stages; at the first or higher stage, expenditure is allocated to broad groups of goods (food, shelter, and entertainment in Figure 5.1), while at the second, or lower stage, group expenditures are allocated to the individual commodities. At each of these stages, information appropriate to that stage only is required. At the first stage, allocation must be possible given knowledge of total expenditure and appropriately defined group prices, while at the second stage, individual expenditures must be functions of group expenditure and prices within the group only. Both of these allocations have to be perfect in the sense that the results of two-stage budgeting must be identical to what would occur if the allocation were made in one step with complete information. This is the basic definition of two-stage budgeting, and it can be extended and modified in a number of ways. One possibility is the requirement that the decision at each stage can be thought of as corresponding to a utility maximization problem of its own. For example, different foods would be chosen so as to maximize a food subutility function subject to a food budget constraint, while the full utility function would determine the allocation to food, clothing, fuel, and so on. Note that for this latter to work we must be able to define appropriate price and quantity indices for the groups so that the synthetic utility maximization procedure can be defined. A weaker form of two-stage budgeting is to require only that once an optimal allocation has been reached, it can be maintained in the face of changes in prices and total expenditure. Note that two-stage budgeting involves both aggregation (to construct the broad groups) and separable decision making (for each of the group subproblems).

The two ideas, of separability of preferences and of two-stage budgeting, are intimately related to one another but are by no means equivalent; it is *not true* that either one implies the other. What is true, however, is that separability as illustrated in (1.4), often called *weak separability*, is both necessary and sufficient for the *second* stage of two-stage budgeting. If any subset of commodities appears only in a separable subutility function, then quantities purchased within the group can always be written as a function of group expenditure and prices within the group alone. This can be seen directly from structures such as (1.4); the maximization of utility in (1.4) must imply that v_F, v_S, and v_E are each maximized subject to whatever is spent on food, shelter, and entertainment. If this were not so, v_F, v_S, or v_E could be increased without violating the budget constraint, so that, since f is increasing in its arguments, utility cannot be maximal. Hence, the expenditures on q_1 and q_2 are the outcome of maximizing $v_F(q_1, q_2)$ subject to $p_1q_1 + p_2q_2 = x_F$, the total expenditure on food, so that we can write

$$q_i = g_{Fi}(x_F, p_1, p_2), \qquad i = 1, 2 \tag{1.5}$$

for the Marshallian subgroup demands. Such demand functions possess all the usual properties of demand functions since they derive from a standard utility maximizing problem. The converse of the result, that the existence of subgroup demand functions as in (1.5) implies weak separability, is also true but is more difficult to prove: an outline is given in the next section.

If we temporarily leave aside the question of the first stage of two-stage budgeting, how useful is the result linking weak separability to the second stage? First, the separability of preferences imposes restrictions on behavior that limit the possible substitution effects between goods in different groups. We shall explore these restrictions further in the next section but the principle is clear enough from Figure 5.1. Apart from income effects, a change in the price of (say) cereals, can only affect the demand for fuel through the same channels and thus in the same way as will a change in the price of any other commodity in the food branch. This is a natural enough restriction, however, and will be satisfied if goods which bear special relationships to one another in consumption either as substitutes or complements are always kept in the same group. Given this, the separability result is important in at least four areas.

First, we may wish to assume that preferences are weakly *intertemporally* separable so that goods in each period form a closely related group with only general relations between periods. The plausibility of this assumption depends very much on the time period adopted, lunch now is rather closely substitutable with lunch twenty minutes hence. Even with longer periods, some goods, holidays, or units in an economics course, may be closely related to one another across time periods thus violating intertemporal separability. Conversely, the adoption of the assumption allows the expression of commodity demands in each period as a function of total outlay and prices in that period alone in accordance with usual practice as discussed in Chapter 3. Note too

that there is no other way to justify this; regressing demands on current variables rules out any specific interaction between commodities consumed in different time periods.

Second, if leisure is weakly separable from goods, the allocation of total expenditure is independent of the decision on hours. This is clearly implausible for goods that are closely complementary (recreational goods, TV, etc.) or substitutable (travel to work, business lunches, etc.) with leisure, but may nevertheless be acceptable for the bulk of consumers' expenditure. However, since second-stage budgeting implies weak separability as well as the other way around, the allocation of *all* expenditure without regard to hours worked is only valid if *all* goods are separable from leisure.

The third important application is the obvious one; detailed commodity expenditure can be related to group outlay and prices alone. This has obvious econometric advantages since it is possible to find an explanation for behavior through a much smaller number of variables.

Fourth, and perhaps least obvious, is the application to rationing. If any good or group of goods is rationed and if some other group of goods is separable from the rationed commodity or commodities, then the only effect of the rationing on the goods in the separable group is through total group expenditure. Consequently, if only one good is rationed and all other goods are separable from it, the expenditure needed to buy the ration is simply deducted from total expenditure and the remainder is allocated amongst the other goods independently of the amount of the ration. The most important example here is leisure. If a consumer has no choice over hours worked and if goods are weakly separable from leisure, the spending of the resulting income is explicable without reference to the number of hours actually worked. The converse is just as important. If goods are not weakly separable from leisure and if the consumer is constrained in hours worked, hours worked must appear as an exogenous argument in the allocation of expenditures. These propositions and closely related arguments play an important part in the rest of this book.

The first stage of two-stage budgeting, the allocation of total expenditure into broad groups using group price indices, is more problematical than the second stage. An *exact* solution that does not require the composite commodity theorem demands conditions considerably stronger (and less plausible) than weak separability alone. Nevertheless, as we shall see in the next section, a number of useful approximations are available so that the usual (and largely inevitable) practice of working with broad groups may not involve too great an error. Even so, the existence of such key concepts as the consumption function depends upon the ability to define such aggregates, so it is worthy of note that there appear to be no theoretical assumptions that are not manifestly implausible and that will guarantee *exactly* the treatment of consumption as a single good with a single price. We shall return to this topic in Chapter 12.

Finally, while leaving the details to the next section, it must be emphasized that there are many types of separability other than the "weak" separability just discussed and that most of these different types imply or are implied by different types of two-stage budgeting. The cataloging of all the possibilities is well beyond our present purpose and is well-documented elsewhere, see Blackorby and others (1978), and only one other type is dealt with here. This is *implicit* or *quasi* separability under which not the quantities in the direct utility function but the prices in the cost function are broken up into separable groups. Implicit separability also allows a form of two-stage budgeting whereby the first stage is accomplished perfectly using group price indices that are functions of the utility level, while, at the second stage, the fractions of group outlay going to each good are a function of total utility and in-group prices alone. This structure has been relatively little used either theoretically or empirically so far, but we feel it is of sufficient potential importance to be included in some detail in the next section.

Exercises

5.1. Prove that $c^*(u, p_1, \theta)$, as defined in (1.2), is homogeneous of degree one and concave in p_1 and θ.

5.2. Assume that the price vector p can be partitioned in the form $p = (p_1, \ldots, p_G, \ldots, p_N)$, where each p_G is a subvector of prices. Assume, for all G, $p_G = \lambda_G p_G^0$, for some scalar $\lambda_G > 0$ and constant base period prices p_G^0. Following the steps in the text, define a new cost function on u and $\lambda_1, \ldots, \lambda_G, \ldots, \lambda_N$ and show that it is homogeneous and concave and that the derivative property holds for suitably defined quantity indices.

5.3. Show that for the linear expenditure system *any* group of commodities is separable from any other nonoverlapping group. What does this imply about substitutability between goods in the model? (See also §5.3 below.)

5.4. (Samuelson, 1956.) Discuss the plausibility of a model of intrafamily decision making in which each member possesses a utility function the value of which, in some suitable normalization, is an argument in an overall family utility function. Show that if no goods are consumed jointly (e.g., heating for the whole family), decentralized budgeting is possible. How might joint consumption be dealt with?

5.5. An intertemporal version of the linear expenditure has the utility function

$$u = \sum_{t=1}^{T} \sum_{k=1}^{n} \beta_{tk} \log (q_{tk} - \gamma_{tk})$$

where the t-suffix indicates the time period and the k-suffix the commodity. This is maximized (with exogenous labor supply) subject to the usual intertemporal budget constraint, see equation (2.7) of Chapter 4

$$\sum_k \sum_t p_{tk}^* q_{tk} = W_1$$

By analogy with the conventional linear expenditure system, write down the demand functions for q_{tk} in terms of the $(Tn \times 1)$ vector p^*, W_1 and the parameters. Solve this for $x_t^* = \Sigma_k p_{tk}^* q_{tk}$ in terms of the same variables, and show, by substitution, that q_{tk} can be expressed in terms of x_t^* and p_t^* alone.

5.6. The single period, linear expenditure system with utility functions $\Sigma\beta_k \log(q_k - \gamma_k)$ is separable. Good 1 is rationed to an amount z. Show that maximizing utility subject to $x = \Sigma p_k q_k$ *and* $q_1 = z$ is equivalent to maximizing $\Sigma_2^n \beta_k \log(q_k - \gamma_k)$ subject to $\Sigma_2^n p_k q_k = x - p_1 z$. Comment on the result.

5.7.* (On the sources of relative price changes.) Suppose that production possibilities in an economy are represented by the cost function $\Sigma c_i(r)e^{-\beta_i t}q_i$, where r is the vector of factor input prices, $c_i(r)e^{-\beta_i t}$ is the unit cost in the ith sector at time t, and β_i is the rate of technical progress in that sector. Show that efficient allocation requires that relative prices of outputs equal relative unit costs. Hence, discuss the role of factor proportions and technical progress in the determination of the relative prices of goods. How might we use an input-output table to measure the consequences for relative prices of changes in prices of primary inputs such as energy?

5.8.* (Exercise 5.7 continued.) If every consumer good price can be written as a linear combination with fixed weights of a few primary input prices, discuss how the Hicks composite commodity theorem as represented in (1.2) can be generalized. Discuss the problems of empirically implementing this approach.

5.2* Separability and two-stage budgeting: further results

The definition of separability

We begin by characterizing separability in terms of the preference ordering. If we write the vector of commodities q in the form $(q_G, q_{\bar{G}})$, where q_G is the vector in the group and $q_{\bar{G}}$ is the vector of excluded commodities, then for any arbitrary fixed vector $\bar{q}_{\bar{G}}$, say, the consumer's preferences over q will define an ordering on q_G. This is a "conditional" ordering on the goods in the group, and, in general, the position of different bundles within the group in the ordering will depend on the preselected values $\bar{q}_{\bar{G}}$. When this is not so, when the conditional ordering on goods in the group is independent of consumption levels outside the group, the group is said to be *separable.* In this case and this case only, the conditional ordering can be represented by a subutility function for the group, $v_G(q_G)$, say. If the whole commodity vector q can be partitioned into N such groups, we say that preferences are (weakly) separable. Separable preferences are represented by a utility function of the form

$$u = f[v_1(q_1), v_2(q_2), \ldots, v_G(q_G), \ldots, v_N(q_N)] \tag{2.1}$$

for subvectors $q_1, \ldots, q_G, \ldots, q_N$ and some function f which is increasing in all its arguments.

Given the argument in §5.1 above, a utility function of the form (2.1) implies subgroup demands or conditional demands of the form, for all i belonging to G

$$q_i = g_{Gi}(x_G, p_G) \tag{2.2}$$

where $x_G = \Sigma_k p_{Gk} q_{Gk}$ is total expenditure on group G. To see that (2.2) implies (2.1), consider the following heuristic argument based on Gorman (1971). Assume that the utility function $v(q)$ is *not* separable but that we partition its arguments to give $v(q_G, q_{\bar{G}})$ and that we maximize this subject to $\Sigma p_k q_k = x$ holding $q_{\bar{G}}$ constant at $\bar{q}_{\bar{G}}$, say. This is equivalent to maximizing the conditional utility function $v_G(q_G; \bar{q}_{\bar{G}}) = v(q_G, \bar{q}_{\bar{G}})$ subject to $\Sigma_{k\in G} p_k q_k = x_G = x - \Sigma_{k\in \bar{G}} p_k q_k$ and will result, for all x_G and p_G in conditional demand functions

$$q_i = f_{Gi}(x_G, p_G; \bar{q}_{\bar{G}}) \tag{2.3}$$

Since (2.2) and (2.3) hold simultaneously everywhere, the conditional demands must be identical to (2.2) and hence independent of the conditioning variables $\bar{q}_{\bar{G}}$. Thus v_G too must be independent of $\bar{q}_{\bar{G}}$ and the utility function is separable, contrary to our original assumption.

Separability and intergroup substitution

Weak separability places severe restrictions on the degree of substitutability between goods in different groups. Again, following Gorman (1971), let $i \in G$ and $j \in H$, where $G \neq H$. If we differentiate (2.2) with respect to p_j holding u constant, the only effect must be through x_G. Hence,

$$s_{ij} = \frac{\partial q_i}{\partial x_G} \cdot \frac{\partial x_G}{\partial p_j}\bigg|_{u=\text{const}} \tag{2.4}$$

But

$$s_{ji} = \frac{\partial q_j}{\partial x_H} \cdot \frac{\partial x_H}{\partial p_i}\bigg|_{u=\text{const}} = s_{ij}, \quad \text{by symmetry} \tag{2.5}$$

Thus, by division,

$$\frac{\partial x_G/\partial p_j}{\partial q_j/\partial x_H}\bigg|_{u=\text{const}} = \frac{\partial x_H/\partial p_i}{\partial q_i/\partial x_G}\bigg|_{u=\text{const}} \tag{2.6}$$

The left-hand side of (2.6) does not involve i, nor does the right-hand side involve j; hence the whole expression is independent of both and may be represented, say, by λ_{GH}. Thus

$$\frac{\partial x_G}{\partial p_j}\bigg|_{u=\text{const}} = \lambda_{GH} \frac{\partial q_j}{\partial x_H} \tag{2.7}$$

and, from (2.4), for $i \in G, j \in H$, and $G \neq H$,

$$s_{ij} = \mu_{GH} \frac{\partial q_i}{\partial x} \cdot \frac{\partial q_j}{\partial x} \tag{2.8}$$

where

$$\mu_{GH} \frac{\partial x_G}{\partial x} \frac{\partial x_H}{\partial x} = \lambda_{GH} \tag{2.9}$$

Although we shall not attempt a proof here, see Gorman (1971), the relationship (2.8) is both necessary and sufficient for weak separability so that this equation summarizes the empirical implications of separability. Note that substitutability between goods in different groups is limited in a very natural way. The quantity μ_{GH} summarizes the interrelation between the groups; individual commodities must conform to this except as modified by the total expenditure responses. Since the branches of the tree are the only means of contact between goods in different groups, responses to price changes and total expenditure changes must use the same channels of communication and hence bear a close relation to one another. Thus, for example, whole groups will be substitutes or complements for one another, and barring inferiority, all pairs of goods, one from each group, must also be substitutes or complements as the case may be.

We can now interpret the terms λ_{GH} in terms of the allocation problem between the broad groups of commodities. If we multiply (2.7) by p_j and sum over all j belonging to H, we get

$$\sum_{j\in H} p_j \frac{\partial x_G}{\partial p_j}\bigg|_{u=\text{const}} = \lambda_{GH} \sum_{j\in H} p_j \frac{\partial q_j}{\partial x_H} = \lambda_{GH} \tag{2.10}$$

Hence, λ_{GH} is the compensated derivative of expenditure on group G with respect to a proportional change of all prices in group H. The λ_{GH}'s are thus the intergroup substitution terms when each group is defined as a Hicks aggregate with fixed relative prices within the groups. If we define the elements λ_{GG} as the compensated own-price responses of each Hicks aggregate G to its own (aggregate) price, we can complete an $N \times N$ matrix out of the λ's that is interpretable as the Slutsky substitution matrix of the group aggregates under the Hicks' condition. Hence weak separability gives us a two-tier structure of substitution matrices. Within each group, no restrictions are placed on substitution and we have N completely general intragroup Slutsky matrices. Between groups, substitution is limited by (2.8), which in turn reflects a completely general pattern of substitution between groups although not between commodities. This "high-level" substitution matrix is represented by the quantities λ_{GH}. Note, however, that although the λ-matrix reflects substitution between the broad groups, we still have no allocation rule for these groups unless the conditions of the composite commodity theorem are met.

The allocation to broad groups

The correspondence between weak separability and two-stage budgeting and the existence of the subutility functions means that we can define group indirect utility functions and cost functions in the usual way. For example, we can define

$$c_G(u_G, p_G) = \min_{q_G} \left[\sum_G p_k q_k ; v_G(q_G) = u_G \right] \tag{2.11}$$

which, in turn yield group indirect utility functions,

$$u_G = \psi_G(x_G, p_G) \tag{2.12}$$

All these functions have properties of homogeneity, concavity, and so on that are the exact counterparts of the properties discussed in Chapter 2. Equations (2.12) and (2.1) allow us to write total utility u as

$$u = f(u_1, u_2, \ldots, u_G, \ldots, u_N) \tag{2.13}$$

The broad group allocation problem is then solved by maximizing (2.13) subject to the budget constraint

$$\sum_1^N c_G(u_G, p_G) = x \tag{2.14}$$

In general, if no restrictions are placed on the functions c_G or ψ_G, solution of this maximization problem requires knowledge of all of the individual prices. It is thus not generally possible to replace (2.13) and (2.14) by a maximization problem involving a single price index and a single quantity index for each of the broad groups. The conditions which do permit this have been derived by Gorman (1959). One possibility is that each group cost function $c_G(u_G, p_G)$ can be written as $\theta_G(u_G)b_G(p_G)$ for some first degree homogeneous function $b_G(p_G)$ and some monotone increasing function $\theta_G(u_G)$. Making the substitution $v_G = \theta_G(u_G)$, the maximization problem becomes

$$\max u = f[\theta_1^{-1}(v_1), \ldots, \theta_G^{-1}(v_G), \ldots, \theta_N^{-1}(v_N)] \tag{2.15}$$

such that

$$\Sigma b_G(p_G)v_G = x$$

Clearly, the v_G's can be thought of as group quantity indices with corresponding price indices $b_G(p_G)$, so that (2.15) gives a perfect utility maximizing solution to the broad group allocation problem. Unfortunately, as we shall see in §5.4, this specific form for c_G implies that the budget shares within each group are independent of total group expenditure. Hence, we may never group together necessities and luxuries; everything in each group must have the same total expenditure elasticity. This condition and the composite commodity theorem are each in their own way rather restrictive and neither provides a satisfactory general basis for the allocation to broad groups.

The second solution proposed by Gorman is that the indirect utility functions ψ_G take the *Gorman generalized polar form*

$$\psi_G(x_G, p_G) = F_G[x_G/b_G(p_G)] + a_G(p_G) \tag{2.16}$$

for some monotone increasing function $F_G(\ )$, while the utility function should take an explicitly additive form between groups, that is,

$$u = v_1(q_1) + v_2(q_2) + \cdots + v_G(q_G) + \cdots + v_N(q_N) \tag{2.17}$$

In this case, (2.13) is also additive, and using essentially the same substitutions as before, that is, $v_G = x_G/b_G(p_G)$, (2.15) becomes

$$\max u = \Sigma F_G(v_G) + \Sigma a_G(p_G)$$

such that

$$\Sigma b_G(p_G)v_G = x$$

Once again, the v_G's and b_G's are interpretable as quantity and price indices while the additive structure of utility means that the functions a_G are irrelevant for the maximization and thus do not need to be known by the consumer. The Gorman generalized polar form is a good deal less restrictive with nonzero a_G's than without. Although it places restrictions on the allowable forms, see Exercise 5.13, nonlinear relationships between within-group expenditures and group outlay are permitted. However, the additivity restriction (2.17) is both severe and, in our view, unrealistic; it will be discussed more fully in §5.3. Gorman shows that, apart from the case where there are only two groups, the conditions of this or the previous paragraph are necessary and sufficient for broad group allocation within the framework of weak separability. A reasonably general solution must thus be either an approximate one or one that abandons weak separability.

An approximate solution can be derived by returning to the original problem, the maximization of (2.13) subject to (2.14). In Chapter 7 we shall discuss how the cost function can be used to define price and quantity indices. Anticipating this, choose some base price vector p^0 with associated subvectors p_G^0 and write

$$c_G(u_G, p_G) = c(u_G, p_G^0) \cdot \frac{c(u_G, p_G)}{c(u_G, p_G^0)} \tag{2.18}$$

The first term on the right-hand side of (2.18) is the money cost of reaching the utility level u_G at base period group prices p_G^0. This is itself a satisfactory measure of utility and we write it v_G so that u_G is given by $\psi(v_G, p_G^0)$. The second term, as we shall see in Chapter 7 is a true cost-of-living price index for group G; we write it as $P_G(p_G, p_G^0; u_G)$ to emphasize its dependence on u_G. The problem, (2.13) to (2.14), is then to maximize

$$u = f\{\psi_1(v_1, p_1^0), \ldots, \psi_G(v_G, p_G^0), \ldots, \psi_N(v_N, p_N^0)\}$$

such that

$$\Sigma v_G P_G(p_G, p_G^0; u_G) = x \tag{2.19}$$

This is now very much in standard form with matching price and quantity indices v_G and P_G. In fact, in empirical work, we rarely have available physical measures of quantity and instead work with quantity and price indices. Hence, it is quite usual to think of the budget constraint in the form of (2.19). The difficulty lies in the fact that P_G is a function of u_G and if we allow for this dependency explicitly we are, of course, back to (2.13) and (2.14). Nevertheless, as we shall see in Chapter 7, it may be that the empirical variation of P_G with u_G is not very great so that indices involved in (2.19) can be approximated by Paasche or Laspeyres indices, which are exactly those indices that are involved when demand systems are estimated on national accounts data for aggregated groups. So the construct (2.19) may provide some justification for standard practice, albeit in an approximate way.

The variation of the group price indices with group utility reflects the fact that within-group Engel curves generally are not straight lines through the origin. If, however, the group cost functions take the form $x_G = a_G(p_G) + \theta_G(u_G)b_G(p_G)$ for functions a_G, θ_G, and b_G, then it will be shown in §5.4 that this gives linear, within-group Engel curves (not necessarily through the origin). With this form and only weak separability, it is easily seen (Exercise 5.14) that total group expenditure can be written as a function of total expenditure and two sets of price indices constructed from the a's and b's. Exactly this idea, assuming weak separability but working with local approximations in terms of derivatives, has been advocated by those who work with differential demand systems, see particularly Barten and Turnovsky (1966), Barten (1970) and Theil (1975). The algebra is simplest using the Rotterdam equations [Chapter 3, (3.3)] although the derivation is quite general since no parameters are being assumed constant. Equations (2.8) and (2.9) can be rewritten, for all $i \in G$ and $j \in H$, and Kronecker delta δ_{GH} (=0 for $G \neq H$ and 1 if $G = H$),

$$p_i p_j s_{ij} = \delta_{GH} s_{ij}^G p_i p_j + \lambda_{GH}\, p_i \frac{\partial q_i}{\partial x_G}\, p_j \frac{\partial q_j}{\partial x_H} \tag{2.20}$$

to give the Rotterdam c_{ij} parameters when divided by x. Application of homogeneity to (2.20) implies that $\Sigma_{i\in G} p_i s_{ij}^G = 0$ for all j and G, see Exercise 5.10, so that substitution of (2.20) into (3.3) from Chapter 3 and summing over $i \in G$ gives

$$\sum_{i\in G} w_i\, d \log q_i = x^{-1} \sum_{i\in G} p_i\, dq_i = b_G(d \log x - \Sigma w_k\, d \log p_k) + \sum_{j\in H}\sum_{H} x^{-1}\lambda_{GH} p_j \frac{\partial q_j}{\partial x_H}\, d \log p_j \tag{2.21}$$

where $b_G = \Sigma_{i\in G} b_i$ is the marginal propensity to spend on the group. We can then define *two* differential group price indices by

$$d \log P_G^1 = \sum_{k\in G} \left(\frac{p_k q_k}{x_G}\right) d \log p_k$$
$$d \log P_G^2 = \sum_{k\in G} \left(p_k \frac{\partial q_k}{\partial x_G}\right) d \log p_k \tag{2.22}$$

one with in-group budget shares as weights, the other with in-group marginal propensities. Equation (2.21) then becomes

$$\sum_{i\in G} w_i\, d \log q_i = x^{-1} \sum_{i\in G} p_i\, dq_i = b_G(d \log x - \Sigma w_G\, d \log P_G^1) + \Sigma \lambda_{GH}^*\, d \log P_H^2 \tag{2.23}$$

where $w_G = \Sigma_{k\in G} p_k q_k / x$ and $\lambda_{GH}^* = \lambda_{GH}/x$. Apart from the presence of the two price indices instead of one, (2.23) is another Rotterdam system for allocating the broad groups, moreover, b_G and λ_{GH}^* may easily be shown to possess all the correct properties of such parameters in the model. Hence, (2.23) offers a method of maintaining an optimal allocation, once it is reached, requiring knowledge of only two price indices for each group. In practice, if prices move roughly together, the two price indices may well be close to one another and might be adequately proxied by a single index. Note that this result requires no restrictions beyond weak separability.

Implicit separability and two-stage budgeting

Weak separability is by no means the only type of separability nor is two-stage budgeting as so far defined the only useful form of hierarchic decision making. Implicit or quasi separability, first described by Gorman (1970), provides an alternative and, in many respects, simpler structure. Preferences are said to be implicitly separable if and only if the cost function can be written in the form

$$c(u, p) = C[u, c_1(u, p_1), \ldots, c_G(u, p_G), \ldots, c_N(u, p_N)] \tag{2.24}$$

where, as before, the goods are partitioned into N groups with price subvectors $p_1, \ldots, p_N$. The functions $c_G(u, p_G)$ are increasing in u and p_G. Note carefully that it is *total* utility that appears in each of the functions $c_G(u, p_G)$; in sharp contrast to weak separability there are no group subutilities.

Differentiating (2.24) with respect to p_i, where good i belongs to group G,

$$q_i = h_i(u, p) = \frac{\partial c(u, p)}{\partial p_i} = \frac{\partial C}{\partial c_G} \cdot \frac{\partial c_G}{\partial p_i} \tag{2.25}$$

Hence group expenditure x_G is given by

$$x_G = \sum_{i\in G} q_i p_i = \frac{\partial C}{\partial c_G} \sum_{i\in G} p_i \frac{\partial c_G}{\partial p_i} = \frac{\partial C}{\partial c_G} \cdot c_G(u, p_G) \tag{2.26}$$

since each $c_G(u, p_G)$ is linear homogeneous. Thus, writing $w_G \equiv x_G/x$ for the group budget shares

$$w_G = \frac{\partial \log C}{\partial \log c_G} \tag{2.27}$$

while for the intragroup budget shares $w_{iG} \equiv p_i q_i / x_G$, from (2.25) and (2.26)

$$w_{iG} = \frac{\partial \log c_G(u, p_G)}{\partial \log p_i} \tag{2.28}$$

Provided we are prepared to revise our concepts somewhat, equations (2.27) and (2.28) show that implicit separability provides an extremely simple and perfectly consistent form of two-stage budgeting. First of all, it is clear from (2.26) that the functions $c_G(u, p_G)$ are *not* equal to group total expenditures x_G; rather they should be thought of as group price indices that depend on the level of utility u. Hence (2.24) tells us that, given implicit separability, we can define group price indices $c_G(u, p_G)$ – properly homogeneous in p_G – over which a macrocost function $C(u, \ )$ is defined. This macrocost function can be used to allocate expenditures to the broad groups in accordance with the derivative property in logarithmic form (2.27). In fact, we can go beyond this, and we shall show next that quantity indices can be defined for each group that are equal to the partial derivatives themselves. Hence, the top stage of two-stage budgeting can take place according to "macro" utility maximization and without error if preferences are such that (2.24) holds. The second, or bottom stage, is given by (2.28). The allocation within the groups can also be derived from appropriately defined cost functions and again the derivative rule works in logarithmic form. From (2.25), we see that q_i is not equal to $\partial c_G / \partial p_i$, so (2.28) is as far as we shall get. But this is really quite enough; the microallocations can be carried out using total utility and group prices alone.

*Implicit separability and the distance function***

The previous discussion can be extended if we examine the implications of implicit separability for the distance function discussed in §2.7. Define (direct) implicit separability as holding if and only if the distance function can be written in the form

$$d(u, q) = D[u, d_1(u, q_1), \ldots, d_G(u, q_G), \ldots, d_N(u, q_N)] \tag{2.29}$$

for a partition of q. This is clearly dual to the definition of implicit separability through the cost function (2.24). It turns out that the two definitions are not only analogous, they are identical. Following Gorman (1976), starting from (7.4) in Chapter 2 and substituting from (2.29)

$$\begin{aligned}
c(u, p) &= \min_q [\Sigma\, p_k q_k;\ d(u, q) = 1] \\
&= \min_q \Big\{\sum_G \sum_{k \in G} p_k q_k;\ D[u, d_1(u, q_1), \ldots, d_N(u, q_N)] = 1\Big\} \\
&= \min_v \Big\{\sum_G \min_{q_G} \Big[\sum_{k \in G} p_k q_k;\ d_G(u, q_G) = v_G\Big];\ D(u, v) = 1\Big\} \\
&= \min_v \Big\{\sum_G \min_{q_G^*} \Big[\sum_{k \in G} p_k q_k^* v_G;\ d_G(u, q_G^*) = 1\Big];\ D(u, v) = 1\Big\} \\
&= \min_v \Big[\sum_G v_G c_G(u, p_G);\ D(u, v) = 1\Big] \\
&= C[u, c_1(u, p_1), \ldots, c_G(u, p_G), \ldots, c_N(u, p_N)]
\end{aligned} \tag{2.30}$$

where $v_1, \ldots, v_N$ are a set of positive scalars, $q_G^* = q_G / v_G$, C is some function, and $c_G(u, p_G)$ is defined by

$$c_G(u, p_G) = \min_{q_G} \Big[\sum_{k \in G} q_k p_k;\ d_G(u, q_G) = 1\Big] \tag{2.31}$$

A precisely similar converse argument takes us from (2.30) to (2.29). Hence, under implicit separability, the structure of separability is identical in both primal and dual formulations. No other form of separability seems to possess this extremely elegant and simple property.

From the definition (2.31), q_G would be the cheapest way of reaching u at p_G if utility were implicitly defined by $d_G(u, q_G) = 1$. Hence, compare with equation (7.3) in Chapter 2,

$$x_G = \sum_{k \in G} p_k q_k = d_G(u, q_G) \cdot c_G(u, p_G) \tag{2.32}$$

Comparing with (2.26) we have at once

$$\frac{\partial C(u, c_1, \ldots, c_G, \ldots, c_N)}{\partial c_G} = d_G(u, q_G) \tag{2.33}$$

Equation (2.33) provides the justification for claims that implicit separability provides a perfect allocation mechanism for broad groups. The expression $d_G(u, q_G)$ is a quantity index for group G, dependent on utility, but linearly homogeneous in q_G. The indices $c_1 \cdots c_N$, are price indices, one for each group, and again dependent on the utility level u. The macrocost function $C(\ )$ is then defined over these N indices and utility u; it too is concave and linearly homogeneous in the indices. Finally, (2.33) shows that the quantity indices are given as the derivatives of the macrocost function with respect to the price indices. Thus, provided we can define the indices appropriately, implicit separability allows us to treat broad groups of goods as if they were single commodities.

Empirical comparisons of weak and implicit separability

We have seen how both weak and implicit separability, although different from one another, each imply schemes for two-stage budgeting. At the first level, the allocation to broad groups, there is perhaps little difference between the two structures. Implicit separability gives a theoretically "perfect" solution, but the price and quantity indices depend on the total utility level and, in empirical work, this is likely to require some approximation. Analogous approximations are required for the weakly separable scheme (2.24). Even so, detailed empirical comparisons may yet reveal unsuspected differences.

At the second stage of budgeting, the situation is quite different. In Hicksian terms, weak separability implies that *quantities* demanded are a function of *group* utility and within-group prices while implicit separability implies that *intragroup budget shares* are a function of *total* utility and within-group prices. This last implies, for example, that a compensated price increase outside the group – no matter what the good – will change all intragroup expenditures by the same proportion. This is matched, for weak separability, by the restrictions on substitution given by (2.8). These differences are important both for theory and for econometrics. In empirical work, the obvious application is to systems of demand equations. As we argued before, relating commodity demands to total expenditure x and current prices p assumes weak intertemporal separability of preferences. If instead, preferences are *implicitly* intertemporally separable, budget shares should be related not to x and p but to intertemporal utility and p. This would demand the introduction of lifetime expected wealth W_1 as well as future anticipated prices into demand systems and might have radical implications for the type of results usually obtained. This is work that remains to be done. The empirical task is complicated by the fact that we are comparing models with the same dependent variables but different explanatory variables. Conventional statistical tests designed for testing restrictions (such as homogeneity or symmetry) do not deal with this case, and instead we require tests for "nonnested" models. These tests have recently been developed in the econometric literature [for an elementary application to demand analysis see Deaton (1978)], but considerably more work needs to be undertaken before we can be certain of their properties in practice.

Exercises

5.9. The marginal rate of substitution between goods i and j is given by $MRS_{ij} = (\partial v/\partial q_i)/(\partial v/\partial q_j)$, which equals p_i/p_j at an internal equilibrium, see §2.2. Show that under weak separability, if i and $j \in G$ and $k \in H \neq G$, then MRS_{ij} is independent of k. The converse of this result is also true and is known as the Leontief separation theorem, see Leontief (1947).

5.10. Using (2.8) and the definition of λ_{GH} given by (2.10), prove that (a) $\Sigma_G \lambda_{GH} = \Sigma_H \lambda_{GH} = 0$, (b) $\lambda_{GH} = \lambda_{HG}$, (c) $\Sigma_{i \in G} p_i s_{ij}^G = 0$.

5.11. Prove that the compensated cross-price elasticity $e_{i1}^* = \partial \log q_i / \partial \log p_1 |u =$ constant, is independent of i if and only if all other goods are implicitly separable from good 1.

5.12. *Indirect* weak separability is said to hold when the indirect utility function $\psi(x, p)$ written in terms of the price-outlay ratios $r = p/x$ takes the form $\psi(x, p) = \psi^*(r) = f[\psi_1(r_1), \ldots, \psi_G(r_G), \ldots, \psi_N(r_N)]$ for some partition of prices. Show that, under indirect separability, the intragroup value shares $w_{iG} = w_i/w_G$ depend only on total expenditure and group prices. What is the difference between this result and the corresponding formulation under implicit separability?

5.13. The Gorman "generalized polar form" of the indirect utility function (at the group level) is $u = \theta[x/b(p)] + a(p)$. Show, by working with the corresponding cost function, that this implies a relationship between the budget shares of the form

$$w_i = A_{ij}(p) w_j + B_{ij}(p)$$

5.14. Suppose the utility function is weakly separable and that, at the group level, preferences are given by $c_G(u_G, p_G) = a_G(p_G) + u_G b_G(p_G)$. Show that group expenditure can be written as a function of total expenditure and the price indices a_K, b_K, $K = 1, \ldots, N$. Is this result more or less general than the Barten and Turnovsky result discussed in the text?

5.15.* Using the distance function $d(u, q) = 1$ to characterize preferences, show that $MRS_{ij} = (\partial d/\partial q_i)/(\partial d/\partial q_j)$. Hence show that for $i, j \in G$ and $k \notin G$, weak separability implies that MRS_{ij} is independent of k while implicit separability implies that MRS_{ij} is independent of k *along an indifference curve*. How are we to interpret the phrase "along an indifference curve" when a single quantity is being varied? (Continued as Exercise 5.28.)

5.3 Strong separability and additive preferences

Definition of additivity

One of the most popular assumptions about preferences, and one of the most restrictive, is strong or additive separability. In this case, the direct utility function is again made up of subutility functions for each group, but unlike weak separability, they are combined additively. Hence, utility is written as

$$u = F[v_1(q_1) + v_2(q_2) + \cdots + v_N(q_N)] \tag{3.1}$$

so that under some monotone transformation, the utility function takes the explicitly additive form. Note carefully that it is preferences that are strongly or additively separable, not the utility function, hence for example

$$u = \Pi \exp[v_k(q_k)] \quad \text{and} \quad u = \Sigma v_k(q_k) \tag{3.2}$$

are both representations of the same additively separable preferences. In the case where there is only one good in each group, preferences are said to be *additive* or occasionally, that "wants are independent." The terms *strong separability* or *block additivity* are usually reserved for the case of multigood groups. In the history of consumer theory, additive preferences have played an important role. The founders of consumer theory, Gossen and Jevons, clearly thought of utility as being generated in this way, and later writers such as Marshall and Pigou made heavy use of the assumption of independent wants. In more recent writing, theorists have continued to find it a very useful assumption, while econometricians have found its restrictiveness an important aid in generating degrees of freedom for estimation. For example, by far and away the most popular demand model, the linear expenditure system, is derived from the utility function (2.5) in Chapter 3, which can easily be seen to represent additive preferences.

Empirical implications of additivity

The restrictiveness of (3.1) lies in the fact that no group occupies any special position. Since the function is additive, we can arbitrarily create new groups by combining any others, and this effectively prevents the existence of any particular relationships between pairs of groups. We shall take up this point formally at the end of this section, where we show that if i and j are two goods in different groups, the substitution effect between them can only take the form

$$s_{ij} = \mu \frac{\partial q_i}{\partial x}\frac{\partial q_j}{\partial x} \qquad (3.3)$$

Note that μ is independent of the groups to which i and j belong. If preferences are additive, then (3.3) holds for all pairs of goods. It is this case that we pursue further; if preferences are block additive, our conclusions will apply to the broad groups regarded as Hicks aggregates.

Equation (3.3) defines the off-diagonal terms of the Slutsky matrix. The diagonal terms can be filled in using the relationship $\Sigma_k s_{ik} p_k = 0$ to give

$$s_{ii} = -\frac{\mu}{p_i}\frac{\partial q_i}{\partial x}\left(1 - p_i \frac{\partial q_i}{\partial x}\right) \qquad (3.4)$$

In elasticity terms, using $e^*_{ij} = s_{ij}p_j/q_i$ and $e_{ij} = e^*_{ij} - e_i w_j$,

$$\begin{aligned} e_{ii} &= \phi e_i - e_i w_i(1 + \phi e_i) \qquad i = 1, \ldots, n \\ e_{ij} &= -\, e_i w_j (1 + \phi e_j) \qquad i \neq j \end{aligned} \qquad (3.5)$$

where $\phi = -\mu/x$ is a scalar not indexed on i. This last relation (3.5), first observed by Frisch (1959), shows why additivity is both so useful and so restrictive. Apart from the parameter ϕ, knowledge of expenditure elasticities alone is sufficient to determine all the own and cross-price elasticities. Consequently, the econometrician needs almost no relative price variation in the data in order to estimate price elasticities so that, for example, knowledge of one price elasticity is sufficient to allow cross-section household budget data to be used to "measure" price elasticities. The measurement is, of course, largely by assumption.

All this information comes at a price. Equations (3.3) and (3.4) give a particularly simple structure to the substitution matrix, and apart from very peculiar cases, the matrix will only be negative semidefinite if μ is positive (so that ϕ is negative) and if each of the expenditure elasticities is positive. Thus, inferior goods are ruled out at once, while (3.3) rules out complements permitting goods to be substitutes only. Even more severe are the consequences of (3.5). Since all the value shares are positive and add to unity, each is of the same order of magnitude as the reciprocal of the number of goods. Hence, if the number of goods – or commodity groups in the case of block additivity – is at all large, we have, to a reasonable degree of approximation, from (3.5)

$$e_{ii} \simeq \phi e_i \qquad (3.6)$$

This approximate proportionality of expenditure and price elasticities, which in practice tends to be quite accurate even for eight or ten commodities, has been called Pigou's Law by Deaton (1974b) after a suggestion on the same lines by Pigou (1910). Even for broad aggregates of goods, we have no reason to suppose a priori that (3.6) is true so that, if we assume additivity, we risk severe distortion of measurement. For a fairly detailed disaggregation of commodities, the distortion is clear and is documented commodity by commodity for the case of the linear expenditure system in Deaton (1975a). In practice, the data is usually such that the variation in total outlay is much greater than that in relative prices, so the price elasticities tend to be distorted most heavily. Figure 5.2, reproduced from Deaton (1975b), illustrates this feature for a 37-commodity disaggregation of British consumers' expenditure. It shows a plot of price elasticities calculated from estimates of the linear expenditure system against those from another model that is not derived from additive preferences. These latter were cross-checked against a second nonadditive model and shown to correspond reasonably closely. Between the linear expenditure system and the nonadditive models, however, the correspondence is virtually nonexistent, as Figure 5.2 shows. This strongly suggests that additivity and its implications, (3.5) and (3.6), far from being an aid to measurement, positively hinder it.

Even so, it can be argued that for sufficiently broad groups of goods, the proportionality restriction and the absence of inferiority and complementarity are not obviously inappropriate. The evidence would seem to be against this. For example, it is possible to test (3.3) and (3.4) within the Rotterdam model. In this case, the C matrix of the Rotterdam model, see (3.6) in Chapter 3, takes the form, see Exercise 5.20,

$$c_{ij} = \phi(\delta_{ij} b_i - b_i b_j) \qquad (3.7)$$

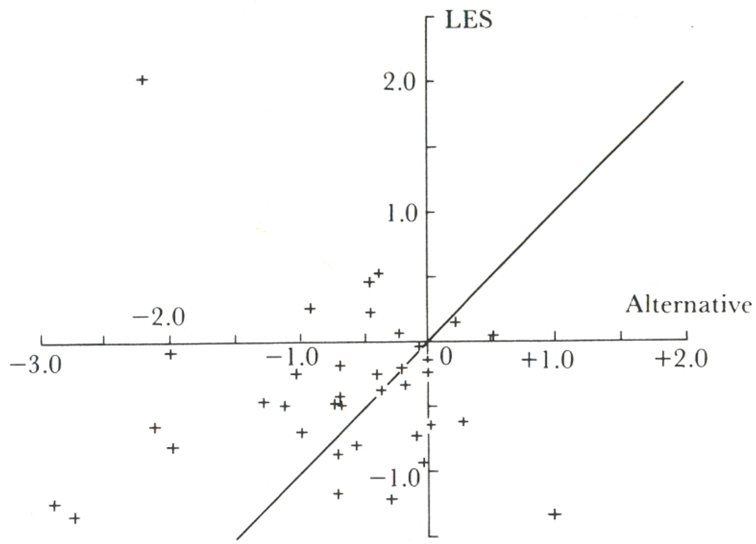

Figure 5.2. Price elasticities estimated from the linear expenditure system and an alternative model. (From A. S. Deaton, *European Economic Review*, Vol. 6, 1975.)

where ϕ is as before and b_i is the marginal propensity to spend on good i. This restriction can be tested in the same way as symmetry and homogeneity was tested in Chapter 3. Barten (1969) has done this for a 16-commodity disaggregation of Dutch data and encountered a very large drop in likelihood, indicating a sharp rejection of the additivity hypothesis. Deaton (1974a), using 9 commodities and British data, came to a similar conclusion and also found that the linear expenditure system, which is also additive, gave similar results to the additive version of the Rotterdam model. Even with highly aggregated commodities and only 4 groups, Theil (1976) does not find additivity acceptable. Further evidence using nonnested testing procedures with the linear expenditure system in Deaton (1978) further supports the conclusion. Clearly then, strong separability and additivity are too strong to be used in empirical work, despite their undoubted econometric advantages.

Additivity and the "measurement" of marginal utility

In the case where the first-order conditions $\partial v/\partial q_i = \lambda p_i$ solve for the demand functions, an alternative derivation of (3.5) is possible and this is the route followed by most previous authors. In particular, the function F in (3.1) is taken as the unit function $F(x) \equiv x$, so that normalization of utility is chosen that gives the explicitly additive form. In this normalization, marginal utility $\partial v/\partial q_i$ is a function of q_i alone, and this can be used to derive (3.5), see Exercise 5.16. Following this derivation, the quantity ϕ turns out to be the reciprocal of $\partial \log \lambda/\partial \log x$, the elasticity of the marginal utility of outlay with respect to outlay. Since the estimation of any demand system based on additive preferences will yield an estimate of ϕ, it would appear that demand analysis can provide a measure, if not of marginal utility, at least of its derivative with respect to total expenditure. As we shall see in Chapter 9, such a measure could be useful in applied welfare economics, and many authors, starting with Frisch (1932), have attempted to use additivity to make such calculations. Such applications have produced a fair uniformity of results; ϕ is usually estimated as close to -0.5, giving an estimate of -2 for the elasticity of marginal utility.

In assessing these claims, note first that our original derivation of (3.5) is quite independent of the choice for F. The equation would be the same and ϕ would be the same if, for example, utility were either $\Sigma v_k(q_k)$ or $\Pi \exp[v_k(q_k)]$. In other words, ϕ is a parameter of the *preference ordering* and is independent of the specific utility function chosen to represent it. It cannot then have anything to do with marginal utility, which varies from one normalization to another. It is certainly true that for $u = \Sigma v_k(q_k)$, $\phi = (\partial \log \lambda/\partial \log x)^{-1}$, but as soon as the utility function is nonlinearly transformed, for example to $\exp[\Sigma v_k(q_k)]$, ϕ no longer has this interpretation, although preferences, observable behavior, and the estimate of ϕ remain unchanged. Hence, we can only accept ϕ as a measure of the reciprocal of the elasticity of marginal utility if (a) we accept that consumers do actually have a utility level (i.e., the utility function has a real significance beyond the convenient mathematical representation of preferences) and (b) that consumers with additive preferences always choose to measure utility using the explicitly additive form. We see no conceivable basis for testing either (a) or (b) so that there seems to be no grounds other than assertion for assigning the marginal utility interpretation to estimates of ϕ. Nor is the uniformity in the estimates of ϕ itself surprising. According to (3.6), ϕ will be estimated at close to the average ratio of price to total expenditure elasticity. Most studies have worked with between 6 and 16 commodities and, at this level of disaggregation, an average of -0.5 is perfectly plausible.

Formal derivation of equation (3.3) *

Start from equation (3.1) and take three goods i, j, and k each belonging to a different group I, J, and K. Since a strongly separable utility function is certainly weakly separable, we can use the results of §5.2, particularly (2.8). We do this in two ways, first, recognizing the three groups and, second, combining J and K into a new group L, which because of strong separability is also separable from I. From (2.8), there exist μ_{IJ}, μ^*_{IL}, μ_{IK}, such that

$$s_{ij} = \mu_{IJ} \frac{\partial q_i}{\partial x} \cdot \frac{\partial q_j}{\partial x} = \mu^*_{IL} \frac{\partial q_i}{\partial x} \cdot \frac{\partial q_j}{\partial x} \quad (3.7)$$

$$s_{ik} = \mu_{IK} \frac{\partial q_i}{\partial x} \cdot \frac{\partial q_k}{\partial x} = \mu^*_{IL} \frac{\partial q_i}{\partial x} \cdot \frac{\partial q_k}{\partial x} \quad (3.8)$$

Hence, dividing s_{ij} by s_{ik}, $\mu_{IJ} = \mu_{IK}$ and, since J and K are arbitrary, μ_{IJ} is dependent on I only. However, by symmetry, $s_{ij} = s_{ji}$, $\mu_{IJ} = \mu_{J.}$ and so μ_{IJ} is independent of both I and J. Hence, writing $\mu_{IJ} = \mu$, (3.7) reduces to (3.3).

Exercises

5.16. Prove (3.5) from the first-order conditions (2.4) in Chapter 2 following the outline given below. Taking logarithms of (2.4) in Chapter 2 gives $\log(\partial v/\partial q_i) = \log \lambda + \log p_i$. Differentiate this with respect to $\log x$ and $\log p_j$ in turn, remembering that, under additivity, $\partial v/\partial q_i$ is a function of q_i alone. Hence, writing $\omega = \partial \log \lambda/\partial \log x$ and $\theta_i = \partial \log \lambda/\partial \log p_i$, show that $\omega e_{ij} = e_i\theta_j + \delta_{ij}e_i$. Finally, use the adding-up restrictions $\Sigma w_i e_{ij} + w_j = 0$ and $\Sigma w_i e_i = 1$ to solve for and eliminate θ_j. Check that $\phi = \omega^{-1}$.
5.17. In his original article, Pigou (1910) asserts that under additivity the proportional effect on the marginal utility of money of a one percentage increase in a single price is negligibly small. Assess this claim in the light of your solution for θ_i in the previous question.
5.18. Show that, under additivity, cross-price elasticities are always small relative to own-price and to total expenditure elasticities.
5.19. Lluch, Powell, and Williams (1978) estimate the linear expenditure system for four commodity groups (one of which is usually all food) for a large number of countries and state that they find little evidence for Deaton's (1974b) "assertion" that $e_{ii} \simeq \phi e_i$. Why might the approximation break down in this case? Does this mean that the linear expenditure system is less restrictive for very broad groups of goods?
5.20. Using (3.5), show that the Rotterdam c_{ij} parameters satisfy $c_{ij} = \phi(\delta_{ij} b_i - b_i b_j)$ under additivity. Check this for the case of the linear expenditure system, and derive an expression for ϕ in this case. How would you expect ϕ to vary over time if calculated on the basis of estimates from the linear expenditure system?
5.21. *Indirect* additivity is said to hold when the indirect utility function in terms of $p/x \equiv r$ can be written $\psi(x, p) \equiv \psi^*(r) = f[\Sigma \psi_k(r_k)]$. Show that the condition e_{ij} independent of j for all $i, j, i \neq j$ is both necessary and sufficient for indirect additivity. (Necessity is straightforward, sufficiency is harder.)

5.4 Homotheticity and quasi homotheticity

Homothetic preferences

We have already had occasion to mention homotheticity several times and the time has come to describe it more fully. Preferences are said to be homothetic, if, for some normalization of the utility function, doubling quantities doubles utility. Drawing the analogy with production theory, preferences are said to

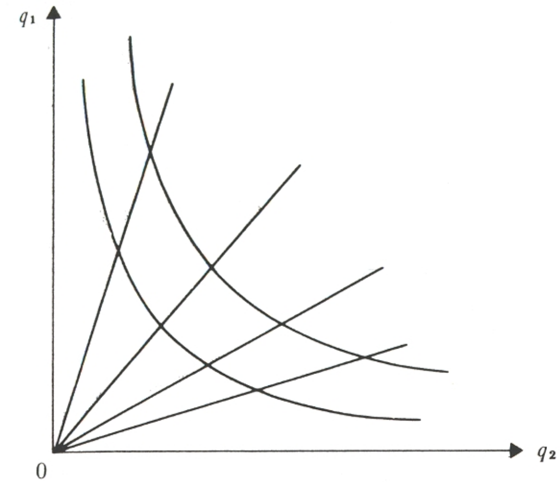

Figure 5.3. Homothetic preferences.

be homothetic if utility can be produced under constant returns to scale. In this case, each indifference curve is simply a magnified or reduced version of every other one. As illustrated in Figure 5.3, any ray through the origin will cut all the indifference curves at points where the slope is the same, which has several immediate consequences. First, because of the constant slope property, the expansion paths given by increasing x with p constant will be straight lines through the origin. Clearly, the implication is that the composition of the budget is independent of total expenditure or of utility. Hence, all expenditure elasticities are unity. The second consequence is for the structure of the cost function. If we label indifference curves along any ray through the origin such that double utility is generated by being twice as far from the origin, then the cost of reaching utility u must be proportional to u.

These two properties, obvious from Figure 5.3, can easily be demonstrated formally. First, preferences are homothetic if and only if we can write, for an arbitrary scalar $\theta > 0$

$$u = F[v(q)]; \quad v(\theta q) = \theta v(q) \quad (4.1)$$

That is, utility is a monotone increasing function of a function that is homogeneous of degree one. By ordinality, we can choose $v(q)$ itself as our representation of preferences. In this normalization $c(u, p) = \min[\Sigma p_k q_k;\ v(q) = u]$. Since $v(q)$ is homogeneous, doubling u will be accomplished by doubling q and, hence, by doubling costs. Then

$$c(u, p) = ub(p) \quad (4.2)$$

for some function $b(p)$ that is linearly homogeneous and concave in p. Again by analogy with production theory, the marginal and average costs of utility are constant and equal to one another. If we take logs of (4.2) and differentiate with respect to $\log p_i$, we have

$$w_i = \frac{\partial \log c(u, p)}{\partial \log p_i} = \frac{\partial \log b(p)}{\partial \log p_i} \tag{4.3}$$

which, since $b(p)$ is independent of u, establishes that the budget shares are independent of u and hence of x. That all expenditure elasticities should be unity contradicts all known household budget studies, not to mention most of the time-series evidence of systematic change in expenditure patterns as total outlays increase. In consequence, the assumption of constant returns to scale is much less useful in consumption than in production theory.

Quasi homotheticity and linear Engel curves

Much of the mathematical convenience of homotheticity may be retained within the economically much more fruitful assumption of "quasi" homotheticity, exploited extensively by Gorman (1961, 1976). In this formulation, a fixed cost element is added to the cost function (4.2), so that we have the *Gorman polar form* [c.f., (2.16) with $F_G(x) \equiv x$]

$$c(u, p) = a(p) + ub(p) \tag{4.4}$$

The quantity $a(p)$ represents the cost of living when u is zero, and may thus be interpreted as subsistence expenditure. For example, in the linear expenditure system, which has a cost function of the form (4.4), $a(p)$ takes the form $\Sigma\gamma_k p_k$, so that subsistence expenditure is the cost of the subsistence quantities γ. From (4.4), the indirect utility function is given by

$$u = \psi(x, p) = \frac{x - a(p)}{b(p)} \tag{4.5}$$

which has an obvious interpretation as the real value of expenditure in excess of that required for subsistence. Differentiating (4.4) and substituting from (4.5) gives

$$q_i = a_i(p) + \frac{b_i(p)}{b(p)}[x - a(p)] \tag{4.6}$$

where $a_i(p)$ and $b_i(p)$ are the ith partial derivatives of $a(p)$ and $b(p)$, respectively. This equation shows how quasi homotheticity restricts behavior; all Engel curves are straight lines. Under homotheticity, Engel curves are straight lines through the origin, so all expenditure elasticities are unity; under "quasi" homotheticity, the straight lines need not go through the origin, and elasticities only tend to unity as total expenditure increases. This is a significant generalization although it is still unlikely to be true for narrowly defined commodities. Even for broad groups such as food, household budget studies tend to give nonlinear Engel curves. However, such data do not, by definition, show the response of single individuals or households as outlays increase. On time-series data, albeit at an aggregate level, it is much less obvious that linear Engel curves are contradicted by the evidence for broad groups of goods.

If quasi homotheticity is valid, it offers a very convenient representation of how the pattern of demand alters with total outlay. In budget share, form (4.6) implies

$$w_i = \left(\frac{p_i a_i}{a}\right)\left(\frac{a}{x}\right) + \left(\frac{p_i b_i}{b}\right)\left(1 - \frac{a}{x}\right) \tag{4.7}$$

so that the ith value share is a weighted average of two expressions, the weights depending on the ratio of fixed costs a to total expenditure. The first expression, $p_i a_i/a$ is the ith value share when $x = a$; this term thus represents the expenditure patterns of those who can do no more than meet fixed costs, and it carries a weight of unity when $x = a$, decreasing to zero as x tends to infinity. By contrast, $p_i b_i/b$ is the ith value share at "bliss," when x is infinite, and this value carries a weight that increases from zero to one as x increases to infinity. Hence, under quasi homotheticity, actual expenditure patterns are a weighted average of value shares appropriate to very rich and very poor consumers.

Note finally that, at any given point, the cost function (4.4) imposes no restriction on price responses. The functions $a(p)$ and $b(p)$ apart from being linearly homogeneous and concave may take any form whatever, and the substitution matrix will be a utility weighted combination of their Hessians. Consequently, if we believe that poor consumers are less able to substitute between goods than are rich consumers (necessities have fewer substitutes than luxuries), it is possible to choose functions a and b so as to allow for this. The only restriction embodied in (4.4) is the linearity of Engel curves, for which quasi homotheticity is necessary as well as sufficient. And even in cases where linearity seems implausible globally, the demand functions (4.6) may still be excellent local approximations; as formally demonstrated by Diewert (1980a), the cost function (4.4) can act as a local second-order approximation to any arbitrary cost function, so that (4.6) can be regarded as a local first-order approximation to any system of demand functions. We shall return to the Gorman polar form in the next chapter, where it plays a crucial role in the theory of aggregation over consumers.

Exercises

5.22. Homotheticity implies that all expenditure elasticities are unity. What restrictions, if any, does it place on price elasticities?
5.23. Consider the quadratic utility function $u = \Sigma\Sigma(q_k - \beta_k)\,\alpha_{kj}(q_j - \beta_j)$ where the matrix (α_{kj}) is symmetric and negative definite. Show that these preferences are quasi-homothetic.
5.24. The marginal utility of money λ can be characterized by its reciprocal, the marginal cost of utility $\partial c/\partial u$. Show that there exists a normalization of utility such that λ is independent of u if and only if preferences are quasi-homothetic.
5.25. Show that the linear expenditure system is quasi-homothetic and that, within the model, potentialities for substitution are higher for better-off households.
5.26. (Samuelson 1965.) By combining (3.5) with the result of Exercise 5.17, show that if the same preference ordering is both directly and indirectly additive, $w_i + \phi e_i w_i$ must be independent of i. The most general condition under which this holds is $e_i = -\phi = 1$ and show that this condition is necessary and sufficient for homotheticity.
5.27.* Show that preferences are homothetic if and only if the distance function takes the form $d(u, q) = \beta(q)u$ where $\beta(q)$ is a linearly homogeneous concave function. [Hint: use equation (7.4) of Chapter 2.]
5.28.* (Exercise 5.15, cont.) Show that, in general,

$$\frac{\partial}{\partial q_k}\left(\frac{\partial v/\partial q_i}{\partial v/\partial q_j}\right) = \frac{\partial}{\partial q_k}\left(\frac{\partial d/\partial q_i}{\partial d/\partial q_j}\right) + \frac{\partial}{\partial u}\left(\frac{\partial d/\partial q_i}{\partial d/\partial q_j}\right)\frac{\partial u}{\partial q_k}$$

Hence, show that weak and implicit separability can only hold simultaneously if the subgroup utility functions are homothetic.
5.29.* By rewriting (3.3) in the form $s_{ij} = \gamma(\partial q_i/\partial u)(\partial q_j/\partial u)$, or otherwise, show that homotheticity and additivity together imply that $c(u, p) = u(\Sigma\beta_i p_i^{\alpha})^{1/\alpha}$. What values can α legitimately take? [For the extension of this to quasi homotheticity, see Pollak (1971a)].

Bibliographical notes

The link between functional structure and aggregation over commodities seems to have been first drawn out in papers by Sono (1962) and Leontief (1947). Strotz's (1957) paper on the empirical implications of the utility tree prompted Gorman's (1959) characterization of the necessary and sufficient conditions for two-stage budgeting. Much of §5.2 is based on unpublished London School of Economics lecture notes by Gorman (1971) who has also published material on preferences and separability [Gorman (1968)]. Goldman and Uzawa (1964) provide a convenient listing of the empirical consequences of different separable structures. The essentially identical topic in production theory is also discussed by Bliss (1975, Chapter V). For a different approach to the same topics, see Green (1964) [summarized in Green (1976)]. A useful summary is also given by Geary and Morishima (1973). The relationship between rationing and separability is drawn out in papers by Pollak (1969, 1971b). The most complete discussion of implicit separability is an unpublished manuscript by Gorman (1970) although much of the material is in Gorman (1976). Other types of separability and their consequences are discussed by Pearce (1961, 1964) and by Blackorby et al. (1978). Additivity has a long intellectual history although the forerunner of modern analysis is undoubtedly Frisch (1932, 1959). Houthakker's (1960) paper is also important in spite of a record number of misprints. Surveys of estimates of ϕ are contained in Brown and Deaton (1972) and in Sato (1972). The empirical consequences of the additivity assumption are discussed by Deaton (1974b, 1975a, b), while the (very) special cases that allow for inferior goods are discussed by Green (1961). Linear Engel curves have their greatest theoretical importance in aggregation theory, where their history can be traced back to Antonelli (1886) and in modern times to Gorman (1953) and Theil (1954). The Gorman polar form appears in Gorman (1961) and more extensively in Gorman (1976). Homotheticity, like additivity, has long been recognized as an important special case, see, for example, the important early contribution by Bergson (1936). For an interesting application of separability theory to the analysis of international trade flows and prices, see the papers by Armington (1969a, b, 1970), Branson (1972b) and Johnson, Grennes, and Thursby (1977).