Gregory Kordas
Last update: March 21, 2023

LECTURE 1
EMPIRICAL DEMAND ANALYSIS

1. Introduction

`"I never listen to long introductions," said Fräulein Bürstner.`

— Franz Kafka, *The Trial*, Chapter 1.

The theory of consumer choice is rightly viewed as one of the major triumphs of economic theory, for it yields conclusions that are not only mathematically elegant and intuitively satisfying, but are also in accordance with empirical evidence. A number of standard and widely known results have been derived from the neoclassical theory of choice, including the most famous result in economics, the “Law of Demand”, which states that the (income-compensated) demand function for a good (that represents quantity demanded as a function of own price) slopes downward. It also yields a near perfect marriage of theory and econometrics, to a degree unparalleled by any other field of economics. The field has attracted a lot of attention since the introduction of the linear expenditure system and its application to British data by Stone (1954), followed by the differential demand system of Barten (1964) and Theil (1965, 1975/76, 1980) and developments thereafter. Dr. Pangloss would be pleased!

This lecture covers demand systems based on specific utility functions (e.g., the linear expenditure system), demand systems based on indirect utility functions (e.g., the translog), those based on cost functions (e.g., the almost ideal demand system) and some others. It then introduces the differential approach to consumption theory, some members of the class of differential demand equations and some new simpler alternatives. This lecture also discusses separable utility functions and demand equations for groups of goods and for goods within a group.

An economic agent, identified as an individual consumer or as a household unit whose members act jointly, is assumed to allocate an income of y over n market goods $q_i, i = 1, ..., n$, which are purchased at unit prices $p_i, i = 1, ..., n$, in such a way that a “utility” function $u(q_1, ..., q_n)$ defined over the n goods is at a maximum. Much of the elegance of this theory is that one can equally well take the demand functions as the starting point (the demand functions, after all, are what is in principle observable), and, under certain conditions, associate them with a utility function. These conditions (the so-called Slutsky conditions, which are both necessary and sufficient) are that the matrix, whose typical element is

$$s_{ij} = \frac{\partial q_i}{\partial p_j} + q_j \frac{\partial q_i}{\partial y}, \qquad i, j = 1, ..., n.$$

be symmetric and of rank $n-1$. Demand functions that satisfy these conditions are said to be *integrable* (that is, theoretically plausible). Slutsky revealed how utility functions, embarrassingly invisible, might actually be constructed from demand functions, visible or otherwise, provided certain conditions were satisfied.

Still another example of consumption expenditures that are of the 'eliminate - discomfort' variety are ones that are socially induced so as to be able "to appear in public without shame," a notion, incidentally, that can be traced to the Talmud (Tamari, 1987). Related notions include Veblen's Conspicuous Consumption (1899), "bandwagon" and "snob" effects of Leibenstein (1950), and "Keeping-up-with-the-Joneses" (Duesenberry, 1949).

Once the utility function is introduced, demand analysis becomes much richer in its implications and applications. The utility framework is the foundation for index number theory, which includes the measurement of real income; the measurement of the effects of distortions such as commodity taxation; and the division of goods into groups which are closely related. In addition, the utility function generates the three major predictions of demand analysis, (i) that the demand equations are homogeneous of degree zero in income and prices; (ii) symmetry of the substitution effects; and (iii) that the substitution matrix is negative semidefinite. These topics are discussed at length in this lecture.

2. The Neoclassical Theory of Consumer Demand

The debate about ordinal versus cardinal utility has a long history dating from the Marginalist Era. Pareto [3, pp. 159, 540–542] is the first to argue that only ordinality matters for consumer choice theory, but does not make use of this observation. The issue was resolved definitively by Slutsky [18] (see Stigler [19] for a discussion). Hicks and Allen [20, 21] construct a theory of consumer choice using the marginal rate of substitution as the primitive concept rather than the utility function, thereby avoiding cardinality.

Given some rather mild regularity conditions on preferences[1] (completeness, transitivity, continuity, reflexivity, strong monotonicity and strict convexity[2]) it can be shown that the *preference relation* $\boldsymbol{q} \succeq \boldsymbol{q}'$ ($\boldsymbol{q}, \boldsymbol{q}' \in \mathbb{R}^n$), which is read as "the bundle $\boldsymbol{q}$ is weakly preferred to bundle $\boldsymbol{q}'$", can be rationalized by a continuous, non-decreasing, strictly quasi-concave[3] *utility function* $u : \mathbb{R}^n_+ \mapsto \mathbb{R}_+$ such that,

$$\boldsymbol{q} \succeq \boldsymbol{q}' \quad \text{if and only if} \quad u(\boldsymbol{q}) \geq u(\boldsymbol{q}').$$

For the derivation of the utility function from primitive preferences see Varian (1992). A counterexample of preferences that cannot be represented by a utility function are *lexicographic* preferences, which are quite extreme and will not concern us here.

---

[1] *De gustibus non est disputandum.* Preferences are taken as a primitive: how they are formed and what they are is not part of the science of economics. The above admonition *not to quarrel over tastes* is commonly interpreted as advice to terminate a dispute when it has been resolved into a difference of tastes. The question of our likes and dislikes, and whether we should allow them to influence our search for truth, is the subject of an anecdote quoted by the mathematician Littlewood: The philosophers Russell and Moore were having one of their many philosophical discussions. Suddenly Russell said : "Moore, you don't like me, do you?" "No," said Moore. The discussion then continued without a further word on the point from either side.

[2] Theorem (Debreu, 1954): *A binary preference relation $\succeq$ can be represented by a continuous real-valued utility function $u : \mathbb{R}^n_+ \to \mathbb{R}$ if and only if it is complete, transitive and continuous.*
For a proof of this theorem see Barten and Böhm (1982) and Jehle and Reny (2011), p.13. It can be shown that if u represents the preference ordering $\succeq$, then $\succeq$ is complete and transitive. But the reverse is not generally true, that is, there are preference orderings $\succeq$ that are complete and transitive but cannot be represented by a utility function. For example, lexicographic preferences are complete and transitive (that's why we have usable dictionaries) but they cannot be represented by a utility function. Continuity of the ordering $\succeq$ rules out lexicographic preferences and makes u a continuous function. The rest of the conditions on $\succeq$ listed above guarantee monotonicity and quasi-concavity of u.(see Mandy, David M - Producers, consumers, and partial equilibrium-Academic Press (2017) p. 190-194 and p.204)

[3] That is, we require that

$$\frac{\partial u}{\partial q_i} \geq 0, \text{ and } \left.\frac{d^2 q_i}{d q_j^2}\right|_{u=u_0} \geq 0, \quad \text{for all } q_i, q_j \geq 0, i, j = 1, ..., n.$$

Quasi-concavity of u guarantees that the corresponding indifference curves (isoquants of u) are convex to the origin, so that the budget line $y = \boldsymbol{p}'\boldsymbol{q}$ is tangent to at most one of these curves, possibly at more that one points. If we strengthen the assumption of quasi-concavity to strict quasi-concavity of u, the tangency occurs at only one point. Recall that indifference curves are solutions to the equation

$$du = \sum_{i=1}^{n} \frac{\partial u(\boldsymbol{q})}{\partial q_i} dq_i = 0, \text{ such that } u(\boldsymbol{q}) = u_0.$$

The utility function $u(\cdot)$ need not convey any *cardinal* information about utility since any strictly increasing transformation of $u(\boldsymbol{q})$, say $\log u(\boldsymbol{q})$ or $\sqrt{u(\boldsymbol{q})}$, yields the same ordering of bundles $\boldsymbol{q}$. We say that u is an *ordinal* function, which is to say that, it is determined only up to a strictly increasing transformation, so that, for example, $u(q_1, q_2) = q_1^a q_2^b$ and $u^*(q_1, q_2) := \log u(q_1, q_2) = a \log q_1 + b \log q_2$ represent the same preferences and is thus the "same" utility function. We say that u and every strictly increasing transformation of it u^* belong to the same *class* of utility functions.

To characterize each class we use an *invariant* of the class, the marginal rate of substitution. To define it, suppose we increase the consumption of good i and ask: how does the consumer has to change his consumption of good j in order to keep utility constant, i.e., stay on the same indifference curve? By assumption, the change in utility must be zero, i.e., $du = 0$ which means that for $i \neq j$,

$$\frac{\partial u(\boldsymbol{q})}{\partial q_i} dq_i + \frac{\partial u(\boldsymbol{q})}{\partial q_j} dq_j = 0 \qquad i, j = 1, ..., n.$$

Therefore[4]

$$MRS_{ij} := \frac{dq_j}{dq_i} = -\frac{\partial u(\boldsymbol{q})}{\partial q_i} \Big/ \frac{\partial u(\boldsymbol{q})}{\partial q_j} = -\frac{u_{[i]}}{u_{[j]}}, \quad i, j = 1, ..., n \tag{2.1}$$

provided that the marginal utilities $u_{[i]} = \partial u(\boldsymbol{q})/\partial q_i$ are strictly positive for all $i = 1, ..., n$. This is the *marginal rate of substitution* of good i with good j, and gives us the amount by which we have to decrease (notice the minus sign) the consumption of good j in order to counterbalance the increase in consumption of good i by dq_i and remain on the same indifference curve, that is $dq_j = MRS_{ij}\, dq_i$. Geometrically, the MRS is the slope of the indifference curve along which we move in our thought experiment above.

The MRS does not depend on the specific utility function used to represent the underlying preferences, but is common across all members of the same class of utilities. To see this, let $u^*(\boldsymbol{q}) = g(u(\boldsymbol{q}))$ be a strictly increasing transformation of $u(\boldsymbol{q})$. The marginal rate of substitution for the two utility functions u and u^* is the same since

$$\begin{aligned} -\frac{\partial u^*(\boldsymbol{q})}{\partial q_i} \Big/ \frac{\partial u^*(\boldsymbol{q})}{\partial q_j} &= -g'(u) \frac{\partial u(\boldsymbol{q})}{\partial q_i} \Big/ \left[g'(u) \frac{\partial u(\boldsymbol{q})}{\partial q_j} \right] \\ &= -\frac{\partial u(\boldsymbol{q})}{\partial q_i} \Big/ \frac{\partial u(\boldsymbol{q})}{\partial q_j}, \end{aligned}$$

[4]This a consequence of the Implicit Function Theorem that, provided $u(\boldsymbol{q})$ is C^1 and strictly quasiconcave in $\boldsymbol{q}$, guaranties the existence of C^1 functions $q_j(\boldsymbol{q}_{-j})$ for all $j = 1, ..., n$, where $\boldsymbol{q}_{-j}$ is the vector $\boldsymbol{q}$ with the jth component excluded, and that their derivatives with respect to q_i are given by $-u_{[i]}/u_{[j]}$.

if and only if $g'(u) > 0$, i.e., g is strictly increasing. This gives a useful way to recognize preferences that are represented by different utility functions: given two utility functions, just compute the MRS for each one to see if they are the same. If they are, then the two utility functions have the same indifference curves, and so the underlying preferences are the same.

Example 1. Consider the following two ways to write the Cobb-Douglas utility function: $u(q_1, q_2) = q_1^a\, q_2^b$, and $u^*(q_1, q_2) = a \log q_1 + b \log q_2$. To see that they represent the same preferences, we compute MRS in each case and verify that they are equal:

$$\begin{aligned} MRS_{12} &= -\frac{\partial u(q_1,q_2)}{\partial q_1} \Big/ \frac{\partial u(q_1,q_2)}{\partial q_2} \\ &= -\frac{a q_1^{a-1} q_2^b}{b q_1^a q_2^{b-1}} = -\frac{a q_2}{b q_1}, \end{aligned}$$

and

$$\begin{aligned} MRS_{12}^* &= -\frac{\partial u^*(q_1,q_2)}{\partial q_1} \Big/ \frac{\partial u^*(q_1,q_2)}{\partial q_2} \\ &= -\frac{a/q_1}{b/q_2} = -\frac{a q_2}{b q_1}. \end{aligned}$$

■

Since several functions may be used to represent the same preferences, it is useful to find the "simplest" function to represent a particular set of preferences. For example, the discussion above shows that Cobb-Douglas preferences depend only on the ratio a/b, and not on a and b separately. This means that, without loss of generality, we may write the general Cobb-Douglas utility function for two goods as $u(q_1, q_2) = q_1^\alpha q_2^{1-\alpha}$, so $MRS_{1,2} = -\alpha/(1-\alpha) \cdot (q_2/q_1)$ (in terms of a and b, $\alpha = a/(a+b)$). In the n commodities case, the Cobb-Douglas utility function may be written as $u(\boldsymbol{q}) = \prod_{i=1}^n q_i^{\alpha_i}$, with $0 < \alpha_i < 1$, $\sum_{i=1}^n \alpha_i = 1$.[5] Moreover, since in its logarithmic form

$$u(\boldsymbol{q}) = \sum_{i=1}^n \alpha_i \log q_i, \quad \text{with} \quad q_i > 0,\ 0 < \alpha_i < 1,\ \sum_{i=1}^n \alpha_i = 1,$$

the Cobb-Douglas function is seen to be *additive-separable*, this is our preferred way of representing the class of Cobb-Douglas preferences.

[5]The Cobb-Douglas power function is also used to represent production functions in production theory, but since production functions are *cardinal*, the normalization $\sum_{i=1}^n \alpha_i = 1$ is not imposed there. In fact, this condition corresponds to constant returns-to-scale which is of course a restriction, since returns-to-scale in any given situation might be increasing, constant, or decreasing.

Before ending our discussion of the ordinality of the utility function, it is expedient to introduce one more concept that we will use shortly, namely the *elasticity of substitution*, σ_{ij}, between goods i and j, introduced by John Hicks (1932) and defined by

$$\sigma_{ij} := \frac{d(q_j/q_i)}{(q_j/q_i)} \Bigg/ \frac{d(dq_j/dq_i)}{(dq_j/dq_i)} = MRS_{ij} \cdot \frac{q_i}{q_j}.$$

It measures the *curvature* of an indifference curve and thus the substitutability between goods, i.e. how easy it is to substitute one good for the other. For the Cobb-Douglas class of utilities, $MRS_{ij} = -(\alpha_j/\alpha_i) \cdot (\alpha_i q_j/\alpha_j q_i)$, so $\sigma_{ij} = 1$.

The elasticity of substitution is defined as the ratio of the proportionate change in good proportions to the proportionate change in the slope of the indifference curve. Good proportions are q_i/q_j and the change in good proportions is $d(q_i/q_j)$, hence the proportionate change in good proportions is $d(q_i/q_j)/(q_i/q_j)$. The slope of the indifference curve is dq_i/dq_j and the change in that slope is $d(dq_i/dq_j)$, hence the proportionate change of slope is $d(dq_i/dq_j)/(dq_i/dq_j)$. Putting all this together, we get that the elasticity of substitution between goods i and j is:

$$\sigma_{ij} := \frac{d\left(\frac{q_i}{q_j}\right) \Big/ \left(\frac{q_i}{q_j}\right)}{d\left(\frac{dq_i}{dq_j}\right) \Big/ \left(\frac{dq_i}{dq_j}\right)}. \tag{2.2}$$

This may look like a fairly complicated expression but as we shall see, it in fact turns out to be very simple when applied to many standard forms of utility function.

2.1. The Primal Consumer Behavior Problem – Marshallian Demands

Given a twice continuously differentiable, strictly increasing, strictly quasiconcave utility function $u(\boldsymbol{q}) : \mathbb{R}^n_+ \to \mathbb{R}_+$ representing preferences over the vector of goods $\boldsymbol{q}$, we may formulate "consumer behavior" as the problem

$$\text{[UMP]:} \qquad \max_{\boldsymbol{q} \in \mathbb{R}^n_+} \; u(\boldsymbol{q}), \quad \text{subject to} \quad \boldsymbol{p}^\mathsf{T} \boldsymbol{q} \leq m, \tag{2.3}$$

called the *Utility Maximization Problem* [UMP], where $\boldsymbol{p} = (p_1, ..., p_n)^\mathsf{T} \in \mathbb{R}^n_{++}$ is a vector of strictly positive prices of the positive coordinate commodities $\boldsymbol{q} = (q_1, ..., q_n)^\mathsf{T} \in \mathbb{R}^n_+$, and $m > 0$ denotes money income. Both prices $\boldsymbol{p} \gg 0$ and money income $m > 0$ are regarded as given from the standpoint of the consumer. We abstract from various complications like nonlinear pricing, intertemporal decision making, conspicuous consumption, etc., etc. We will assume that the utility function is twice continuously differentiable and that there is *nonsatiation*, so

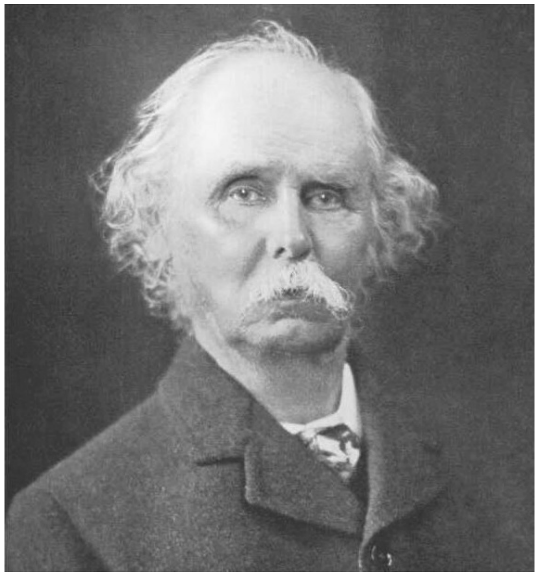


Figure 1. Alfred Marshall (1842–1924)

that the marginal utility for each product is positive,

$$\frac{\partial u(\boldsymbol{q})}{\partial q_i} > 0, \qquad i = 1, ..., n.$$

We further assume that there is *generalized diminishing marginal utility*, so that the Hessian matrix of second derivatives of u

$$\boldsymbol{U} := \left[\frac{\partial^2 u(\boldsymbol{q})}{\partial q_i \partial q_j}\right]_{i,j=1,\ldots,n} = \nabla_{\boldsymbol{q}\boldsymbol{q}^\mathsf{T}} u(\boldsymbol{q}) \tag{2.4}$$

is a symmetric negative definite $n \times n$ matrix.

Since a continuous function attains its maximum on a compact set, and (i) u is continuous and (ii) the set $\{\boldsymbol{q} \in \mathbb{R}^n_+ : \boldsymbol{p}^\mathsf{T}\boldsymbol{q} \leq m\}$ defining the budget constraint is compact (i.e., closed and bounded), the set of solutions of the problem in (2.3) is nonempty. Furthermore, by the strict convexity of the preference relation $\succeq$ assumed here, u is strictly quasi-concave, and by nonsatiation it is also monotone increasing, and thus the solution is unique.[6]

[6]If preferences are only convex but not strictly convex, the utility function u is quasi-concave but not strictly so, in which case the utility maximization problem in (2.3) may have multiple solutions. If solutions are non-unique, the *demand functions* studied below become *demand correspondences*, the study of which requires more advanced mathematical methods than the elementary calculus-based methods used here. We will not consider this more general setting here, however, since our goal is to arrive at empirically implementable equations and not to give the most general theoretical results.

Nonsatiation means that the budget constraint will be satisfied with equality, and the Lagrangian of the so-called "primal" problem in (2.3) is given by

$$\mathscr{L}(\boldsymbol{q},\lambda;\boldsymbol{p},m)=u(\boldsymbol{q})+\lambda(m-\boldsymbol{p}^{\mathsf{T}}\boldsymbol{q}), \tag{2.5}$$

where λ is a scalar Lagrange multiplier. Differentiation with respect to the elements of $\boldsymbol{q}$ and λ yields the $(n+1)$ system of simultaneous equations

$$\begin{aligned}\frac{\partial u(\boldsymbol{q}^*)}{\partial q_i}-\lambda^* p_i &= 0, \qquad i=1,...,n, \\ m-\sum_{i=1}^{n} p_i q_i^* &= 0.\end{aligned} \tag{2.6}$$

Dividing the ith by the jth equation we get

$$\frac{\partial u(\boldsymbol{q}^*)/\partial q_i}{\partial u(\boldsymbol{q}^*)/\partial q_j}=\frac{p_i}{p_j}, \quad i,j=1,...,n,i\neq j,$$

which says that the optimal consumption bundle $\boldsymbol{q}^*=(q_1^*,q_2^*,...,q_n^*)^{\mathsf{T}}$ is characterized by the conditions

$$MRS_{ij}|_{\boldsymbol{q}^*}=-\frac{p_i}{p_j}, \qquad i\neq j,\ i,j=1,...,n. \tag{2.7}$$

Here $MRS_{ij}|_{\boldsymbol{q}^*}$ is the slope of the indifference curves at the optimal consumption $\boldsymbol{q}^*$ and $-p_i/p_j$ is the slope of the budget constraint, so (2.7) says that at the optimal level of consumption the highest attainable indifference curve should be *tangent* to the budget constraint. Note that the optimality characterization of $\boldsymbol{q}^*$ depends on $u()$ only through its MRS, and is thus invariant to a strictly increasing transformation $g(u(\boldsymbol{q}))$, as the ordinality of our concept of utility requires.

It is customary to write the first-order conditions for utility maximization simply as

$$\begin{aligned}u_{[i]} &= \lambda^* p_i, \qquad i=1,...,n, \\ m &= \sum_{i=1}^{n} p_i q_i^*,\end{aligned} \tag{2.8}$$

where $u_{[i]}$ denotes the partial derivative of $u(\boldsymbol{q})$ with respect to q_i evaluated at $\boldsymbol{q}^*$. In words, this system of equations says that at the optimal consumption $\boldsymbol{q}^*$, the marginal utility from consuming the ith good, u_i, is proportional to its price p_i, with the constant of proportionality $\lambda^*>0$ being equal to the marginal utility of income (see below) for all $i=1,...,n$, and that this consumption is feasible, i.e., it exhausts the available money income m.

The second-order conditions for a *local maximum* require the Hessian matrix of the Lagrangian function $\mathscr{L}$ with respect to $\boldsymbol{q}$ at $(\boldsymbol{q}^*,\lambda^*)$ be *negative definite* for all directions $\boldsymbol{z}_1$ that are *orthogonal* to the gradient of the constraint function, that is,

$$\boldsymbol{z}_1^\top (\nabla_{\boldsymbol{q}\boldsymbol{q}^\top}\mathscr{L})\,\boldsymbol{z}_1 = \boldsymbol{z}_1^\top \boldsymbol{U}\,\boldsymbol{z}_1 < 0 \text{ for all } \boldsymbol{z}_1 \in \mathbb{R}^n_{++} \text{ such that } \boldsymbol{z}_1^\top \nabla_{\boldsymbol{q}}\eta(\boldsymbol{q}) = -\boldsymbol{z}_1^\top \boldsymbol{p} = 0, \tag{2.9}$$

where $\eta(\boldsymbol{q}) = m - \boldsymbol{p}^\top \boldsymbol{q}$ is the budget line. Thus, we require that $d^2u < 0$ for all $\boldsymbol{dq}$ satisfying $d\eta = -\boldsymbol{p}^\top \boldsymbol{dq} = 0$. In words, this means that utility is decreasing as we move away from the optimal $\boldsymbol{q^*}$ in any direction along the budget line $\eta(\boldsymbol{q})$, or equivalently, in any direction orthogonal to the gradient of the budget line $\nabla_{\boldsymbol{q}}\eta = -\boldsymbol{p}$.

Consider the Hessian matrix of the Lagrangian $\mathscr{L}$ evaluated at $(\boldsymbol{q}^*, \lambda^*)$, given by

$$\bar{\boldsymbol{H}} = \nabla_{(\lambda,\boldsymbol{q})(\lambda,\boldsymbol{q})^\top}\mathscr{L}(\boldsymbol{q}^*, \lambda^*; \boldsymbol{p}, y) = \begin{bmatrix} 0 & \eta_{[1]} & \eta_{[2]} & \cdots & \eta_{[n]} \\ \eta_{[1]} & u_{[11]} & u_{[12]} & \cdots & u_{[1n]} \\ \eta_{[2]} & u_{[21]} & u_{[22]} & \cdots & u_{[2n]} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \eta_{[n]} & u_{[n1]} & u_{[n2]} & \cdots & u_{[nn]} \end{bmatrix} \tag{2.10}$$

$$= \begin{bmatrix} 0 & -p_1 & -p_2 & \cdots & -p_n \\ -p_1 & u_{[11]} & u_{[12]} & \cdots & u_{[1n]} \\ -p_2 & u_{[21]} & u_{[22]} & \cdots & u_{[2n]} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -p_n & u_{[n1]} & u_{[n2]} & \cdots & u_{[nn]} \end{bmatrix} = \begin{bmatrix} 0 & -\boldsymbol{p}^\top \\ -\boldsymbol{p} & \boldsymbol{U}, \end{bmatrix}, \tag{2.11}$$

where function subscripts in square brackets $[\cdot]$ denote partial differentiation with respect to the corresponding q_i evaluated at $(\boldsymbol{q}^*, \boldsymbol{\lambda}^*)$. The $(n+1) \times (n+1)$ matrix $\bar{\boldsymbol{H}}$ is called *bordered Hessian* (and we write it as $\boldsymbol{H}$ with a bar above it) because the submatrix $\nabla_{\boldsymbol{q}\boldsymbol{q}^\top}\mathscr{L} = \boldsymbol{U}$ is "bordered" by the gradient vector of the constraint function $\nabla_{\boldsymbol{q}}\eta = -\boldsymbol{p}$. Recall[7] that the matrix $\boldsymbol{U} = \nabla_{\boldsymbol{q}\boldsymbol{q}^\top}\mathscr{L}$ is negative definite on the linear subspace $\{\boldsymbol{z} \in \mathbb{R}^n_{++} : \boldsymbol{p}^\top \boldsymbol{z} = 0\}$ if the last $n-1$ *leading principal submatrices* of $\bar{\boldsymbol{H}}$ have determinants (called *leading principal minors*) that are nonzero and alternate in sign starting from a positive sign, as follows:

$$|\bar{\boldsymbol{H}}_3| = \begin{vmatrix} 0 & -p_1 & -p_2 \\ -p_1 & u_{[11]} & u_{[12]} \\ -p_2 & u_{[21]} & u_{[22]} \end{vmatrix} > 0, \quad |\bar{\boldsymbol{H}}_4| = \begin{vmatrix} 0 & -p_1 & -p_2 & -p_3 \\ -p_1 & u_{[11]} & u_{[12]} & u_{[13]} \\ -p_2 & u_{[21]} & u_{[22]} & u_{[23]} \\ -p_3 & u_{[31]} & u_{[32]} & u_{[33]} \end{vmatrix} < 0, \quad \text{and so on...}, \tag{2.12}$$

with the last determinant $|\bar{\boldsymbol{H}}_{n+1}| = |\bar{\boldsymbol{H}}|$ being nonzero and having the same sign as $(-1)^n$. [8]

[7] See Theorem 16.4 (p. 389), Theorem 16.5 (p. 391), and Theorem 19.6 (p. 460) of Simon and Blume (1994).

[8] Recall that if $f : U \to \mathbb{R}$ is a C^2 function of n variables $\boldsymbol{x} = (x_1, ..., x_n)^\top$ and $\boldsymbol{x}^*$ is an interior point of U, an open subset of $\mathbb{R}^n$, such that $\nabla f(\boldsymbol{x}^*) = \boldsymbol{0}$ then

Aside. A common error[9] is to state the second order conditions for a constrained local maximum as requiring that $\bar{\boldsymbol{H}} = \nabla_{(\lambda,\boldsymbol{q})(\lambda,\boldsymbol{q})^\mathsf{T}} \mathscr{L}(\boldsymbol{q}^*, \lambda^*; \boldsymbol{p}, y)$ be *negative semidefinite*, i.e., that $\boldsymbol{z}^\mathsf{T}\, \bar{\boldsymbol{H}}\, \boldsymbol{z} \leq 0$ for all $\boldsymbol{z} \in \mathbb{R}^{n+1}_{++}$. The correct condition involves checking *only the last* $n-1$ of the $n+1$ leading principal minors of $\bar{\boldsymbol{H}}$, and verifying that they are nonzero and alternate in sign as indicated above, and *not all of them*, as the mistaken condition that $\bar{\boldsymbol{H}}$ be negative semidefinite would require, which also allows them to be zero. Also, the mistaken condition that $\bar{\boldsymbol{H}}$ be *negative semidefinite*, sounds a lot like the correct condition that $\boldsymbol{U} = \nabla_{\boldsymbol{q}\boldsymbol{q}^\mathsf{T}} \mathscr{L}$ be negative definite on the linear subspace $\{\boldsymbol{z} \in \mathbb{R}^n_{++} : \boldsymbol{z}^\mathsf{T} \nabla_{\boldsymbol{q}} \eta(\boldsymbol{q})) = 0\}$.

Thus, we must check only the minors starting from 3 and higher $|\bar{\boldsymbol{H}}_3|, |\bar{\boldsymbol{H}}_4|, ..., |\bar{\boldsymbol{H}}_{n+1}|$. The condition for a local maximum says nothing about the first 2 minors $|\bar{\boldsymbol{H}}_1| = 0$ and $|\bar{\boldsymbol{H}}_2| = (-\eta_{[1]})^2 = -p_1^2 < 0$, that are *always* 0 and negative, respectively. This *cannot* be summarized by requiring that $\bar{\boldsymbol{H}}$ be *negative semidefinite* because then $|\bar{\boldsymbol{H}}_3|, |\bar{\boldsymbol{H}}_4|, ..., |\bar{\boldsymbol{H}}_{n+1}|$ would also be allowed to be zero, as $|\bar{\boldsymbol{H}}_1| = 0$ is, which is *not* the case: $|\bar{\boldsymbol{H}}_k|, k = 3, 4, ..., n+1$ must be nonzero and alternate in sign staring from positive. The error stems from the desire to sum-up in an easy condition what it is that is required, but the negative semidefiniteness condition doesn't accomplish this, since in order to accommodate the zero in the 1st minor, also allows zero minors from 3 and up, which is not correct.

This discussion generalizes directly to, say, m constraints by requiring that we check that the last $n-m$ of the $n+m$ leading principal minors of the $(n+m) \times (n+m)$ border Hessian $\bar{\boldsymbol{H}}$, given by $|\bar{\boldsymbol{H}}_k|, k = m+2, ..., m+n$, be nonzero and alternate in sign starting from $(-1)^{m+1}$ and ending in $(-1)^n$. Again, this *cannot* be summarized by requiring that $\bar{\boldsymbol{H}}$ be negative semidefinite.

What is not widely known is that it is indeed possible to write a condition in terms of the definiteness of some matrix that summarizes correctly the sufficient conditions for a constrained local maximum, and thus avoid all the discussion about which principal minors need be considered and their signs. Consider the constrained maximization problem

$$\max_{\boldsymbol{x} \in A \subset \mathbb{R}^n} f(\boldsymbol{x}; \boldsymbol{\alpha}) \quad \text{s.t.} \quad \boldsymbol{g}(\boldsymbol{x}; \boldsymbol{\alpha}) \leq \boldsymbol{y}, \tag{2.13}$$

(a) If the Hessian $\nabla_{\boldsymbol{x}\boldsymbol{x}} f(\boldsymbol{x}^*)$ is a negative definite symmetric matrix, $\boldsymbol{x}^*$ is a strict local maximum of f;

(b) If the Hessian $\nabla_{\boldsymbol{x}\boldsymbol{x}} f(\boldsymbol{x}^*)$ is a positive definite symmetric matrix, $\boldsymbol{x}^*$ is a strict local minimum of f; and

(c) If $\nabla_{\boldsymbol{x}\boldsymbol{x}} f(\boldsymbol{x}^*)$ is indefinite, then $\boldsymbol{x}^*$ is neither a local maximum nor a local minimum of f;

[9]For an instance of this error see p. 481 of Kalman, P.J. and Intriligator, M.D. (1973), "Generalized Comparative Statics with Applications to Consumer Theory and Producer Theory", *International Economic Review*, **14**:473–486. Caputo (2000).

and its Lagrangian

$$\mathscr{L}(\boldsymbol{x},\boldsymbol{\lambda};\boldsymbol{\alpha},\boldsymbol{y}) = f(\boldsymbol{x};\boldsymbol{\alpha}) + \boldsymbol{\lambda}^{\mathsf{T}}[\boldsymbol{y} - \boldsymbol{g}(\boldsymbol{x};\boldsymbol{\alpha})]. \tag{2.14}$$

The bordered Hessian for of the problem is then given by

$$\bar{\boldsymbol{H}} = \begin{bmatrix} \mathbf{0} & (-\nabla_{\boldsymbol{x}}\boldsymbol{g})^{\mathsf{T}} \\ -\nabla_{\boldsymbol{x}}\boldsymbol{g} & \nabla_{\boldsymbol{x}\boldsymbol{x}^{\mathsf{T}}}\mathscr{L} \end{bmatrix}. \tag{2.15}$$

Assuming that $\bar{\boldsymbol{H}}$ is nonsingular, the inverse $\bar{\boldsymbol{H}}^{-1}$ exists and is given by

$$\begin{aligned} \bar{\boldsymbol{H}}^{-1} &= \begin{bmatrix} \mathbf{0} & (-\nabla_{\boldsymbol{x}}\boldsymbol{g})^{\mathsf{T}} \\ -\nabla_{\boldsymbol{x}}\boldsymbol{g} & \nabla_{\boldsymbol{x}\boldsymbol{x}^{\mathsf{T}}}\mathscr{L} \end{bmatrix}^{-1} \\ &= \begin{bmatrix} \boldsymbol{A} & \boldsymbol{A}\,(\nabla_{\boldsymbol{x}}\boldsymbol{g})^{\mathsf{T}}\,(\nabla_{\boldsymbol{x}\boldsymbol{x}^{\mathsf{T}}}\mathscr{L})^{-1} \\ \boldsymbol{A}\,(\nabla_{\boldsymbol{x}\boldsymbol{x}^{\mathsf{T}}}\mathscr{L})^{-1}\,(\nabla_{\boldsymbol{x}}\boldsymbol{g}) & \boldsymbol{D} \end{bmatrix} \end{aligned}$$

where

$$\boldsymbol{A} = -\left[(\nabla_{\boldsymbol{x}}\boldsymbol{g})^{\mathsf{T}}\,(\nabla_{\boldsymbol{x}\boldsymbol{x}^{\mathsf{T}}}\mathscr{L})^{-1}\,(\nabla_{\boldsymbol{x}}\boldsymbol{g})\right]^{-1}, \tag{2.16}$$

and

$$\boldsymbol{D} = (\nabla_{\boldsymbol{x}\boldsymbol{x}^{\mathsf{T}}}\mathscr{L})^{-1} + \boldsymbol{A}\,(\nabla_{\boldsymbol{x}\boldsymbol{x}^{\mathsf{T}}}\mathscr{L})^{-1}\,(\nabla_{\boldsymbol{x}}\boldsymbol{g})\,(\nabla_{\boldsymbol{x}}\boldsymbol{g})^{\mathsf{T}}\,(\nabla_{\boldsymbol{x}\boldsymbol{x}^{\mathsf{T}}}\mathscr{L})^{-1}. \tag{2.17}$$

Since the second-order sufficient condition holds at the optimum, it follows from a theorem in Takayama (1985, footnote 16, p. 166) that $\boldsymbol{D}$ is symmetric and negative semidefinite.

Theorem 1. *The $n \times n$ matrix $\nabla^2_{\boldsymbol{x}\boldsymbol{x}^{\mathsf{T}}}\mathscr{L}$ is negative (positive) definite on the constraint set $\{\boldsymbol{x} : g(\boldsymbol{x}) = y\}$ if and only if $\bar{\boldsymbol{H}}^{-1}$ exists, and the $n \times n$ symmetric submatrix $\boldsymbol{D}$ has rank $(n-m)$ and is negative (positive) semidefinite.*

Proof: See Samuelson, p. 378-379. ■

This concludes this rather along Aside, and we turn to applying it to our [UMP]. ■

That $\bar{\boldsymbol{H}}$ is indeed negative semidefinite is guaranteed by our assumption that $u(\boldsymbol{q})$ is strictly increasing and strictly quasiconcave in $\boldsymbol{q}$. To see this note that by the FOC in (2.8) we have that at the optimal consumption $\boldsymbol{q}^*$, $p_i = u_{[i]}/\lambda$, so that

$$|\bar{\boldsymbol{H}}| = \begin{vmatrix} 0 & -p_1 & \cdots & -p_n \\ -p_1 & u_{[11]} & \cdots & u_{[1n]} \\ \vdots & \vdots & \ddots & \vdots \\ -p_n & u_{[n1]} & \cdots & u_{[nn]} \end{vmatrix} = \begin{vmatrix} 0 & -u_{[1]}/\lambda & \cdots & -u_{[n]}/\lambda \\ -u_{[1]}/\lambda & u_{[11]} & \cdots & u_{[1n]} \\ \vdots & \vdots & \ddots & \vdots \\ -u_{[n]}/\lambda & u_{[n1]} & \cdots & u_{[nn]} \end{vmatrix} \tag{2.18}$$

$$= \frac{1}{\lambda^2} \begin{vmatrix} 0 & u_{[1]} & \cdots & u_{[n]} \\ u_{[1]} & u_{[11]} & \cdots & u_{[1n]} \\ \vdots & \vdots & \ddots & \vdots \\ u_{[n]} & u_{[n1]} & \cdots & u_{[nn]} \end{vmatrix} = \frac{1}{\lambda^2} \begin{vmatrix} 0 & \boldsymbol{u}^\top \\ \boldsymbol{u} & \boldsymbol{U} \end{vmatrix}, \tag{2.19}$$

where we have pulled one $-1/\lambda$ out from the last column and another $-1/\lambda$ from the last row of $|\bar{\boldsymbol{H}}|$, and $\boldsymbol{u} = \nabla_{\boldsymbol{q}} u(\boldsymbol{q})$ is the $n \times 1$ gradient vector of $u(\boldsymbol{q})$ and $\boldsymbol{U} = \nabla^2_{\boldsymbol{q}\boldsymbol{q}^\top} u(\boldsymbol{q})$ is the $n \times n$ Hessian matrix of $u(\boldsymbol{q})$, as above. Therefore,

$$|\bar{\boldsymbol{H}}| = \frac{1}{\lambda^2} |\bar{\boldsymbol{U}}|, \tag{2.20}$$

where $\bar{\boldsymbol{U}}$ is the *bordered Hessian* of $u(\boldsymbol{q})$. *Therefore, if $u(\boldsymbol{q})$ is (strictly) quasiconcave the Lagrangian $\mathscr{L}$ is also (strictly) quasiconcave.* A necessary condition for u to be a quasi-concave function is that the even-numbered principle minors of the bordered Hessian $\bar{\boldsymbol{U}}$ be non-negative and the odd-numbered principle minors of $\bar{\boldsymbol{U}}$ be non-positive. A sufficient condition for u to be quasi-concave is that the even-numbered principle minors of the bordered Hessian be strictly positive and the odd-numbered principle minors be strictly negative.

3.4 Quasiconcavity and quasiconvexity

https://mjo.osborne.economics.utoronto.ca/index.php/tutorial/index/1/qcc/t

Simon, Blume (1994) – Mathematics for economists, theorem 22.1, p.545

Arrow and Enthoven (1961), Theorem 5 (p. 797).

https://en.wikipedia.org/wiki/Concavification

The first-order conditions for a maximum in (2.8) is a system of $(n+1)$ nonlinear equations in $2(n+1)$ unknowns, namely $(\boldsymbol{q}^\top, \boldsymbol{p}^\top, \lambda, m)^\top$. The second-order conditions (2.12) for a clean interior maximum yield that the following Jacobian is nonzero

$$J = |\bar{\boldsymbol{H}}_n| = \begin{vmatrix} \boldsymbol{U} & -\boldsymbol{p} \\ -\boldsymbol{p}^\top & 0, \end{vmatrix} \neq \boldsymbol{0}, \tag{2.21}$$

where $\boldsymbol{U}$ is defined in (2.4). Taking the $(n+1)$ variables $(\boldsymbol{p}^{\mathsf{T}}, m)^{\mathsf{T}}$ as fixed data (or parameters), the *Implicit Function Theorem* (IFT) now yields that the remaining $(n+1)$ variables $(\boldsymbol{q}^{\mathsf{T}}, \lambda)^{\mathsf{T}}$ may be written uniquely as functions of the parameters $(\boldsymbol{p}^{\mathsf{T}}, m)^{\mathsf{T}}$ in a neighborhood of the currently prevailing prices and income $(\boldsymbol{p}^0, m^0)^{\mathsf{T}}$ as

$$\begin{aligned} \boldsymbol{q}^* &= \boldsymbol{g}(\boldsymbol{p}, m) = (g_1(\boldsymbol{p}, y), ..., g_n(\boldsymbol{p}, m))^{\mathsf{T}} \\ \lambda^* &= f(\boldsymbol{p}, m), \end{aligned} \tag{2.22}$$

where $\boldsymbol{g}\,(\boldsymbol{p}, m) = (g_1(\boldsymbol{p}, m), ..., g_n(\boldsymbol{p}, m))^{\mathsf{T}}$ is a $n \times 1$ vector of functions called the *uncompensated*, or *ordinary*, or *Marshallian demands*, and λ^* is the *marginal utility of income.*[10] Moreover, these functions possess continuous first partial derivatives in this neighborhood of $(\boldsymbol{p}^0, m^0)^{\mathsf{T}}$.

The Lagrange multiplier λ^* has an interesting economic interpretation: it is, as Alfred Marshall called it, the *marginal utility of income.* To see what this means, note that from the first order conditions

$$\frac{\partial u}{\partial q_i} = \lambda^* p_i, \quad i = 1, ..., n,$$

so that

$$\lambda^* = \frac{\partial u(\boldsymbol{q}^*)}{\partial (q_i p_i)}, \quad i = 1, ..., n.$$

Since $(p_i q_i)$ is the expenditure on good i, the above says that, at the optimal level of consumption, an additional euro spent on *any* of the goods yields the same increase λ in utility. We can therefore say that λ is the change in the maximized value of utility as income changes, or that, at the optimal level of consumption

$$\lambda^* = \frac{\partial u(\boldsymbol{q}^*)}{\partial m}.$$

Thus, λ is the increase in utility resulting from an additional euro of income, justifying Marshall's interpretation of λ^* as the marginal utility of income[11].

---

[10] The custom of settling the problem of existence of an equilibrium by counting equations and unknowns, that is, by verifying that they are equal, and then appealing to the *IFT* to obtain a solution is very old. In the present case the *IFT* is adequate, but in more general situations such simple calculus-based methods are not powerful enough. Modern general equilibrium theory utilizes more advanced (topological and analytical) concepts, such as *fixed points* and *convex sets*, so that *fixed-point theory* (FPT) and *convexity theory* (CT) are the cornerstones of modern economic theory.

[11] This interpretation of the Lagrange multipliers generalizes to any constrained optimization problem, in that the Lagrange multiplier attached to each of the constraints measures, at the optimum, the increase in the objective that results from relaxing the corresponding constraint by a "small" (marginal) amount. In this sense λ_j attached to constraint j is the *shadow price* of the variable restricted under the constraint j. For example, in a production optimization problem over the inputs of capital K and labor L, the multiplier λ_K attached to

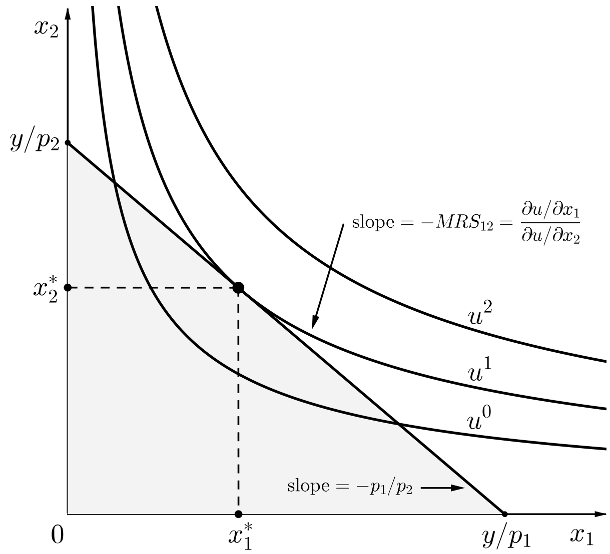


FIGURE 2. Utility maximization subject to the budget constraint.

Aside. The Lagrange condition

$$\nabla u = \lambda \nabla g$$

for the [UMP] can be restated using differentials and the wedge product as

$$du \wedge dg = 0,$$

which means that at the optimal point the two differentials (co-vectors) du and dg are linearly dependent (parallel). We compute,

$$du = \sum_i \frac{\partial u}{\partial q_i}\, dq_i, \qquad dg = \sum_i p_i\, dq_i$$

so that, using the anti-commutativity of the wedge product $a \wedge b = -b \wedge a$ (which implies that $a \wedge a = 0$),

$$\begin{aligned} du \wedge dg &= \left(\sum_i \frac{\partial u}{\partial q_i}\, dq_i\right) \wedge \left(\sum_i p_i\, dq_i\right) \\ &= \sum_{i<j} \left(\frac{\partial u}{\partial q_i}\, p_j - \frac{\partial u}{\partial q_j}\, p_i\right) dq_i \wedge dq_j. \end{aligned}$$

The condition $du \wedge dg = 0$ thus yields

$$\frac{\partial u}{\partial q_i} \Big/ \frac{\partial u}{\partial q_j} = p_i / p_j,$$

the capital constraint is interpreted as the shadow price of capital, and the multiplier λ_L attached to the labor constraint is interpreted as the shadow price of labor.

which is exactly the solution we obtained from the Lagrange condition. For more on this alternative derivation of the optimal solution, see Frank Zizza (1998) – "Differential Forms for Constrained Max-Min Problems: Eliminating Lagrange Multipliers", *The College Mathematics Journal*, vol. 29, p. 387 – 396. ∎

2.1.1. The Marshallian Demand Functions

The following theorem provides the properties of the Marshallian demand functions.

Theorem 2. *Suppose that $u(\cdot)$ is a strictly quasiconcave twice continuously differentiable utility function representing a complete, transitive, strictly convex preference relation $\succeq$ defined on the consumption set $Q = \mathbb{R}^n_+$. Then the Marshallian demand functions $\boldsymbol{g}(\boldsymbol{p}, m)$ exist and are:*

(i) Homogeneous of degree 0 in $(\boldsymbol{p}, m)$, i.e., for any $\boldsymbol{p}, m$ and scalar $\alpha > 0$, we have $\boldsymbol{g}(\alpha\boldsymbol{p}, \alpha m) = \boldsymbol{g}(\boldsymbol{p}, m)$.

(ii) Satisfy Walra's Law (also called the Adding-Up Condition): $\boldsymbol{p}^\mathsf{T}\boldsymbol{g}(\boldsymbol{p}, m) = m$.

(iii) Continuously differentiable in $\boldsymbol{p}$ and m.

Proof: (i) The homogeneity of degree 0 follows easily from the invariance of the budget constraint (feasibility set) to a proportional increase in prices and income. Since the following two sets are equal

$$\{\boldsymbol{q} \in \mathbb{R}^n_+ : \alpha\boldsymbol{p}^\mathsf{T}\boldsymbol{q} = \alpha m\} = \{\boldsymbol{q} \in \mathbb{R}^n_+ : \boldsymbol{p}^\mathsf{T}\boldsymbol{q} = m\}$$

the two demand functions $\boldsymbol{g}(\boldsymbol{p}, m)$ and $\boldsymbol{g}(\alpha\boldsymbol{p}, \alpha m)$, $\alpha > 0$, solve the same maximization problem and are thus equal.

(ii) Follows directly from the nonsatiation property of the preferences and the strict quasiconcavity of the utility function $u(\cdot)$. Let $\boldsymbol{q}^* = \boldsymbol{g}(\boldsymbol{p}, m)$ denote the optimal consumption. If $\boldsymbol{p}^\mathsf{T}\boldsymbol{q}^* < m$ there exists another bundle $\boldsymbol{q}'$ sufficiently close to $\boldsymbol{q}^*$ with $\boldsymbol{p}^\mathsf{T}\boldsymbol{q}' < m$ and $u(\boldsymbol{q}') > u(\boldsymbol{q}^*)$, which contradicts the optimality of $\boldsymbol{q}^*$.

(iii) Existence and continuous differentiability follow from our assumptions regarding $u(\boldsymbol{q})$ and the Implicit Function Theorem (see the Appendix). ∎

The fact that the functions $\boldsymbol{g}(\boldsymbol{p}, m) : \mathbb{R}^{n+1}_+ \to \mathbb{R}^n_+$ are homogeneous of degree 0 in $(\boldsymbol{p}, m)$ means that optimizing rational consumers do not suffer from *money illusion*[12], and thus their

[12] The existence or not of money illusion on the part of economic agents has been a very old preoccupation in economics. In the 1930's Irving Fisher expressed the opinion that people are more sensitive to changes in money income than to changes in prices, and he considered the consequent failure to adjust interest rates to rising prices as one cause of inflationary profits. J. M. Keynes' theory is based on similar observations, referring both to the demand for goods and to the supply of labor. See Marschak Jacob (1943) – Money Illusion and

demands depend on $\boldsymbol{p}$ and m only through the *normalized prices*

$$\mathfrak{p} := \boldsymbol{p}/m = (p_1/m, p_2/m, ..., p_n/m)^\mathsf{T}.$$

Therefore, there exist functions $\mathfrak{g}(\mathfrak{p}) : \mathbb{R}^n_+ \to \mathbb{R}^n_+$ such that

$$\mathfrak{g}(\mathfrak{p}) := \boldsymbol{g}(\boldsymbol{p}/m, 1). \tag{2.23}$$

These functions are called *normalized Marshallian demands*, and are the solutions to the optimization problem

$$\max_{\boldsymbol{q}\in\mathbb{R}^n_+} \; u(\boldsymbol{q}), \quad \text{subject to} \quad \mathfrak{p}^\mathsf{T}\boldsymbol{q} \leq 1. \tag{2.24}$$

While the functions $\boldsymbol{g}(\boldsymbol{p}, m_j)$ describe demand in a world where everyone faces the same prices $\boldsymbol{p}$ but has a different income m_j, say, the functions $\mathfrak{g}(\mathfrak{p}_j)$ describe demand in world where everyone faces different normalized prices $\mathfrak{p}_j = \boldsymbol{p}/m_j$. In this world, rich people have the 'right' to purchase goods cheaply, while poor people must purchase the same goods dearly, so as everyone's total expenditure is less that or equal to 1 monetary unit.

In applications, we almost always use the regular Marshallian demands $\boldsymbol{g}(\boldsymbol{p}, m)$ and talk of the effect of changes in $\boldsymbol{p}$ on the quantities demanded as *price effects*, and the effect of changes in m on the quantities demanded as *income effects*. Since price changes are not compensated by a proportional change (in the other direction) in income, the (regular) Marshallian demands $\boldsymbol{g}(\boldsymbol{p}, m)$ are also called *uncompensated demands*. As it turns out, normalizing by income is not quite enough to guarantee that a change in the price of a good is exactly compensated by a corresponding opposite change in income, since this compensation should also depend on the utility function and the satisfaction the consumer derives from each good. A price increase in a good that yields a lot of utility is not the same as a price increase in a good that matters less, and compensation should be analogous to the decrease in utility each price increase results in. In the next section we will introduced the compensated demands.

2.1.2. The Indirect Utility Function

Substituting the Marshallian demands $\boldsymbol{g}(\boldsymbol{p}, m)$ back into the utility function $u(\cdot)$ we obtain the *indirect utility function*

$$v(\boldsymbol{p}, m) := u(\boldsymbol{g}(\boldsymbol{p}, m)), \tag{2.25}$$

which expresses the maximum achievable utility under *price-income state* $v(\boldsymbol{p}, m)$. Equivalently we may define

$$v(\mathfrak{p}) := u(\mathfrak{g}(\mathfrak{p})), \tag{2.26}$$

Demand Analysis, Review of Economics and Statistics Volume 25 issue 1 1943, for an early demonstration of the existence of non-homogeneous demand functions in the market for meat [p. 46].

to be the *same* function as above since $v(\boldsymbol{p}, m)$ can only depend $\boldsymbol{p}$ and m only through the normalized prices $\mathfrak{p} = \boldsymbol{p}/m$.

The following theorem gives the main properties of the indirect utility function.

Theorem 3. (Properties of the Indirect Utility Function) *Suppose that $u(\cdot)$ is a twice continuously differentiable utility function representing a locally nonsatiated preference relation $\succeq$ defined on the consumption set $Q = \mathbb{R}^n_+$. Then the indirect utility function*

$$v(\boldsymbol{p}, m) = \max_{\boldsymbol{q} \in \mathbb{R}^n_+} \{u(\boldsymbol{q}) \text{ s.t. } \boldsymbol{p}^\mathsf{T} \boldsymbol{q} = m\}$$

for given $\boldsymbol{p} \gg 0$ and $m > 0$ is:

(i) Homogeneous of degree 0 in $(\boldsymbol{p}, m)$, i.e., for any $\boldsymbol{p} \gg 0, m > 0$ and scalar $\alpha > 0$, we have $v(\alpha \boldsymbol{p}, \alpha m) = v(\boldsymbol{p}, m)$.

(ii) Strictly increasing in m and decreasing (but not necessarily strictly) in $\boldsymbol{p}$.

(iii) Quasiconvex in in $(\boldsymbol{p}, m)$, i.e., the set $\{(\boldsymbol{p}, m) \in \mathbb{R}_{++} \times \mathbb{R}_+ : v(\boldsymbol{p}, m) \leq v_0\}$ is convex for any $v_0 \in \mathbb{R}$, or equivalently, for every $\alpha \in (0,1)$ and any distinct price-income states $(\boldsymbol{p}^0, m^0)$ and $(\boldsymbol{p}^1, m^1) \in \mathbb{R}_{++} \times \mathbb{R}_+$,

$$v\left(\alpha \boldsymbol{p}^0 + (1-\alpha)\boldsymbol{p}^1, \alpha m^0 + (1-\alpha) m^1\right) \leq \max\left\{v(\boldsymbol{p}^0, m^0), v(\boldsymbol{p}^1, m^1)\right\}.$$

(iv) Continuously differentiable in $\boldsymbol{p}$ and m.

Proof: (iv) Follows from the continuity and differentiability of $u(\cdot)$ and $\boldsymbol{g}(\boldsymbol{p}, m)$, since it is their composition. ■

Note that while the direct utility function $u(\boldsymbol{q})$ is (assumed!) quasi*concave* in $\boldsymbol{q}$, the indirect utility function $v(\boldsymbol{p}, m)$ is quasi*convex* in $(\boldsymbol{p}, m)$, and it is so even if $u(\cdot)$ is not quasiconcave (see Mas-Colell p.56, footnote 11). Also note that the indirect utility function depends on the utility representation chosen, in that if $u(\boldsymbol{q})$ corresponds to $v(\boldsymbol{p}, m)$ then $g(u(\boldsymbol{q}))$ corresponds to $g(v(\boldsymbol{p}, m))$, where $g(\cdot)$ is a strictly increasing function. For this reason $v(\boldsymbol{p}, m)$ and all strictly increasing transformations of it $g(v(\boldsymbol{p}, m))$, are considered the 'same' indirect utility function.

The direct utility function $u(\boldsymbol{q})$ and the indirect utility function $v(\mathfrak{p})$ are equivalent descriptions of the underlying preference ordering. Choice under a budget constraint can be analyzed either as the maximization of the direct utility function with given prices and income, or as the minimization of the indirect utility function with given quantities: the observable consequences are the same. This means that there is a *duality* relationship between the direct and the indirect utility functions, in the sense that maximization of $u(\boldsymbol{q})$ with respect to $\boldsymbol{q}$ given $\boldsymbol{p}$ and m, and minimization of $v(\mathfrak{p})$ with respect to $\mathfrak{p}$ given $\boldsymbol{q}$, leads to the same demand equations. We say the the direct utility $u(\boldsymbol{q})$ and the indirect utility $v(\mathfrak{p})$ are *dual functions* to each other.

While the direct utility function probably has greater intuitive appeal, the indirect utility function is not without its claims to interest also, for it is the foundation of "constant-utility" index numbers of the cost of living. If we try to determine what change in income is necessary to compensate for a given change in prices (to mention one of the problems to which such index numbers can be applied) we are in effect trying to keep the indirect utility function constant.

Example 2. The *constant-elasticity-of-substitution* (CES) utility function is defined as

$$u(\boldsymbol{q}) = \Big(\sum_{i=1}^n \alpha_i q_i^\rho\Big)^{1/\rho}, \qquad -\infty < \rho < 1,\ 0 < \alpha_i < 1,\ \sum_{i=1}^n \alpha_i = 1,\ q_i > 0,\ i = 1, ..., n. \tag{2.27}$$

As the name suggests, this specification has a constant $\sigma = 1/(1-\rho)$ *elasticity of substitution* for all levels of the q's, and it nests a number of interesting cases:

(i) $\rho \to 1, \sigma \to \infty \Rightarrow u(\boldsymbol{q}) = \sum_{i=1}^n \alpha_i q_i$ [linear utility, *perfect substitutes*].

(ii) $\rho \to 0, \sigma \to 1 \Rightarrow u(\boldsymbol{q}) = \sum_{i=1}^n \alpha_i \ln(q_i)$ [*Cobb-Douglas* utility].

(iii) $\rho \to -\infty, \sigma \to 0 \Rightarrow u(\boldsymbol{q}) = \min_{i=1,\dots,n} \{\alpha_i q_i\}$ [Leontief utility, *perfect complements*].

For income $y > 0$ and prices $p_i > 0, i = 1, ..., n$, the Lagrangian of the problem is

$$\mathscr{L} = \Big(\sum_{i=1}^n \alpha_i q_i^\rho\Big)^{1/\rho} + \lambda\Big(m - \sum_{i=1}^n p_i q_i\Big).$$

The first-order conditions for an interior solution are

$$\begin{aligned} \frac{\partial \mathscr{L}}{\partial q_i} &= 0 \Rightarrow \frac{1}{\rho}\Big(\sum_{i=1}^n \alpha_i q_i^\rho\Big)^{\frac{1}{\rho}-1} \alpha_i \rho q_i^{\rho-1} = \lambda p_i, \qquad i = 1, ..., n, \\ \frac{\partial \mathscr{L}}{\partial \lambda} &= 0 \Rightarrow m = \sum_{i=1}^n p_i q_i. \end{aligned}$$

Taking ratios to eliminate λ, we obtain

$$\frac{p_i}{p_j} = \frac{\alpha_i q_i^{\rho-1}}{\alpha_j q_j^{\rho-1}}, \qquad \text{or} \qquad q_j = \Big(\frac{\alpha_j p_i}{\alpha_i p_j}\Big)^\sigma q_i, \qquad i, j = 1, ..., n. \tag{2.28}$$

Substituting the last expression into the budget constraint and gathering terms we obtain

$$m = \sum_{j=1}^n p_j q_j = \sum_{j=1}^n p_j \Big(\frac{\alpha_j p_i}{\alpha_i p_j}\Big)^\sigma q_i = q_i \Big(\frac{p_i}{\alpha_i}\Big)^\sigma \sum_{j=1}^n \alpha_j^\sigma p_j^{1-\sigma},$$

or

$$\frac{m}{q_i} = \left(\frac{p_i}{\alpha_i}\right)^{\sigma} I(\boldsymbol{p}), \qquad i = 1, ..., n$$

where,

$$I(\boldsymbol{p}) = \sum_{i=1}^{n} \alpha_i^{\sigma} p_i^{1-\sigma}.$$

$I(\boldsymbol{p})$ is a *utility-weighted price index*. Such indices play an important role in measuring welfare effects of price changes and will be discussed later. Solving for q_i, we obtain the *CES Marshallian demand system*

$$q_i^* = g_i(\boldsymbol{p}, m) = \frac{m}{I(\boldsymbol{p})}\left(\frac{\alpha_i}{p_i}\right)^{\sigma}, \qquad i = 1, ..., n. \tag{2.29}$$

Plugging these demands back into the utility function, we derive the corresponding indirect utility function

$$v(\boldsymbol{p}, m) = m\, I(\boldsymbol{p})^{1/(1-\sigma)}. \tag{2.30}$$

We see that the indirect utility function is a function of real income m and the utility-weighted price index $I(\boldsymbol{p})^{1/(1-\sigma)}$, that is, it is equal to the *utility-weighted real income* given by $m/I(\boldsymbol{p})^{1/(\sigma-1)}$. ■

2.2. The Transposed (aka Dual) Problem – Hicksian Demands

An alternative approach, which yields equivalent results, is to formulate the *transposed problem* [13]

$$\text{[EMP]:} \qquad \min_{\boldsymbol{q}\in\mathbb{R}^n} \; e(\boldsymbol{q}; \boldsymbol{p}) = \boldsymbol{p}^{\mathsf{T}}\boldsymbol{q} \quad \text{subject to} \quad u(\boldsymbol{q}) \geq u, \tag{2.31}$$

called the *Expenditure Minimization Problem* [EMP], where, $e(\boldsymbol{q}; \boldsymbol{p})$ is the total *expenditure* or *cost* incurred by the consumer as a function of $\boldsymbol{q}$, taking $\boldsymbol{p}$ as given. For the sake of definiteness, we could think of u as the level of utility achieved by solving the primary [UMP] problem in (2.3).

[13] Quite often what is here called a *transposed* problem has been said by economists to be a problem *dual* to the original problem; see, e.g. Debreu (1951, p. 279), Arrow and Debreu (1954, pp. 285-286), Baumol (1977, p. 355), Deaton and Muellbauer (1980, pp. 37ff), Layard and Walters (1978, p. 143), and Theil (1980, p. 21). But to say this is to be inconsistent with normal mathematical usage, where a dual problem is one whose solutions are located in a vector space that is "in duality with" –and so in principle different from– the vector space of the solutions to the primal problem, whereas a transposed problem has its solution in the same space as the original.

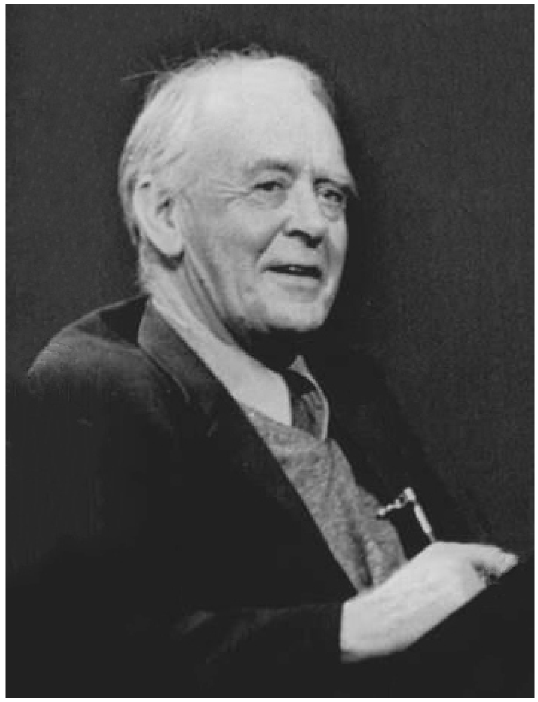


FIGURE 3. John Hicks (1904–1989)

Nonsatiation means that the utility constraint will be satisfied with equality, and the Lagrangian of the so-called "dual" problem in (2.31) is given by

$$\mathscr{M}(\boldsymbol{q}, \mu; \boldsymbol{p}, m) = \boldsymbol{p}^{\mathsf{T}}\boldsymbol{q} + \mu\,(u - u(\boldsymbol{q})), \tag{2.32}$$

where μ is a scalar Lagrange multiplier. Differentiation with respect to the elements of $\boldsymbol{q}$ and μ yields the $(n+1)$ system of simultaneous equations

$$\begin{aligned} p_i - \mu^* \frac{\partial u(\boldsymbol{q}^*)}{\partial q_i} &= 0, \qquad i = 1, ..., n, \\ u - u(\boldsymbol{q}) &= 0. \end{aligned} \tag{2.33}$$

These are exactly the same first order conditions (dividing the ith and the jth equation we obtain that the ratio of prices equals the ratio of marginal utilities) that we obtained from the [UMP] except that now the restriction is on u and not on m. Consequently, the second order conditions are also the same and thus the two problems are equivalent. By this we mean that if q^* solves the [UMP] it also solves the [EMP], or put another way, the solutions to the two problems are equal.

The [EMP] problem in (2.31) has solution

$$\boldsymbol{q}^* = \boldsymbol{h}(\boldsymbol{p}, u) = (h_1(\boldsymbol{p}, u), ..., h_n(\boldsymbol{p}, u))^{\mathsf{T}}, \tag{2.34}$$

$$\mu^* = 1/\lambda^*. \tag{2.35}$$

Here $\boldsymbol{h}(\boldsymbol{p}, u)$ is a $n \times 1$ vector-valued function called the *compensated* or *Hicksian demands*[14] and they represent optimal consumption behavior as a function of prices $\boldsymbol{p}$ and the utility level u, in contrast to the Marshallian formulation which is in terms of prices $\boldsymbol{p}$ and the (*observable*!) level of income M. Also, μ^* is the *marginal cost of utility* and is equal to the reciprocal of the marginal utility of income λ^*.

Since the tangency conditions of [UMP] and [EMP] are identical, we have

$$\boldsymbol{q}^* = \boldsymbol{g}(\boldsymbol{p}, m) = \boldsymbol{h}(\boldsymbol{p}, u),$$

i.e., the solution of both the maximization problem and minimization problem produce identical demands $\boldsymbol{q}^*$. However, the solutions are functions of different variables, so the comparative-statics exercises will generally produce different results.

Aside. The following theorem is proven in Panik, M.J. *Classical Optimization: Foundations and Extensions*, New York: North-Company, 1976 (pp. 207). See also, Caputo M.R. (2000), "Lagrangian transposition identities and reciprocal pairs of constrained optimization problems", *Economics Letters*, 66, 265–273, and the references therein. See also Varian H.R. (1992) *Microeconomics Analysis*, p. 113.

Theorem (Lagrangian Transposition Principle): The solution $\boldsymbol{x}^*(\boldsymbol{\alpha}; z)$ to the problem

$$\max_{\boldsymbol{x}} \{f(\boldsymbol{x}; \boldsymbol{\alpha}) \text{ s.t. } g(\boldsymbol{x}; \boldsymbol{\alpha}) \leq z\}$$

with Lagrangian

$$\mathscr{L}(\boldsymbol{x}, \lambda) = f(\boldsymbol{x}; \boldsymbol{\alpha}) + \lambda(z - g(\boldsymbol{x}; \boldsymbol{\alpha}))$$

for fixed z, is identical to the solution $\hat{\boldsymbol{x}}(\boldsymbol{\alpha}; y)$ of the "transposed problem"

$$\min_{\boldsymbol{x}} \{g(\boldsymbol{x}; \boldsymbol{\alpha}) \text{ s.t. } f(\boldsymbol{x}; \boldsymbol{\alpha}) \geq y\}$$

with Lagrangian

$$\mathscr{M}(\boldsymbol{x}, \mu) = g(\boldsymbol{x}; \boldsymbol{\alpha}) + \mu(y - f(\boldsymbol{x}; \boldsymbol{\alpha}))$$

for fixed $y = f(\boldsymbol{x}^*; \boldsymbol{\alpha})$, where $\boldsymbol{x}^*$ solves the first problem.

[14] As the name suggests the dual consumer problem and the corresponding compensated demand functions were introduced by Sir John Richard Hicks (1939).

Assume that $y = f(\boldsymbol{x};\boldsymbol{\alpha})$ is quasiconcave and $z = g(\boldsymbol{x};\boldsymbol{\alpha})$ is quasiconvex. ■

Again substituting the solution back into the objective function, we get the minimum-value function

$$c(\boldsymbol{p}, u) := \boldsymbol{p}^\top \boldsymbol{h}(\boldsymbol{p}, u),$$

which is usually called the *indirect cost* or *expenditure function*: it gives the minimum expenditure needed to obtain utility level u under prices $\boldsymbol{p}$. The indirect cost function $c(\boldsymbol{p}, u)$ is the *dual function* of the direct cost function $e(\boldsymbol{q};\boldsymbol{p}) = \boldsymbol{p}^\top \boldsymbol{q}$. While in the direct cost function $e(\boldsymbol{q};\boldsymbol{p})$, $\boldsymbol{q}$ is variable and $\boldsymbol{p}$ is fixed, in the dual function $c(\boldsymbol{p}, u)$ there is no $\boldsymbol{q}$ anymore and both $\boldsymbol{p}$ and u are variable.

Since the solutions $\boldsymbol{q}^*$ in the primal and dual problems are determined by the tangency point of the same utility curve and budget constraint line, it follows that the minimized expenditure in the dual problem is equal to the budget constraint in the primal, i.e.,

$$c(\boldsymbol{p}, u) = \boldsymbol{p}^\top \boldsymbol{h}(\boldsymbol{p}, u) = \boldsymbol{p}^\top \boldsymbol{g}(\boldsymbol{p}, m) = \boldsymbol{p}^\top \boldsymbol{q}^* = m.$$

The expenditure function was introduced into the literature by Lionel McKenzie (1957), and plays an important role in Welfare Economics.

Example 3. The Lagrangian of the dual problem for the CES utility function in (2.27) is given by

$$\mathscr{M} = \sum_{i=1}^{n} p_i q_i + \mu \left[u - \Big(\sum_{i=1}^{n} \alpha_i q_i^\rho \Big)^{1/\rho} \right].$$

The first-order conditions for an interior solution yield again eq. (2.28). Substituting back into the constraint and solving for each of the q's in turn, we obtain the *CES Hicksian demand functions*

$$q_i^* = h_i(\boldsymbol{p}, u) = \frac{u}{I(\boldsymbol{p})^{1/\rho}} \Big(\frac{\alpha_i}{p_i}\Big)^{\sigma}, \qquad i = 1, ..., n. \tag{2.36}$$

Plugging the solutions into the individual's budget, we get the expenditure function

$$c(\boldsymbol{p}, u) = u\, I(\boldsymbol{p})^{1-\sigma}. \tag{2.37}$$

Question: Could you have guessed the form of the expenditure function $e(u, \boldsymbol{p})$ in the CES case from the derivations in Example 1?

We see that the expenditure function $c(\boldsymbol{p}, u)$ here is homogeneous of degree 1 in u (actually, this is true for any utility function not just the CES, see Theorem 1 below), which permits us to form an exact *price index* corresponding to the *cost of a unit of utility* given by

$$\frac{\partial c(\boldsymbol{p}, u)}{\partial u} = I(\boldsymbol{p})^{1-\sigma} = \Big(\sum_{i=1}^{n} \alpha_i^{\sigma} p_i^{1-\sigma}\Big)^{1-\sigma}.$$

■

2.3. A Third Variant – Frisch Demands

Yet a third variant of the foregoing is the so-called *Frisch demands*. Recall that solving the primal problem yields the first order conditions

$$\frac{\partial u(\boldsymbol{q}^*)}{\partial q_i} = \lambda^* p_i, \qquad i = 1, ..., n \tag{2.38}$$

where λ^* is a Lagrange multiplier on the budget constraint that may be interpreted as the *marginal utility of income*. When utility is *additive separable*, i.e.,

$$u(\boldsymbol{q}) = \sum_{i=1}^{n} u_i(q_i),$$

eq. (2.38) becomes

$$u_i'(q_i^*) = p_i/r^*, \qquad i = 1, ..., n,$$

where $r^* = 1/\lambda^*$ can be interpreted as the *marginal cost of utility at current prices*. Inverting this (recall that $u_i(q_i)$ must be monotonic by nonsatiation so the inverse is well-defined) we obtain the *Frisch demands*

$$q_i^* = f_i(p_i/r), \qquad i = 1, ..., n,$$

where $f_i(\cdot) = (u_i')^{-1}(\cdot)$. It follows that demand for the ith good depends only on r and the ith price. For general utility functions this is not true, but under additive separability (which holds in many intertemporal models, for example) it is, and as it turns out it is quite useful.

An important example of the foregoing theory is the *linear expenditure system*, developed by Stone, Gorman, Samuelson and others. The utility function takes the additive form

$$u(\boldsymbol{q}) = \sum_{i=1}^{n} u_i(q_i) := \sum_{i=1}^{n} \beta_i \log(q_i - \gamma_i), \quad \text{with } q_i > \gamma_i > 0,\ \beta_i > 0,\ \sum_{j=1}^{n} \beta_j = 1.$$

This is the so-called *Stone-Geary utility function*, and it was first used by Klein and Rubin (1948). To obtain it, start from the simple product utility $u(\boldsymbol{q}) = \prod_i q_i$ and replace q_i by $(q_i - \gamma_i)$. Then raise each term to the power β_i to obtain the utility $u(\boldsymbol{q}) = \prod_i (q_i - \gamma_i)^{\beta_i}$. Finally, take logs to get another member of this class of utility functions, and set $\sum_i \beta_i = 1$ to normalize it. Note that the only difference between this and the Cobb-Douglas utility function discussed above is the presence of the γ_i parameters that are interpreted below as 'committed' or 'subsistence' quantities of good i.

The Lagrangian of the primary problem is given by

$$\mathscr{L} = \sum_{i=1}^{n} u_i(q_i) + \lambda\Big(m - \sum_{i=1}^{n} p_i q_i\Big),$$

which yields the first order conditions

$$u_i'(q_i^*) = \frac{\beta_i}{q_i^* - \gamma_i} = \lambda^* p_i, \quad i = 1, ..., n.$$

Using the 'β constraint' $\sum_j \beta_j = 1$ and the budget constraint, we get

$$1 = \sum_{j=1}^{n} \beta_j = \lambda^* \sum_{j=1}^{n} p_j(q_j^* - \gamma_j),$$

so

$$\lambda^* = \frac{1}{m - \sum_{j=1}^{n} p_j \gamma_j}.$$

Substituting this expression for λ^* back and solving for q_i^* we obtain the demand system

$$q_i^* =: f_i(\boldsymbol{p}, m) = \gamma_i + \frac{\beta_i}{p_i}\Big(m - \sum_{j=1}^{n} p_j \gamma_j\Big), \qquad i = 1, ..., n, \tag{2.39}$$

$$\text{such that} \qquad q_i^* > \gamma_i > 0, \beta_i > 0, \text{ and } \sum_{i=1}^{n} \beta_i = 1.$$

This is the famous *linear expenditure demand system* (LES). Each quantity demanded q_i is a function of income m, own-price p_i, and *all* the prices p_j of the other goods. The parameters γ_i and β_i in this formulation have a convenient economic interpretation:

> γ_i is the *committed quantity* of consumption of good i purchased regardless of the currently prevailing prices that may be interpreted as 'subsistence' quantities, and
>
> β_i is the *marginal budget share* of good i, i.e., the rate of change in the share of good i as money income m changes.

Thus, we may view the consumer as first deciding to purchase γ_i of q_i, then computing his remaining, "supernumerary" income $(m - \sum_i p_i \gamma_i)$, and allocating this income according to the β_i's. Since its first use by Stone (1954), LES has probably been the most popular demand system. This system is convenient for its simplicity. However, it is also very restrictive. For instance, it imposes the restriction that all the goods are complements in consumption. This is not realistic in most applications, particularly when the goods under study are varieties of a differentiated product.

Substituting $\boldsymbol{f}(\boldsymbol{p}, m)$ back into $u(\boldsymbol{q})$, we obtain the *LES indirect utility function*

$$v(\boldsymbol{p}, m) = \sum_{i=1}^{n} \beta_i \log \left[\frac{\beta_i}{p_i} \Big(m - \sum_{j=1}^{n} p_i \gamma_i \Big) \right].$$

Exponentiating we obtain the equivalent form

$$\nu(\boldsymbol{p}, m) = e^{v(\boldsymbol{p}, m)} = \Big(m - \sum_{i=1}^{n} p_i \gamma_i \Big) \prod_{j=1}^{n} (\beta_i / p_i)^{\beta_i}. \tag{2.40}$$

Setting the expenditure function equal to a fixed level of utility u and solving for m, we obtain the *LES expenditure function*

$$c(\boldsymbol{p}, u) = u \prod_{i=1}^{n} (p_i / \beta_i)^{\beta_i} + \sum_{i=1}^{n} p_i \gamma_i. \tag{2.41}$$

This provides a very convenient way to compute true *cost of living indices* or exact *compensating variations* required to change prices from $\boldsymbol{p}^0$ to $\boldsymbol{p}^1$. We will return to this later when we discuss welfare measures.

As it turns out, the LES model cannot describe a system which contains inferior or complementary commodities or groups of commodities.

Multiplying (2.39) through by p_i, and letting $c_i = p_i q_i$ be the expenditure on good i, we obtain

$$\begin{aligned} c_i &= p_i \gamma_i + \beta_i \left(m - \sum_{j=1}^{n} p_j \gamma_j \right) && (2.42) \\ &= \beta_i \, m + \left(p_i \gamma_i - \beta_i \sum_{j=1}^{n} p_j \gamma_j \right), \qquad i = 1, ..., n. \end{aligned}$$

Letting $\boldsymbol{\Gamma} = \mathrm{diag}\{\boldsymbol{\gamma}\}$ be an $n \times n$ diagonal matrix and $\boldsymbol{\iota} = (1, 1, ..., 1)^\mathsf{T}$ be an n-vector of ones, we can write

$$\boldsymbol{c} = \boldsymbol{\beta} \, m + (\boldsymbol{I} - \boldsymbol{\beta} \boldsymbol{\iota}^\mathsf{T}) \boldsymbol{\Gamma} \, \boldsymbol{p}, \tag{2.43}$$

where, $\boldsymbol{c} = (p_1 q_1, p_2 q_2, ..., p_n q_n)^\mathsf{T}$ and $\boldsymbol{p} = (p_1, ..., p_n)^\mathsf{T}$ are the n-vector of expenditures and prices, respectively. Finally, letting $\boldsymbol{D} = (\boldsymbol{I} - \boldsymbol{\beta} \boldsymbol{\iota}^\mathsf{T}) \boldsymbol{\Gamma}$ be an $n \times n$ matrix, we can write

$$\boldsymbol{c} = \boldsymbol{\beta} \, m + \boldsymbol{D} \, \boldsymbol{p}. \tag{2.44}$$

We see that the LES model expresses *expenditures* as a *linear* function of *prices* and *income*, which justifies its name. However, although linear in the variables m and $\boldsymbol{p}$, the model is *nonlinear in the structural parameters* $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$, and thus to estimate it we will need to use nonlinear econometric methods.

To be more precise, the model is *linear in the extended-form parameters* $\boldsymbol{\beta}$ and $\boldsymbol{D}$, but this would entail estimation of the n parameters in $\boldsymbol{\beta}$ and the n^2 parameters in $\boldsymbol{D}$, which gives a total of $n(n+1)$ parameters, instead of the only $2n$ structural parameters in $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$. In fact, there are only $2n-1$ structural parameters to be estimated, since the summing-up constraint $m = \boldsymbol{\iota}^\top \boldsymbol{c}$ implies that one (any) of the equations can be deleted in estimation, and its parameters uniquely determined once we obtain estimates of the other equations.

2.3.1. Estimation of the LES model

Assume now that we have a sample of household consumption data $\{\boldsymbol{q}_t, \boldsymbol{p}_t, m_t\}$, for households $t = 1, ..., T$, and define the vector of budget shares $\boldsymbol{w}_t$ for household t as

$$\boldsymbol{w}_t = \begin{bmatrix} p_{1t}q_{1t}/m_t \\ p_{2t}q_{2t}/m_t \\ \vdots \\ p_{nt}q_{nt}/m_t \end{bmatrix}_{n\times 1}, \qquad t = 1, ..., T.$$

We can write the regression model for $\boldsymbol{w}_t$ as

$$\boldsymbol{w}_t = \boldsymbol{\beta} + \boldsymbol{D}\,\boldsymbol{p}_t/m_t + \boldsymbol{u}_t, \qquad t = 1, ..., T, \tag{2.45}$$

such that

$$\begin{aligned} \boldsymbol{D} &= (\boldsymbol{I} - \boldsymbol{\beta}\boldsymbol{\iota}^\top)\boldsymbol{\Gamma}; \\ \boldsymbol{\Gamma} &= \mathrm{diag}\{\gamma\}; \\ \boldsymbol{\gamma} &> \boldsymbol{0}; \boldsymbol{\beta} > \boldsymbol{0}; \\ \boldsymbol{\iota}^\top\boldsymbol{\beta} &= 1; \\ \boldsymbol{\iota}^\top\boldsymbol{D} &= \boldsymbol{0}^\top, \end{aligned}$$

where $\boldsymbol{u}_t = (u_{t1}, u_{t2}, ..., u_{tn})^\top$ is a $n \times 1$ error vector. Because of the summing-up condition $m_t = \boldsymbol{\iota}^\top \boldsymbol{c}_t$ (which implies the $\boldsymbol{\iota}^\top\boldsymbol{D} = \boldsymbol{0}^\top$ condition), the variance-covariance matrix of $\boldsymbol{u}_t$, given by $E(\boldsymbol{u}_t\boldsymbol{u}_t^\top)$, is singular. To achieve identification of the model, we will drop the equation of one of the n goods and consider the remaining $(n-1)$-dimensional system of demands. Note that we may drop *anyone* of the equations without affecting the estimation results, so we will drop the nth good, and consider estimation of the remaining $(n-1)$ demand equations. **In**

what follows, all quantities that up to now had dimension n will, from here on and until the end of this section, have dimension $(n-1)$.

We will assume that

$$\boldsymbol{u}_t = \begin{bmatrix} u_{t1} \\ u_{t2} \\ \vdots \\ u_{t(n-1)} \end{bmatrix} \sim MVN_{(n-1)}\left(\mathbf{0}_{(n-1)\times 1}, \boldsymbol{\Sigma}_{(n-1)\times(n-1)}\right), \quad t = 1, ..., T, \tag{2.46}$$

where

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{1(n-1)} \\ \sigma_{21} & \sigma_{22} & \sigma_{2(n-1)} \\ \vdots & & \\ \sigma_{(n-1)1} & \sigma_{(n-1)2} & \sigma_{(n-1)(n-1)} \end{bmatrix}_{(n-1)\times(n-1)} . \tag{2.47}$$

Now let $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\gamma})$ be the parameters in the model and define

$$\boldsymbol{x}_t(\boldsymbol{\theta}) = \boldsymbol{\beta}\, m_t + \boldsymbol{D}\, \boldsymbol{p}_t, \tag{2.48}$$

to be the nonlinear mean function we wish to estimate, so that

$$\boldsymbol{c}_t = \boldsymbol{x}_t(\boldsymbol{\theta}) + \boldsymbol{u}_t, \qquad t = 1, ..., T. \tag{2.49}$$

The density of $\boldsymbol{u}_t$ is given by

$$p(\boldsymbol{u}_t|\boldsymbol{\Sigma}) = (2\pi)^{-(n-1)/2}\, |\boldsymbol{\Sigma}|^{-1/2}\, \exp\{-\tfrac{1}{2}\boldsymbol{u}_t^{\mathsf{T}}\, \boldsymbol{\Sigma}^{-1}\, \boldsymbol{u}_t\}.$$

Therefore, the density of $\boldsymbol{c}_t$ is

$$p(\boldsymbol{c}_t|\boldsymbol{\theta}, \boldsymbol{\Sigma}) = (2\pi)^{-(n-1)/2}\, |\boldsymbol{\Sigma}|^{-1/2}\, \exp\left\{-\tfrac{1}{2}(\boldsymbol{c}_t - \boldsymbol{x}_t(\boldsymbol{\theta}))^{\mathsf{T}}\, \boldsymbol{\Sigma}^{-1}\, (\boldsymbol{c}_t - \boldsymbol{x}_t(\boldsymbol{\theta}))\right\}\, |\boldsymbol{J}|,$$

where the Jacobian of the transformation $|\boldsymbol{J}| = 1$. Hence, the log-likelihood function of the sample is

$$\ell(\boldsymbol{\theta}, \boldsymbol{\Sigma}) = -\frac{T(n-1)}{2}\log(2\pi) - \frac{T}{2}\log|\boldsymbol{\Sigma}| - \frac{1}{2}\sum_{t=1}^{T}(\boldsymbol{c}_t - \boldsymbol{x}_t(\boldsymbol{\theta}))^{\mathsf{T}}\, \boldsymbol{\Sigma}^{-1}\, (\boldsymbol{c}_t - \boldsymbol{x}_t(\boldsymbol{\theta})). \tag{2.50}$$

To maximize the likelihood in (2.50) we will follow a four-step procedure that is standard in dealing with likelihoods of normal (linear and nonlinear) regression models: First, we will obtain the ML estimator of the variance-covariance matrix of the errors $\boldsymbol{\Sigma}$ as a function of the mean function parameter vector $\boldsymbol{\theta}$. Second, we will use the ML estimator of $\boldsymbol{\Sigma}$ to concentrate the likelihood in terms of the parameter vector $\boldsymbol{\theta}$ only. Third, we will maximize the concentrated likelihood to obtain the ML estimator of $\boldsymbol{\theta}$. Fourth, we will obtain the Information Matrix of the model and use standard asymptotic theory to obtain the asymptotic distribution of the ML estimators along with consistent estimators for the quantities in this distribution. Because of the importance of this four-step procedure in applied work where normal errors are frequently assumed, we will describe it in detail. In many papers of the literature (for example Parks (1971)) certain steps of this procedure are omitted as “obvious” or “well-known”, assuming that the reader can provide the omitted steps for him- or herself.

1. Since $|\boldsymbol{\Sigma}| = |\boldsymbol{\Sigma}^{-1}|^{-1}$, (2.50) can be expressed purely as a function of the inverse matrix $\boldsymbol{\Sigma}^{-1}$ as

$$\ell(\boldsymbol{\theta}, \boldsymbol{\Sigma}) = -\frac{T(n-1)}{2}\log(2\pi) + \frac{T}{2}\log|\boldsymbol{\Sigma}^{-1}| - \frac{1}{2}\sum_{t=1}^{T}(\boldsymbol{c}_t - \boldsymbol{x}_t(\boldsymbol{\theta}))^{\mathsf{T}}\,\boldsymbol{\Sigma}^{-1}\,(\boldsymbol{c}_t - \boldsymbol{x}_t(\boldsymbol{\theta})). \quad (2.51)$$

By the invariance of the ML estimator to monotone transformations, the ML estimator of $\boldsymbol{\Sigma}$ is the inverse of the ML estimator of $\boldsymbol{\Sigma}^{-1}$, so we can compute the ML estimator of $\boldsymbol{\Sigma}$ by partially differentiating (2.51) with respect to the inverse $\boldsymbol{\Sigma}^{-1}$ which is simpler, and solve the resulting equation. We then have

$$\frac{\partial \ell}{\partial(\boldsymbol{\Sigma}^{-1})} = \frac{T}{2}\boldsymbol{\Sigma} - \frac{1}{2}\sum_{t=1}^{T}(\boldsymbol{c}_t - \boldsymbol{x}_t(\boldsymbol{\theta}))\,(\boldsymbol{c}_t - \boldsymbol{x}_t(\boldsymbol{\theta}))^{\mathsf{T}}. \quad (2.52)$$

Setting this equal to zero and solving for $\boldsymbol{\Sigma}$, we obtain the ML estimator of the variance-covariance of the errors as

$$\hat{\boldsymbol{\Sigma}}(\boldsymbol{\theta})_{mle} = \frac{1}{T}\sum_{t=1}^{T}(\boldsymbol{c}_t - \boldsymbol{x}_t(\boldsymbol{\theta}))\,(\boldsymbol{c}_t - \boldsymbol{x}_t(\boldsymbol{\theta}))^{\mathsf{T}}. \quad (2.53)$$

The ML estimator of $\boldsymbol{\Sigma}$ is exactly what we would expect it to be, namely the matrix of sums of squares and cross-products of the residuals, divided by the sample size (without correction for the degrees of freedom).

2. We now wish to substitute the ML estimator in (2.53) back into the log-likelihood function in (2.51). Observing that the trace of a scalar is just the scalar itself and that the trace of a matrix product is invariant to a cyclic permutation of the factors of the product, we obtain

$$\begin{aligned}(\boldsymbol{c}_t - \boldsymbol{x}_t(\boldsymbol{\theta}))^{\mathsf{T}}\,\boldsymbol{\Sigma}^{-1}\,(\boldsymbol{c}_t - \boldsymbol{x}_t(\boldsymbol{\theta})) &= \operatorname{Tr}\Big((\boldsymbol{c}_t - \boldsymbol{x}_t(\boldsymbol{\theta}))^{\mathsf{T}}\,\boldsymbol{\Sigma}^{-1}\,(\boldsymbol{c}_t - \boldsymbol{x}_t(\boldsymbol{\theta})\Big) \\ &= \operatorname{Tr}\Big(\boldsymbol{\Sigma}^{-1}(\boldsymbol{c}_t - \boldsymbol{x}_t(\boldsymbol{\theta}))\,(\boldsymbol{c}_t - \boldsymbol{x}_t(\boldsymbol{\theta})^{\mathsf{T}}\Big).\end{aligned}$$

Summing over t yields

$$\begin{aligned}\sum_{t=1}^{T}(\boldsymbol{c}_t-\boldsymbol{x}_t(\boldsymbol{\theta}))^{\mathsf{T}}\,\boldsymbol{\Sigma}^{-1}\,(\boldsymbol{c}_t-\boldsymbol{x}_t(\boldsymbol{\theta})) &= \sum_{t=1}^{T}\mathrm{Tr}\left((\boldsymbol{c}_t-\boldsymbol{x}_t(\boldsymbol{\theta}))^{\mathsf{T}}\,\boldsymbol{\Sigma}^{-1}\,(\boldsymbol{c}_t-\boldsymbol{x}_t(\boldsymbol{\theta})\right)\\ &= \mathrm{Tr}\left(\boldsymbol{\Sigma}^{-1}\sum_{t=1}^{T}(\boldsymbol{c}_t-\boldsymbol{x}_t(\boldsymbol{\theta}))\,(\boldsymbol{c}_t-\boldsymbol{x}_t(\boldsymbol{\theta})^{\mathsf{T}}\right)\\ &= \mathrm{Tr}\left(\boldsymbol{\Sigma}^{-1}T\boldsymbol{\Sigma}\right)=T(n-1).\end{aligned}$$

Thus the concentrated log-likelihood function that corresponds to (2.50) is

$$\begin{aligned}\ell^c(\boldsymbol{\theta}) &= -\frac{T(n-1)}{2}(\log(2\pi)+1)-\frac{T}{2}\log\left|\frac{1}{T}\sum_{t=1}^{T}(\boldsymbol{c}_t-\boldsymbol{x}_t(\boldsymbol{\theta}))\,(\boldsymbol{c}_t-\boldsymbol{x}_t(\boldsymbol{\theta})^{\mathsf{T}})\right|\\ &= -\frac{T(n-1)}{2}(\log(2\pi)+1)-\frac{T}{2}\log|\hat{\boldsymbol{\Sigma}}(\boldsymbol{\theta})|.\end{aligned}$$

It is customary to aggregate all constants that do not depend on the parameter $\boldsymbol{\theta}$ into a constant C and write the concentrated log-likelihood function as

$$\ell^c(\boldsymbol{\theta})=C-\frac{T}{2}\log|\hat{\boldsymbol{\Sigma}}(\boldsymbol{\theta})|. \tag{2.54}$$

3. Next we wish to obtain the ML estimator of the mean function parameter vector $\boldsymbol{\theta}=(\boldsymbol{\beta},\boldsymbol{\gamma})$. From (2.54) we see that to obtain $\hat{\boldsymbol{\theta}}_{mle}$ we will need to minimize the logarithm of the determinant of the contemporaneous covariance matrix $\hat{\boldsymbol{\Sigma}}(\boldsymbol{\theta})$. This can be done quite easily be using the rule for computing derivatives of logarithms of determinants given in the Appendix. This rule states that if $\boldsymbol{A}$ is a nonsingular square matrix, then the derivative of $\log|\boldsymbol{A}|$ with respect to the ij-th element of $\boldsymbol{A}$ is the ji-th element of $\boldsymbol{A}^{-1}$. By the chain rule, the derivative of $\log|\hat{\boldsymbol{\Sigma}}(\boldsymbol{\theta})|$ with respect to θ_h is

$$\begin{aligned}\frac{\partial\log|\hat{\boldsymbol{\Sigma}}(\boldsymbol{\theta})|}{\partial\theta_h} &= \sum_{i=1}^{n-1}\sum_{j=1}^{n-1}\frac{\partial\log|\hat{\boldsymbol{\Sigma}}(\boldsymbol{\theta})|}{\partial\sigma_{ij}}\,\frac{\partial\sigma_{ij}(\boldsymbol{\theta})}{\partial\theta_h}\\ &= \sum_{i=1}^{n-1}\sum_{j=1}^{n-1}\left(\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\right)_{ji}\,\frac{\partial\sigma_{ij}(\boldsymbol{\theta})}{\partial\theta_h}\\ &= \mathrm{Tr}\left(\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\,\frac{\partial\boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial\theta_h}\right).\end{aligned}$$

It is now easy to see that

$$\frac{\partial\boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial\theta_h}=-\frac{2}{T}\sum_{t=1}^{T}\boldsymbol{u}_t(\boldsymbol{\theta})\,\frac{\partial\boldsymbol{x}_t(\boldsymbol{\theta})}{\partial\theta_h},$$

from which the gradient of (2.54) can be seen to be

$$\nabla_{\boldsymbol{\theta}} \ell^c(\boldsymbol{\theta}) = \sum_{t=1}^{T} X_t^\top(\boldsymbol{\theta})\ \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\ \left(\boldsymbol{c}_t - \boldsymbol{x}_t(\boldsymbol{\theta})\right), \tag{2.55}$$

where $X_t(\boldsymbol{\theta})$ is the $(n-1) \times 2n-1$ matrix with typical element

$$X_{t,ji}(\boldsymbol{\theta}) \equiv \frac{\partial x_{ti}(\boldsymbol{\theta})}{\partial \theta_j}.$$

$$\boldsymbol{\theta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_{n-1} \\ \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_n \end{bmatrix}_{2n-1}, \qquad \boldsymbol{x}_t(\boldsymbol{\theta}) = \begin{bmatrix} p_{1t}\gamma_1 + \beta_1 \left(m_t - \sum_{j=1}^{n} p_{jt}\gamma_j\right) \\ p_{2t}\gamma_2 + \beta_2 \left(m_t - \sum_{j=1}^{n} p_{jt}\gamma_j\right) \\ \vdots \\ p_{n-1,t}\gamma_{n-1} + \beta_{n-1} \left(m_t - \sum_{j=1}^{n} p_{jt}\gamma_j\right) \end{bmatrix}_{(n-1)\times 1} \tag{2.56}$$

Given $\boldsymbol{\beta}$, we can write

$$\boldsymbol{c}_h - \boldsymbol{\beta}\, m_h = (\boldsymbol{I} - \boldsymbol{\beta}\boldsymbol{\iota}^\top)\, \boldsymbol{p}_h\, \boldsymbol{\gamma} + \boldsymbol{u}_h \tag{2.57}$$

or

$$\boldsymbol{y}_h = \boldsymbol{x}_h\, \boldsymbol{\gamma} + \boldsymbol{u}_h. \tag{2.58}$$

This a linear regression-through-the-origin model. Stacking the data and error terms across households, we can write

$$\boldsymbol{Y} = \boldsymbol{X}\, \boldsymbol{\Gamma} + \boldsymbol{U}, \tag{2.59}$$

where, $\boldsymbol{\Gamma} = \text{diag}\{\boldsymbol{\gamma}\}$ is an $(n-1) \times (n-1)$ matrix and

$$\boldsymbol{Y} = \begin{bmatrix} \boldsymbol{y}_1^\top \\ \boldsymbol{y}_2^\top \\ \vdots \\ \boldsymbol{y}_N^\top \end{bmatrix}_{N\times(n-1)}, \quad \boldsymbol{X} = \begin{bmatrix} \boldsymbol{x}_1^\top \\ \boldsymbol{x}_2^\top \\ \vdots \\ \boldsymbol{x}_N^\top \end{bmatrix}_{N\times(n-1)}, \quad \boldsymbol{U} = \begin{bmatrix} \boldsymbol{u}_1^\top \\ \boldsymbol{u}_2^\top \\ \vdots \\ \boldsymbol{u}_N^\top \end{bmatrix}_{N\times(n-1)}. \tag{2.60}$$

Assuming that the errors across households are iid, our assumption regarding $\boldsymbol{u}_h$ yields that

$$\boldsymbol{U} \sim MVN_{N(n-1)}\left(\boldsymbol{0}_{N(n-1)\times 1}, \boldsymbol{I}_N \otimes \boldsymbol{\Sigma}\right). \tag{2.61}$$

Under these assumptions, the equation in (2.59) is a (linear) Seemingly Unrelated Regression (SUR) model, and can thus be estimated by Generalized Least Squares (GLS) methods.

2.3.2. An Alternative Strategy for the Estimation of the LES model

Assume now that we have an iid sample of household consumption data for N households $\{\boldsymbol{c}_i, \boldsymbol{p}_i, \boldsymbol{m}\}$, where $\boldsymbol{c}_i$ is the N-vector of observations on the expenditure for the ith commodity, $\boldsymbol{p}_i$ is the N-vector of prices for the ith commodity, $\boldsymbol{m}$ is the N-vector of observations on total expenditure (income). Using (2.42) we can write the regression model for the expenditure on the ith commodity as

$$\boldsymbol{c}_i = \gamma_i \boldsymbol{p}_i + \beta_i \left(\boldsymbol{m} - \sum_{j=1}^{n} \gamma_j \boldsymbol{p}_j \right) + \boldsymbol{u}_i, \qquad i = 1, ..., n. \tag{2.62}$$

where $\boldsymbol{u}_i$ is a N-vector of unobserved random disturbances, and β_i and γ_i are are unknown scalar parameters to be estimated, subject to the constraint that $\sum_{i=1}^{n} \beta_i = 1$. This is a system of n equations (one for each commodity) with N observations for each equation.

The constraint on the β_i's and the fact that total expenditure $\boldsymbol{m}$ is the sum of the $\boldsymbol{c}_i$'s imply that $\sum_{i=1}^{n} \boldsymbol{u}_i = \boldsymbol{0}$. Thus, one of the equations (2.62) is completely redundant in the sense that using the information contained in any $n-1$ of the equations we can obtain the nth equation by an appropriate linear combination. We are dealing with an allocation problem, and it suffices to allocate $n-1$ of the commodities, the last is a residual. For this reason we shall consider the reduced system which consists of the equations (2.62), but with the nth equation deleted. This choice is arbitrary, but since the equations can be put in any order the choice does not matter.[15] **For the rest of this section, all quantities that up to now had dimension n will, from here on and until the end of this section, have dimension** $(n-1)$.

It will be convenient to express the reduced system in two different forms which alternatively focus on the set of $\boldsymbol{\beta}$ coefficients and on the $\boldsymbol{\gamma}$ coefficients.

1. A typical equation in the system may be written as follows to provide a regression equation for γ_i's given the β_i's

$$(\boldsymbol{c}_i - \beta_i \boldsymbol{m}) = \begin{bmatrix} -\beta_i \boldsymbol{p}_1, & \cdots, & (1-\beta_i)\boldsymbol{p}_i, & -\beta_i \boldsymbol{p}_{i+1}, & \cdots, & -\beta_i \boldsymbol{p}_n \end{bmatrix} \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_n \end{bmatrix} + \boldsymbol{u}_i. \tag{2.63}$$

[15] The linear restriction on the $\boldsymbol{u}_i$'s means that the covariance matrix for the full system is singular. Barten [1] shows for a different demand model that maximum likelihood estimation of the full system reduces to maximum likelihood for the reduced system. A similar formal argument can be made in the case of the linear Linear Expenditure System.

This equation can be written in a more compact form as

$$\boldsymbol{y}_i = \boldsymbol{x}_i \boldsymbol{\gamma} + \boldsymbol{u}_i, \qquad i = 1, ..., n. \tag{2.64}$$

The complete statistical system excluding the last equation can be written as

$$\begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \\ \vdots \\ \boldsymbol{y}_{n-1} \end{bmatrix}_{N(n-1)\times 1} = \begin{bmatrix} \boldsymbol{x}_1 & & \mathbf{0} & \\ & \boldsymbol{x}_2 & & \\ & \mathbf{0} & \ddots & \\ & & & \boldsymbol{x}_{n-1} \end{bmatrix}_{N(n-1)\times N(n-1)} \begin{bmatrix} \gamma \\ \gamma \\ \vdots \\ \gamma \end{bmatrix}_{N(n-1)\times 1} + \begin{bmatrix} \boldsymbol{u}_1 \\ \boldsymbol{u}_2 \\ \vdots \\ \boldsymbol{u}_{n-1} \end{bmatrix}_{N(n-1)\times 1} \tag{2.65}$$

or as

$$\boldsymbol{Y} = \boldsymbol{X}(\boldsymbol{\beta})\,\boldsymbol{\Gamma} + \boldsymbol{U}, \tag{2.66}$$

where $\boldsymbol{\Gamma} = \boldsymbol{\iota} \otimes \boldsymbol{\gamma}$, and we have written the matrix $\boldsymbol{X}$ as $\boldsymbol{X}(\boldsymbol{\beta})$ to remind us that the data in it are conditional on the β_i's.

2. Alternatively, a typical equation may be written as follows to provide an equation for the β_i's given the γ_i's

$$(\boldsymbol{c}_i - \gamma_i \boldsymbol{p}_i) = (\boldsymbol{m} - \sum_{j=1}^{n} \gamma_j \boldsymbol{p}_j)\,\beta_i + \boldsymbol{u}_i, \qquad i = 1, ..., n, \tag{2.67}$$

or more concisely,

$$\boldsymbol{w}_i = \boldsymbol{z}\beta_i + \boldsymbol{u}_i, \qquad i = 1, ..., n. \tag{2.68}$$

The complete statistical system excluding the last equation can be written as

$$\begin{bmatrix} \boldsymbol{w}_1 \\ \boldsymbol{w}_2 \\ \vdots \\ \boldsymbol{w}_{n-1} \end{bmatrix}_{N(n-1)\times 1} = \begin{bmatrix} \boldsymbol{z} & & & \\ & \boldsymbol{z} & \mathbf{0} & \\ & \mathbf{0} & \ddots & \\ & & & \boldsymbol{z} \end{bmatrix}_{N(n-1)\times (n-1)} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_{n-1} \end{bmatrix}_{(n-1)\times 1} + \begin{bmatrix} \boldsymbol{u}_1 \\ \boldsymbol{u}_2 \\ \vdots \\ \boldsymbol{u}_{n-1} \end{bmatrix}_{N(n-1)\times 1} \tag{2.69}$$

or as

$$\boldsymbol{W} = \boldsymbol{Z}(\boldsymbol{\gamma})\,\boldsymbol{\beta} + \boldsymbol{U}, \tag{2.70}$$

where we have written the matrix $\boldsymbol{Z}$ as $\boldsymbol{Z}(\boldsymbol{\gamma})$ to remind us that the data in it are conditional on the γ_i's.

The random disturbances $(u_{1h}, ..., u_{(n-1)h})$ for observation h are assumed to come from a multivariate normal distribution with mean zero and covariance matrix $\boldsymbol{\Sigma}$. Disturbances across different households are assumed to be uncorrelated. Thus, the random disturbance vector $\boldsymbol{U}$ has $E(\boldsymbol{U}) = \mathbf{0}$ and $E(\boldsymbol{U}\boldsymbol{U}^\top) = \boldsymbol{\Omega} = \boldsymbol{\Sigma} \otimes \boldsymbol{I}$, where $\boldsymbol{\Sigma}$ is the $(n-1) \times (n-1)$ covariance matrix of the multivariate normal distribution, where $\boldsymbol{I}$ is a $N \times N$ identity matrix and where $\otimes$ denotes the Kronecker product operation. The matrix $\boldsymbol{\Sigma}$ given by

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{1(n-1)} \\ \sigma_{21} & \sigma_{22} & \sigma_{2(n-1)} \\ \vdots & & \\ \sigma_{(n-1)1} & \sigma_{(n-1)2} & \sigma_{(n-1)(n-1)} \end{bmatrix}. \tag{2.71}$$

is assumed to be positive definite and symmetric.

Under these assumptions for $\boldsymbol{U}$, both the "$\boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{\gamma}$ given $\boldsymbol{\beta}$" model in (2.66) and the "$\boldsymbol{W}, \boldsymbol{Z}, \boldsymbol{\beta}$ given $\boldsymbol{\gamma}$" model in (2.70) are Nonlinear Seemingly Unrelated Regression (NSUR) models, and can thus be estimated by Nonlinear Generalized Least Squares (NGLS) methods.

2.4. Duality

In mathematics there are many notions of duality or conjugacy, among which is Lagrange duality, Frechnel conjugasy. There are dual problems, dual functions, dual vector spaces, and so on.

The following theorem describes the way in which the [UMP] and the [EMP] are dual problems to each other.

Theorem 4. (Duality I) *Assume that the utility function $u(\boldsymbol{q})$ is continuous and locally nonsatiated, and that prices are strictly positive, i.e., $\boldsymbol{p} \gg 0$.*

(i) If $\boldsymbol{q}^* = \boldsymbol{g}(\boldsymbol{p}, m)$ solves the [UMP] for some money income $m > 0$, then
- *$\boldsymbol{q}^*$ also solve the [EMP] for utility level $u = u(\boldsymbol{q}^*)$; and*
- *$c(\boldsymbol{p}, v(\boldsymbol{p}, m)) = m$.*

(ii) If $\boldsymbol{q}^* = \boldsymbol{h}(\boldsymbol{p}, u)$ solves the [EMP] for some utility level $u > 0$, then
- *$\boldsymbol{q}^*$ also solve the [UMP] for money income $m = \boldsymbol{p}^\top \boldsymbol{q}^*$; and*
- *$v(\boldsymbol{p}, c(\boldsymbol{p}, u)) = u$.*

Theorem 5. (Duality II) *Assume that the utility function $u(\boldsymbol{q})$ is continuous, and let $u > u(\boldsymbol{0})$ and $m > 0$. Then*

(i) $\boldsymbol{g}(\boldsymbol{p}, m) = \boldsymbol{h}(\boldsymbol{p}, v(\boldsymbol{p}, m))$; and
(ii) $\boldsymbol{h}(\boldsymbol{p}, u) = \boldsymbol{g}(\boldsymbol{p}, c(\boldsymbol{p}, u))$.

Theorem 6. (Properties of the Expenditure Function) *If $u(\boldsymbol{q})$ is continuous, strictly quasi-concave and locally non-satiated, then the associated expenditure function $c(\boldsymbol{p}, u)$ is*

(i) continuously differentiable in $\boldsymbol{p}$ and u,

(ii) homogeneous of degree 1 in $\boldsymbol{p}$,

(iii) concave in $\boldsymbol{p}$,

(iv) strictly increasing in u, and nondecreasing in p_j for any $j = 1, ..., n$,

(v) convex in u if $u(q)$ is concave,

(vi) Shephard's Lemma: has own-price partial derivatives which are the compensated (Hicksian) demand functions, i.e.,

$$h_i(\boldsymbol{p}, u) = \frac{\partial c(\boldsymbol{p}, u)}{\partial p_i}, \qquad i = 1, ..., n.$$

(vii) Logarithmic form of Shephard's Lemma: the budget share $w_i = p_i q_i^* / m$ of good i is given by

$$w_i(\boldsymbol{p}, u) = \frac{\partial \log c(\boldsymbol{p}, u)}{\partial \log p_i}, \qquad i = 1, ..., n.$$

Proof: We prove each property in turn.

(i) Follows directly from the continuity and differentiability of u.

(ii) *$H1^\circ$ in $\boldsymbol{p}$.* $c(\boldsymbol{p}, u) = \min\{\boldsymbol{p}^\intercal \boldsymbol{q} | u(\boldsymbol{q}) = u\}$ so $c(\theta \boldsymbol{p}, u) = \theta c(\boldsymbol{p}, u)$ for all $\theta > 0$.

(iii) *$\cap$ in $\boldsymbol{p}$.* Take any two price vectors $\boldsymbol{p}^0$ and $\boldsymbol{p}^1$ and set $\boldsymbol{p}^\theta = \theta \boldsymbol{p}^0 + (1-\theta)\boldsymbol{p}^1$, for some $\theta \in (0, 1)$. Let $\boldsymbol{q}^\theta$ be optimal for $\boldsymbol{p}^\theta$ and u, i.e., $\boldsymbol{x}^\theta = \boldsymbol{h}(\boldsymbol{p}^\theta, u)$. Then

$$c(\boldsymbol{p}^\theta, u) = {\boldsymbol{q}^\theta}^\intercal \boldsymbol{p}^\theta = \theta {\boldsymbol{q}^\theta}^\intercal \boldsymbol{p}^0 + (1-\theta){\boldsymbol{q}^\theta}^\intercal \boldsymbol{p}^1.$$

The result now follows from the observation that $\boldsymbol{q}^\theta$ is not optimal at $\boldsymbol{p}^0$ or $\boldsymbol{p}^1$, i.e.,

$$c(\boldsymbol{p}^j, u) \leq {\boldsymbol{q}^\theta}^\intercal \boldsymbol{p}^j, \quad j = 0, 1.$$

(iv) *$\nearrow$ in* u. Follows directly from nonsatiation since more u requires more $\boldsymbol{q}$ in at least one coordinate commodity.

(v) *Shephard's Lemma.* Since $c(\boldsymbol{p}, u)$ is concave in $\boldsymbol{p}$, it's partial derivatives with respect to $\boldsymbol{p}$ exist. Let

$$z(\boldsymbol{p}) = \boldsymbol{p}^\intercal \boldsymbol{q}^0 - c(\boldsymbol{p}, u)$$

where $\boldsymbol{q}^0 = \boldsymbol{h}(\boldsymbol{p}^0, u)$. Now $z(\boldsymbol{p}) \geq 0$ by construction, since $\boldsymbol{p}^\intercal \boldsymbol{q}^0$ is always greater or equal to the minimal cost of achieving u under prices $\boldsymbol{p}$. But $z(\boldsymbol{p})$ is known to achieve a minimum of 0 at $\boldsymbol{p} = \boldsymbol{p}^0$, so $z(\boldsymbol{p})$ has a stationary point at $\boldsymbol{p}^0$, i.e.,

$$\left.\frac{\partial z(\boldsymbol{p})}{\partial p_i}\right|_{\boldsymbol{p}=\boldsymbol{p}^0} = q_i^0 - \left.\frac{\partial c(\boldsymbol{p}, u)}{\partial p_i}\right|_{\boldsymbol{p}=\boldsymbol{p}^0} = 0.$$

Note that this equality depends on the strict convexity of preferences which we have assumed. Thus, $\partial c(\boldsymbol{p}, u)/\partial p_i = h_i(\boldsymbol{p}, u)$ as asserted. ∎

The property of the expenditure function $c(\boldsymbol{p}, u)$ that its derivative with respect to the ith price p_i yields the Hicksian demand for good i, $h_i(\boldsymbol{p}, u)$, is extremely useful in applications. *Exercise:* Check that the CES preferences in Examples 1 and 2 satisfy Theorem 1. Verify Shephard's Lemma for this specification.

In applications, one starts by estimating Marshallian demands $\boldsymbol{q}^* = \boldsymbol{g}(\boldsymbol{p}, m)$, as the quantities in this expression (quantity demanded, prices, and income) are observable. For welfare analysis, however, one invariably needs the compensated Hicksian demands $\boldsymbol{q}^* = \boldsymbol{h}(\boldsymbol{p}, u)$, since in these later quantities utility appears explicitly, and we can control its level. In short, Marshallian demands are estimable from data but not very useful for welfare analysis, while Hicksian demands are appropriate for welfare analysis but un-estimable directly as we lack data for utilities.

The strategy usually employed is to first estimate the Marshallian demands $\boldsymbol{g}(\boldsymbol{p}, m)$ from the observable price-income data and then derive the corresponding Hicksian demands $\boldsymbol{h}(\boldsymbol{p}, u)$. Given the Marshallian demand functions it is straightforward to obtain Hicksian demands by simply substituting m in $\boldsymbol{g}(\boldsymbol{p}, m)$ with the expenditure function $c(\boldsymbol{p}, u)$,

$$\boldsymbol{q}^* = \boldsymbol{g}(\boldsymbol{p}, m) = \boldsymbol{g}(\boldsymbol{p}, c(\boldsymbol{p}, u)) = \boldsymbol{h}(\boldsymbol{p}, u).$$

The reverse can also be done by substituting u in the Hicksian demands with the indirect utility function $v(\boldsymbol{p}, m)$,

$$\boldsymbol{q}^* = \boldsymbol{h}(\boldsymbol{p}, u) = \boldsymbol{h}(\boldsymbol{p}, v(\boldsymbol{p}, m)) = \boldsymbol{g}(\boldsymbol{p}, m).$$

This means that the two demand functions $\boldsymbol{g}(\boldsymbol{p}, m)$ and $\boldsymbol{h}(\boldsymbol{p}, u)$ *cross* at the current price level $\boldsymbol{p}$. This fact will be used latter on.

2.4.1. Roy's Identity

Finally, we may observe that the Marshallian demands $\boldsymbol{g}(\boldsymbol{p}, m)$ may be obtained from the indirect utility function $v(\boldsymbol{p}, m)$

Exercise: Verify Roy's identity for the CES preferences in Example 1.

Theorem 7. *The following is true*

(i) Roy's Identity:

$$g_i(\boldsymbol{p}, m) = -\frac{\partial v(\boldsymbol{p}, m)/\partial p_i}{\partial v(\boldsymbol{p}, m)/\partial m}, \qquad i = 1, ..., n. \tag{2.72}$$

(ii) Logarithmic Form of Roy's Identity: Let $w_i(\boldsymbol{p}, m) = p_i g_i(\boldsymbol{p}, m)/m$ be the budget share of good i. Then

$$w_i(\boldsymbol{p}, m) = -\frac{\partial \log v(\boldsymbol{p}, m)/\partial \log p_i}{\partial \log v(\boldsymbol{p}, m)/\partial \log m}, \qquad i = 1, ..., n. \tag{2.73}$$

(iii) Modified Form of Roy's Identity:

$$w_i(\boldsymbol{p}, m) = -\frac{\partial v(\boldsymbol{p}, m)/\partial \log p_i}{\partial v(\boldsymbol{p}, m)/\partial \log m}, \qquad i = 1, ..., n. \tag{2.74}$$

(iv) Diewert's Modified Form of Roy's Identity:

$$w_i(\mathfrak{p}) = \frac{\mathfrak{p}_i \partial v(\mathfrak{p})/\partial \mathfrak{p}_i}{\sum_{j=1}^n \mathfrak{p}_j \partial v(\mathfrak{p})/\partial \mathfrak{p}_j}, \qquad i = 1, ..., n. \tag{2.75}$$

Proof: (i) Differentiate the identity

$$v(\boldsymbol{p}, c(\boldsymbol{p}, u)) \equiv u$$

with respect to p_i to obtain

$$\frac{\partial v}{\partial p_i} + \frac{\partial v}{\partial m}\frac{\partial c}{\partial p_i} \equiv 0,$$

for $m = c$, and use that $q_i^* = \partial c/\partial p_i$ to obtain Roy's identity

$$q_i^* = \frac{\partial c}{\partial p_i} = -\frac{\partial v/\partial p_i}{\partial v/\partial m} = g_i(\boldsymbol{p}, m), \qquad i = 1, ..., n.$$

Alternatively, consider the Lagrangian of the [UMP], evaluated at the optimal values $\boldsymbol{q}^*$ and λ^* and considered as a function of $\boldsymbol{p}$ and m:

$$\mathscr{L}(\boldsymbol{p}, m; \boldsymbol{q}^*, \lambda^*) = u(\boldsymbol{q}^*) + \lambda^*(m - \boldsymbol{p}^\mathsf{T}\boldsymbol{q}^*).$$

Assuming an interior solution, the Envelope Theorem yields

$$\begin{aligned}
\frac{\partial v(\boldsymbol{p}, m)}{\partial p_i} = \frac{\partial \mathscr{L}(\boldsymbol{p}, m; \boldsymbol{q}^*, \lambda^*)}{\partial p_i} &= -\lambda^* q_i^*, \qquad i = 1, ..., n; \\
\frac{\partial v(\boldsymbol{p}, m)}{\partial m} = \frac{\partial \mathscr{L}(\boldsymbol{p}, m; \boldsymbol{q}^*, \lambda^*)}{\partial m} &= \lambda^*.
\end{aligned}$$

Recall that, although $\boldsymbol{q}^*$ and λ^* depend on $\boldsymbol{p}$ and m, the Envelope Theorem says that we don't need to consider this dependence when we differentiate the value function (here the Lagrangian) at the optimum, and we can treat $\boldsymbol{q}^*$ and λ^* as fixed! Dividing the two equations yields the result.

For a third derivation of the result (which is also the original derivation offered by Roy, 1942), note that at equilibrium, we must have

$$dv = 0 \quad \text{and} \quad \sum_{i=1}^n q_i^* dp_i = dm,$$

or

$$\frac{\partial v}{\partial p_1}dp_1 + \frac{\partial v}{\partial p_2}dp_2 + \cdots + \frac{\partial v}{\partial p_n}dp_n = -\frac{\partial v}{\partial m}dm$$

and

$$q_1^* dp_1 + q_2^* dp_2 + \cdots + q_n^* dp_n = dm,$$

which taken together imply

$$\frac{\partial v/\partial p_1}{q_1^*} = \cdots = \frac{\partial v/\partial p_n}{q_n^*} = -\frac{\partial v}{\partial m}$$

or

$$q_i^* = -\frac{\partial v/\partial p_i}{\partial v/\partial m}, \qquad i = 1, ..., n.$$

(ii) Recall that

$$\frac{\partial \log v}{\partial \log p_i} = \frac{\partial v}{\partial p_i}\frac{p_i}{v} \qquad \text{and} \qquad \frac{\partial \log v}{\partial \log m} = \frac{\partial v}{\partial m}\frac{m}{v},$$

so that, using the result in part (i),

$$\begin{aligned} w_i \equiv \frac{p_i q_i^*}{m} &= -\frac{p_i}{m}\frac{\partial v/\partial p_i}{\partial v/\partial m} \\ &= -\frac{(\partial v/\partial p_i)(p_i/v)}{(\partial v/\partial m)(m/v)} \\ &= -\frac{\partial \log v/\partial \log p_i}{\partial \log v/\partial \log m}, \qquad i = 1, ..., n. \end{aligned}$$

(iii) This part is considerably harder to prove rigorously. It was first stated and proved by Diewert (1974, p. 126) using *duality theory*. For now we can satisfy ourselves by appealing to the 'invariance' of the direct and indirect utility functions to strictly increasing transformations (if preferences are represented by u and v, they are also represented by $g(u)$ and $g(v)$ provided that g is strictly increasing) and noting that $\exp(\log v) = v$ is in this sense 'equivalent' to $\log v$ since $g(\cdot) = \exp(\cdot)$ is strictly increasing. ■

To summarize, by differentiating the expenditure function with respect to prices we obtain the Hicksian demand functions via Shephard's Lemma, while by differentiating the indirect utility function we get the Marshallian demands via Roy's identity. For empirical purposes, it is often convenient to start with estimating either $c(\boldsymbol{p}, u)$ or $v(\boldsymbol{p}, m)$, and then derive estimates of demands $\boldsymbol{h}(\boldsymbol{p}, u)$ and $\boldsymbol{g}(\boldsymbol{p}, m)$ by elementary differentiation or direct substitution.

As a matter of terminology we have the following conventions for cross derivatives introduced by Hicks:

$$\begin{aligned} &\partial h_i/\partial p_j > 0 && q_i, q_j \text{ are } \textit{substitutes} \\ &\partial h_i/\partial p_j < 0 && q_i, q_j \text{ are } \textit{complements} \\ &\partial g_i/\partial p_j > 0 && q_i, q_j \text{ are } \textit{gross substitutes} \\ &\partial g_i/\partial p_j < 0 && q_i, q_j \text{ are } \textit{gross complements.} \end{aligned}$$

We state a general theorem regarding various properties of the demand functions and then discuss their implications.

THEOREM 8. *Hicksian and Marshallian demands have the following properties:*

(i) *Adding-up:*

$$\boldsymbol{p}^\top \boldsymbol{h}(\boldsymbol{p}, u) = \boldsymbol{p}^\top \boldsymbol{g}(\boldsymbol{p}, m) = m,$$

i.e., the consumer spends all his income on the available commodities.

(ii) *Homogeneity of degree 0 in p: For all $\theta \in \mathbb{R}$,*

$$\boldsymbol{h}(\theta\boldsymbol{p}, u) = \boldsymbol{h}(\boldsymbol{p}, u) \quad \textit{and} \quad \boldsymbol{g}(\theta\boldsymbol{p}, \theta m) = \boldsymbol{g}(\boldsymbol{p}, m),$$

i.e., the consumer does not suffer from money illusion.

(iii) *Symmetry: For all $i, j = 1, ..., n$,*

$$\frac{\partial h_i(\boldsymbol{p}, u)}{\partial p_j} = \frac{\partial h_j(\boldsymbol{p}, u)}{\partial p_i}.$$

(iv) *Negative semi-definiteness: For all $\boldsymbol{\xi} \in \mathbb{R}^n$,*

$$\sum_{i=1}^n \sum_{j=1}^n \xi_i \xi_j \frac{\partial h_i(p, u)}{\partial p_j} \leq 0, \quad \textit{and} \quad \sum_{i=1}^n p_i \frac{\partial h_j(p, u)}{\partial p_i} = 0.$$

Proof: (i) Follows directly from nonsatiation which implies that optimal demands exhaust the available income.

(ii) Since $c(\boldsymbol{p}, u)$ is homogeneous of degree one in $\boldsymbol{p}$, its derivatives, $\boldsymbol{h}(\boldsymbol{p}, u)$, are homogeneous of degree zero in $\boldsymbol{p}$: $f(\theta\boldsymbol{p}) = \theta f(\boldsymbol{p}) \Rightarrow \nabla_{\boldsymbol{p}}(f(\theta\boldsymbol{p})) = \nabla_{\boldsymbol{p}} f(\theta\boldsymbol{p}) \times \nabla_{\boldsymbol{p}}(\theta\boldsymbol{p}) = \theta \nabla_{\boldsymbol{p}} f(\theta\boldsymbol{p}) = \theta \nabla_{\boldsymbol{p}} f(\boldsymbol{p})$, so $\nabla_{\boldsymbol{p}} f(\theta\boldsymbol{p}) = \nabla_{\boldsymbol{p}} f(\boldsymbol{p})$.

(iii) Symmetry is a trivial consequence of the analytic fact that the order of differentiation in the cross partial $\partial^2 c(\boldsymbol{p}, u)/\partial p_i \partial p_j$ doesn't matter[16].

[16]Symmetry of second derivatives, also called the equality of mixed partials, is also known as Young's theorem, or Schwarz's theorem, or Clairaut's theorem.

(iv) Negative semi-definiteness is a consequence of the concavity in $\boldsymbol{p}$ of $c(\boldsymbol{p}, u)$, while the singularity constraint is a consequence of (i). ∎

The first two results in Theorem 2 yield the following results.

Theorem 9. *(see Deaton and Muelbauer (1980) page 16) Differentiating the adding-up property*

$$\boldsymbol{p}^{\mathsf{T}}\boldsymbol{g}(\boldsymbol{p}, m) = m$$

(i) with respect to m we obtain

$$\textit{Engel Aggregation:} \qquad \sum_{i=1}^{n} p_i \frac{\partial g_i(\boldsymbol{p}, m)}{\partial m} = 1,$$

and (ii) with respect to p_i we obtain

$$\textit{Cournot Aggregation:} \qquad q_i + \sum_{j=1}^{n} p_j \frac{\partial g_j(\boldsymbol{p}, m)}{\partial p_i} = 0, \quad i = 1, ..., n.$$

(iii) Also, Euler's theorem and the homogeneity of degree 0 property of $\boldsymbol{g}(\boldsymbol{p}, m)$ yields

$$\textit{Euler Aggregation:} \qquad m\,\frac{\partial g_i(\boldsymbol{p}, m)}{\partial m} + \sum_{j=1}^{n} p_j\,\frac{\partial g_i(\boldsymbol{p}, m)}{\partial p_j} = 0, \quad i = 1, ..., n.$$

In terms of price and income elasticities e_{ij} and η_i respectively, and budget shares $w_i = p_i q_i^*/m$ the above identities can be restated as follows:

Theorem 10. *The following identities hold:*

$$\textit{(i) Engel Aggregation:} \qquad \sum_{i=1}^{n} w_i \eta_i = 1, \tag{2.76}$$

$$\textit{(ii) Cournot Aggregation:} \qquad w_i + \sum_{j=1}^{n} w_j e_{ij}^{g} = 0, \qquad i = 1, ..., n. \tag{2.77}$$

$$\textit{(iii) Euler Aggregation:} \qquad \eta_i + \sum_{j=1}^{n} e_{ij}^{g} = 0, \qquad i = 1, ..., n. \tag{2.78}$$

Example 4. The following Table presents demand price and income elasticity estimates for 10 commodities (9 food items and a nonfood aggregate) for the U.S., along with the budget shares for each commodity. Verify that the equalities of Theorem 4 are satisfied.

∎

TABLE 1. Demand price and income elasticity estimates for 10 commodities (9 food items and a nonfood aggregate) in the U.S..

Commodity	Uncompensated Own- and Cross- Price Elasticities, e_{ij} 1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	Income Elast. η_i	Budget Share w_i
1. Cereals & Bakery	**−0.93**	0.04	0.02	0.14	0.13	0.45	−0.04	−0.42	−0.06	0.39	**0.28**	.0151
2. Meat	0.02	**−0.40**	0.05	0.00	0.16	−0.12	−0.09	0.23	0.20	−0.69	**0.64**	.0252
3. Eggs	0.24	1.00	**−0.73**	0.66	−0.47	−0.54	0.27	0.25	−0.20	0.22	**−0.69**	.0012
4. Dairy	0.16	0.00	0.08	**−0.91**	−0.09	0.26	0.20	-0.26	0.17	−0.59	**0.97**	.0117
5. Fruits & Vegetables	0.14	0.32	−0.05	−0.07	**−0.58**	−0.15	0.11	0.20	−0.03	−0.16	**0.27**	.0183
6. Other foods	0.33	−0.17	−0.04	0.15	−0.11	**−0.62**	0.05	0.12	0.00	−0.50	**0.79**	.0244
7. Nonalcoholic Beverages	−0.06	−0.22	0.03	0.21	0.13	0.08	**−0.77**	−0.08	0.18	−0.37	**0.86**	.0093
8. Food Away From Home	−0.15	0.13	0.01	−0.07	0.06	0.05	−0.02	**−0.55**	−0.12	−0.19	**0.84**	.0666
9. Alcoholic Beverages	−0.05	0.24	−0.02	0.10	−0.02	0.00	0.10	−0.22	**−0.50**	−0.13	**0.50**	.0119
10. Nonfood	0.00	−0.03	0.00	−0.01	−0.01	−0.02	−0.01	−0.02	−0.02	**−0.94**	**1.07**	.8162

From Tables 13 and 20 of Okrent and Alston (2011). We omit the estimates' standard errors to avoid cluttering the table.

Aside. (Homothetic preferences.) Note that the symmetry of the Hicksian demands in part (iii) of Theorem 2 is *not* in general shared by the Marshallian demands. But if the utility function is *homothetic*[17], that is, if for every $\alpha > 0$, $u(\alpha\,\boldsymbol{q}) = \alpha\,u(\boldsymbol{q})$, then the Marshallian demands satisfy a similar symmetry, i.e.,

$$\frac{\partial g_i(\boldsymbol{p}, m)}{\partial p_j} = \frac{\partial g_j(\boldsymbol{p}, m)}{\partial p_i}. \tag{2.79}$$

For example, the CES utility function is homothetic and its Marshallian demands obey this symmetry condition.

Recall that the Lagrange multiplier λ of the primal problem [UMP] is equal to the marginal utility of income, and that this is generally a function of both prices $\boldsymbol{p}$ and income m, that is

$$\frac{\partial v(\boldsymbol{p}, m)}{\partial m} = \lambda(\boldsymbol{p}, m).$$

Thus, we can generally rewrite Roy's identity as

$$g_i(\boldsymbol{p}, m)\,\lambda(\boldsymbol{p}, m) = -\frac{\partial v(\boldsymbol{p}, m)}{\partial p_i}.$$

For homothetic preferences, however, λ is independent of the prices $\boldsymbol{p}$, so for homothetic preferences Roy's identity becomes

$$g_i(\boldsymbol{p}, m)\,\lambda(m) = -\frac{\partial v(\boldsymbol{p}, m)}{\partial p_i}.$$

The symmetry condition (2.79) now follows by the invariance of mixed partial derivatives of $v(\boldsymbol{p}, m)$ to the order of differentiation. That λ is independent of $\boldsymbol{p}$ implies a homothetic

[17] A homothetic function is an ordinal version of a homogeneous function of degree 1, i.e., it is a homogeneous function of degree 1 and all its monotonic transformations.

preference ordering with corresponding linear Engel curves and unitary income elasticities for all goods. (see Salvador Barbera, Peter Hammond, Christian Seidl (1999), Handbook of Utility Theory: Principles, Springer, p. 497-498.)

We know that homothetic preferences lead to constant budget shares given prices. Therefore, $\boldsymbol{g}(\boldsymbol{p}, m) = m\beta(\boldsymbol{p})$ for some scalar function $\beta(\boldsymbol{p})$ and, if we were to take the homogeneous representation of utility, which necessarily exists, then doubling m doubles demands and would double utility. Thus, more generally, $v(\boldsymbol{p}, m) = \phi(m/a(\boldsymbol{p}))$ for some scalar function $a(\boldsymbol{p})$ and some increasing $\phi(\cdot)$, and $\boldsymbol{c}(\boldsymbol{p}, u) = \phi^{-1}(u)a(\boldsymbol{p})$. The function $a(\boldsymbol{p})$ can be interpreted as a *price index* which is independent of utility u.

Preferences $\succeq$ are homothetic if $q^1 \succeq q^2 \Leftrightarrow \beta q^1 \succeq \beta q^2$ for any $\beta > 0$. With homothetic preferences there is really only one indifference curve: any indifference curve is a "radial blow-up" of any other. It is intuitively obvious but surprisingly hard to prove that the demand system in this case can be written as $\boldsymbol{g}(\boldsymbol{p}, m) = m\,\boldsymbol{h}(\boldsymbol{p}, 1) = m\,\beta(\boldsymbol{p})$ if and only if preferences are homothetic. With homothetic preferences all income elasticities are equal to 1 – a restriction that appears to be false for many goods. A classic result is that with identical homothetic preferences, aggregate demand is "as if" there were a single consumer with the same preferences and the total income of all consumers. (The proof is easy and left as an exercise). A more subtle result is that if different consumers have different homothetic preferences, and each consumer has a fixed share of total aggregate income (as prices and total income are varied) then aggregate demand is "as if" there were a single consumer with some homothetic preference ordering (see John Chipman, 2006 "Aggregation and Estimation in the Theory of Demand" History of Political Economy 38 (annual supplement), pp. 106-125.). ■

The $n \times n$ Jacobian matrix of the compensated demands, or Hessian matrix of the expenditure function, with respect to $\boldsymbol{p}$

$$\boldsymbol{S} = \left[s_{ij} \equiv \frac{\partial h_i(\boldsymbol{p}, u)}{\partial p_j} \right]_{i,j=1,\ldots,n} = \frac{\partial \boldsymbol{h}(\boldsymbol{p}, u)}{\partial \boldsymbol{p}^{\mathsf{T}}} = \frac{\partial^2 c(\boldsymbol{p}, u)}{\partial \boldsymbol{p}\, \partial \boldsymbol{p}^{\mathsf{T}}}, \tag{2.80}$$

is often called the *Slutsky matrix*, plays an extremely important role: it represents the demand response to changes in prices *holding utility constant.*

It is useful to have an expression for $\boldsymbol{S}$ in terms of Marshallian demands, since they can be estimated. This can be done by exploiting the fact we have already mentioned in our discussion above that the Marshallian and Hicksian demands cross at $\boldsymbol{p}$,

$$g_i(\boldsymbol{p}, c(\boldsymbol{p}, u)) = h_i(\boldsymbol{p}, u), \qquad i = 1, \ldots, n,$$

so

$$\frac{\partial g_i}{\partial p_j} + \frac{\partial g_i}{\partial m}\frac{\partial c}{\partial p_j} = \frac{\partial h_i}{\partial p_j} \equiv s_{ij}, \qquad i, j = 1, ..., n$$

and thus

$$s_{ij} = \frac{\partial g_i}{\partial p_j} + \frac{\partial g_i}{\partial m}\, q_j, \qquad i, j = 1, ..., n. \tag{2.81}$$

The effect of an infinitesimal change in p_j to the Marshallian demand for good i is thus give by

$$\frac{\partial g_i}{\partial p_j} = \frac{\partial h_i}{\partial p_j} - \frac{\partial g_i}{\partial m}\, q_j, \qquad i, j = 1, ..., n.$$

We see that this effect has been decomposed into two parts: (a) a *pure substitution effect*, and (b) a *pure income effect*. This is the classical *Slutsky decomposition*. If p_j, say, increases, the optimizing consumer increases his consumption of substitutes and lowers his consumption of complements, i.e. adjusts along the *same* indifference curve according to $\partial h_i/\partial p_j$, and also lowers proportionally – according to $\partial g_i/\partial m$ – his consumption of all goods by $q_j \partial g_i/\partial m$, reflecting the fact his *real* income is now lower.

The following theorem gives the result.

THEOREM 11. *The $n \times n$ matrix of substitution effects defined in (2.80) can be expressed in terms of the Marshallian demands as*

$$\boldsymbol{S} = \frac{\partial \boldsymbol{g}(\boldsymbol{p}, m)}{\partial \boldsymbol{p}^{\mathsf{T}}} + \frac{\partial \boldsymbol{g}(\boldsymbol{p}, m)}{\partial m} \boldsymbol{g}(\boldsymbol{p}, m)^{\mathsf{T}}, \tag{2.82}$$

and is (i) negative semidefinite and (ii) symmetric, (iii) has rank $n-1$, and (iv) its diagonal elements are negative.

Proof: Since the expenditure function is concave in prices, the Slutsky matrix is symmetric and negative semidefinite. ■

The Slutsky decomposition is somewhat more convenient in elasticity form, which can be easily obtained by multiplying (2.81) through by p_j/q_i

$$e^h_{ij} = \frac{\partial g_i}{\partial p_j}\frac{p_j}{q_i} + \frac{\partial g_i}{\partial m}\frac{m}{q_i}\frac{p_j q_j}{m} = e^g_{ij} + w_j\, \eta_i, \qquad i, j = 1, ..., n \tag{2.83}$$

where e^g_{ij} and e^h_{ij} are the Marshallian and Hicksian price elasticities respectively, η_i is the income elasticity of commodity i, and $w_j = p_j q_j/m$ is the budget share of commodity j. Note that the symmetry requirement $s_{ij} = s_{ji}$ does not imply that the corresponding Hicksian demand elasticities are equal, unless the budget shares of the two goods w_i and w_j are equal.

Equation (2.83) allows us to compute the compensated price elasticities given the uncompensated price and income elasticities. For normal goods $\eta_i > 0$ and thus $e^h_{ij} > e^g_{ij}$, while for inferior goods $\eta_i < 0$ and thus $e_{ij}^h < e_{ij}^g$. Focusing only on the own-price elasticities e_i^g and e_i^h, we have that for normal goods $e_i^g < e_i^h < 0$, while for inferior goods $e_i^h < e_i^g$. Interestingly, e_i^g could even be positive in the latter case (see the discussion of the Giffen good below), whereas utility maximization implies that e_i^h is always negative. We see that for normal goods, the Hicksian compensated demand curve is less responsive to price changes than is the Marshallian uncompensated demand curve, since the uncompensated demand curve reflects both income and substitution effects, while the compensated demand curve reflects only substitution effects.

2.4.2. Barten's Fundamental Matrix Equation of Demand Theory

The symmetry of the Slutsky matrix $s_{ij} = s_{ji}$ yields the *Slutsky equation*

$$\frac{\partial g_i}{\partial p_j} + \frac{\partial g_i}{\partial m} q_j = \frac{\partial g_j}{\partial p_i} + \frac{\partial g_j}{\partial m} q_i, \qquad i, j = 1, ..., n. \tag{2.84}$$

2.4.3. The Demand Integrability Problem

A problem that occupied economic theorists for several decades was to identify the restrictions that the assumption of utility maximization placed on demand functions. Another way to phrase this issue is: given a demand system is it the case that there is a utility function that generates it, and if so, how can it be recovered? Applied mathematicians tend to call this the *inverse optimization problem.* What was discovered is that under certain conditions, it is possible to solve differential equations to recover a utility function from a demand system. The following reasonably general result is taken from Hurwicz and Uzawa (1971).

Definition 1. *A system of Marshallian demands $\boldsymbol{g}(\boldsymbol{p}, m) : \mathbb{R}_{++}^n \times \mathbb{R}_+^n \to \mathbb{R}_+^n$ is called* **integrable** *if*

(i) they are positive;
(ii) they satisfy the adding-up condition $\boldsymbol{p}^\top \boldsymbol{g}(\boldsymbol{p}, m) = m$;
(iii) they are homogeneous of degree 0 in $(\boldsymbol{p}, m)$;
(iv) the Slutsky matrix of substitution effects

$$\boldsymbol{S} = \frac{\partial \boldsymbol{g}(\boldsymbol{p}, m)}{\partial \boldsymbol{p}^\top} + \frac{\partial \boldsymbol{g}(\boldsymbol{p}, m)}{\partial m} \boldsymbol{g}(\boldsymbol{p}, m)^\top$$

is symmetric and negative semidefinite.

Conditions (i)-(iv) are called the **integrability conditions**.

THEOREM 12 (Hurwicz–Uzawa Integrability Theorem). *Assume that the demand system $\boldsymbol{g}(\boldsymbol{p}, m) : \mathbb{R}_{++}^n \times \mathbb{R}_+^n \to \mathbb{R}_+^n$ is integrable and that it satisfies the following boundedness condition on the* *partial derivative with respect to income: For every* $0 \ll \underline{\boldsymbol{p}} \ll \overline{\boldsymbol{p}} \in \mathbb{R}^n_{++}$*, there exists a finite real number* C *such that for all* $m \gg 0$

$$\underline{\boldsymbol{p}} \leq \boldsymbol{p} \leq \overline{\boldsymbol{p}} \quad \Rightarrow \quad \left| \frac{\partial g_i(\boldsymbol{p}, m)}{\partial m} \right| \leq C.$$

Let Q *denote the range of* $\boldsymbol{g}$*, i.e.,*

$$Q = \{\boldsymbol{g}(\boldsymbol{p}, m) \subset \mathbb{R}^n_+ : (\boldsymbol{p}, m) \in \mathbb{R}^n_{++} \times \mathbb{R}^n_+\}.$$

Then there exists a utility function $u : Q \to \mathbb{R}$ *on the range* Q *such that for each* $(\boldsymbol{p}, m) \in \mathbb{R}^n_{++} \times \mathbb{R}^n_+$*,* $\boldsymbol{g}(\boldsymbol{p}, m)$ *is the unique maximizer of* $u(\boldsymbol{q})$ *over the budget set* $\{\boldsymbol{q} \in Q : \boldsymbol{p}^\mathsf{T}\boldsymbol{q} \leq m\}$*.*

We know from the Support Function Theorem or the Envelope Theorem that

$$\frac{\partial c(\boldsymbol{p}, u)}{\partial p_i} = h_i(\boldsymbol{p}, u) = g_i(\boldsymbol{p}, c(\boldsymbol{p}, v)).$$

Ignoring v for the moment, we have the **total differential equation**

$$c'(\boldsymbol{p}) = \boldsymbol{g}(\boldsymbol{p}, c(\boldsymbol{p})). \tag{2.85}$$

Following Hurwicz and Uzawa (1971), define the **income compensation function** as

$$\mu(\boldsymbol{p}; \boldsymbol{p}^0, m^0) = c(\boldsymbol{p}, v(\boldsymbol{p}^0, m^0)),$$

which is a function $\boldsymbol{p}$ alone ($\boldsymbol{p}^0$ and m^0 are fixed parameters). Now note that

$$\mu(\boldsymbol{p}^0; \boldsymbol{p}^0, m^0) = m^0,$$

and that

$$\frac{\partial \mu(\boldsymbol{p}; \boldsymbol{p}^0, m^0)}{\partial p_i} = \frac{\partial c(\boldsymbol{p}; v^0)}{\partial p_i} = h_i(\boldsymbol{p}; v^0) = g_i(\boldsymbol{p}; c(\boldsymbol{p}; v^0)) = g_i(\boldsymbol{p}; \mu(\boldsymbol{p}; p^0, m^0)).$$

We have proven the following theorem.

LEMMA 1. *The function* $c : \mathbb{R}^n_{++} \to \mathbb{R}$ *defined by* $c(\boldsymbol{p}) := \mu(\boldsymbol{p}; \boldsymbol{p}^0, m^0)$ *is the solution to the differential equation*

$$c'(\boldsymbol{p}) = \boldsymbol{g}(\boldsymbol{p}, c(\boldsymbol{p}))$$

that satisfies the initial condition $c(\boldsymbol{p}^0) = m^0$*.*

To summarize, the properties of adding up, homogeneity, negativity and symmetry are not only necessary but also sufficient for consumer optimization. If Marshallian demands satisfy these restrictions then there is a utility function $u(\boldsymbol{q})$ which they maximize subject to the budget constraint. We say in such a case that demands are integrable. The implied Hicksian demands define the expenditure function through a soluble set of differential equations by Shephard's Lemma. We know a system of demands is integrable if *any* of the following hold:

- They were derived as solutions to the primal or dual problem given a well specified direct utility function.
- They were derived by Shephard's Lemma from a cost function satisfying the appropriate requirements.
- They were derived by Roy's identity from an indirect utility function satisfying the appropriate requirements.
- They satisfy adding up, homogeneity, symmetry and negativity.

Aside. The symmetry of the Slutsky matrix implies that $\partial h_i/\partial p_j = \partial h_j/\partial p_i$, and thus the compensated demand vector $\boldsymbol{h}(\boldsymbol{p}, u)$ has

$$\operatorname{curl} \boldsymbol{h} = 0.$$

This means that the compensated demand vector $\boldsymbol{h}(\boldsymbol{p}, u)$ forms a *conservative vector field*. This in turn implies that there exists a *scalar potential energy function* $c(\boldsymbol{p}, u)$ such that

$$\boldsymbol{h} = \nabla_{\boldsymbol{p}}\, c.$$

In economics this scalar "energy" function is the expenditure function, and the above relation is Sheppard's Lemma. ■

Afriat (1980), Demand Functions and the Slutsky Matrix, p.3:
The Slutsky theory is a familiar topic in economics. Also, it has its own mathematical interest. But it has no value for a heavy matter such as the empirical foundation of classical utility as Slutsky and others thought. Slutsky even imagined that the immateriality of the order of utility differentiations, instead of being merely a consequence of the continuity of second derivatives, expressed indifference to the order of consumption (whether the main course comes before the dessert or vice versa). Pareto's notion (1897, pp. 251, 270, 539 ff.) was similar, as can be gathered from Stigler (1965, p. 124), though for him it is instead a matter of integrations.
p. 8
Convexity [of the utility surfaces of a consumer] is characterized by the existence of a linear support at any point...

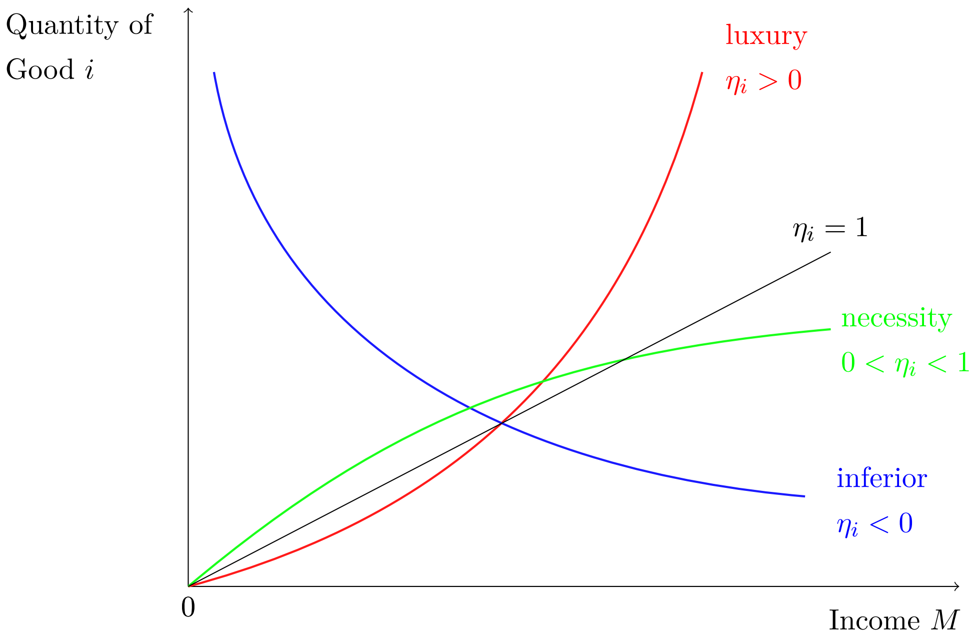


FIGURE 4. Engel curves.

2.5. THE GIFFEN GOOD

The own-Slutsky effect, s_{ii}, is necessarily negative since it is the demand response along an indifference curve to a change in a good's own price (to see this, let $\boldsymbol{\xi} = (0, ..., 1, ..., 0)^{\mathsf{T}}$ be the n vector with the ith element equal to 1 and all other elements equal to 0, in part (iv) of Theorem 2). But note that, notoriously, the derivative of the Marshallian demand with respect to own price

$$\frac{\partial g_i}{\partial p_i} = s_{ii} - q_i \frac{\partial g_i}{\partial m}, \qquad i = 1, ..., n, \tag{2.86}$$

can be positive if the very last term of the equation above, $\partial g_i/\partial m$, is sufficiently negative! This is the infamous *Giffen effect* which Marshall introduced:

$$\text{good } i \text{ is } \textit{Giffen} \text{ if its Marshallian demand slopes } \textit{upward}\text{, i.e., } \frac{\partial g_i}{\partial p_i} > 0.$$

Obviously, this can only happen if x_i is an *inferior* good, i.e., when $\partial g_i/\partial m < 0$, so much so, that the second term in (2.86) overtakes s_{ii}. So a Giffen good must necessarily be inferior, but an inferior good need not be Giffen.

Of course, most goods have positive income effects, hence the usage *normal* goods to describe cases in which $\partial g_i/\partial m > 0$. In terms of income elasticities, the following terminology is in use:

$\eta_i < 0$	i is an *inferior* good
$\eta_i > 0$	i is a *normal* good
$0 < \eta_i < 1$	i is a *necessity*
$\eta_i > 1$	i is a *luxury*.

It should be emphasized that "inferiority" can only be a *local* property: A good cannot be inferior over the whole range of consumption or else it is not a good (i.e., a desirable) thing but a "bad" thing that would never be consumed anyway (for example, garbage). This means that it is impossible to generate Giffen behavior using commonly used specifications like Cobb-Douglas and CES utility functions, since in these specifications either a commodity is globally good or it is globally bad. As it turns out, it is quite hard to analytically construct utility functions that exhibit the Giffen property. For this reason most authors of microeconomics textbooks limit themselves to graphical representations and general arms waving, and do not give any analytic examples of utility functions with Giffen behavior in a part of their domain. The following example presents one such function.

Ernst Engel, 1821-1896, was a Prussian statistician, founder of the International Statistical Institute and from 1860 to 1882 he was director of the Prussian statistical bureau in Berlin. However, he is best known for the formulation of his Engel's law, deriving on what is known as the Engel curve.

Engel developed his famous curve in his book "Die Productions- und Consumtionsverhältnisse des Königreichs Sachsens" 1857, from observing and collecting data of the consumption patterns of Belgian working-class families, and he related their level of income with their expenditure in food and other goods. He observed that households with higher incomes tended to allocate a lower share of their income to food than poorer households. The Engel curve captures this inverse relation. His law is a reflection of this phenomenon and states this same relation. As Engel himself expressed, the implication of this law is very interesting in the macroeconomic sense as it implies that the higher the economic development of a country, the lower the share of agriculture will be in aggregate production.

Example 5. The Wold-Jurèen (1953) utility function is given by

$$u(x_1, x_2) = \frac{(x_1 - 1)}{(2 - x_2)^2},$$

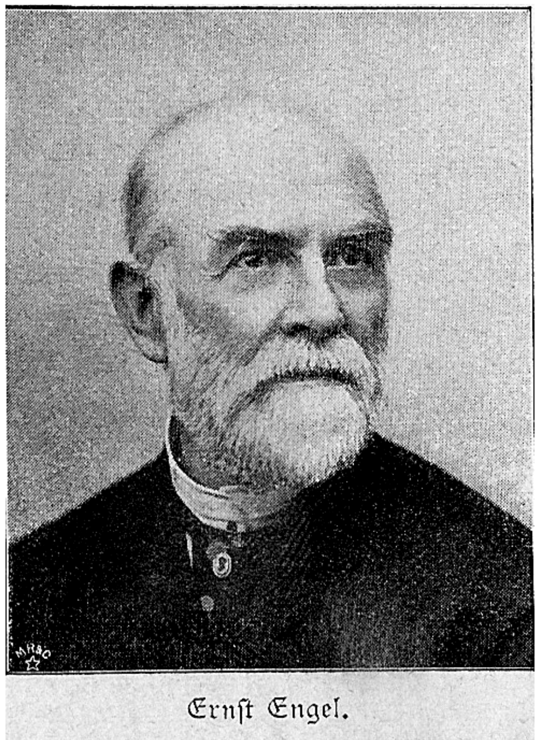


FIGURE 5. Ernst Engel (1821–1896)

with domain $x_1 > 1$ and $0 \leq x_2 < 2$. Figure 1 plots the indifference curves for the region of interest $x_1 > 1$, $0 < x_2 \leq 2$. In this region, x_1 is a Giffen good: an increase in the price of x_1 shifts the budget constraint to the dashed line and results in an increase in the quantity demanded of x_1. Note also the huge drop in the demand for the luxury good x_2.

■

In empirical applications we are often interested in. Let p^0 and p^1 be two price levels and y be the consumer's income. The *money metric indirect utility function* defined by

$$\psi(p^1; p^0, m) = c(p^1, \upsilon(p^0, m)),$$

measures the monetary compensation a consumer with money income m would require in order to be indifferent between current prices p^0 and new prices p^1. Letting (p^0, m^0) and (p^1, m^1) be two price-income states (say, the old and the new, respectively), an obvious measure of welfare change is just the difference in indirect utility

$$\upsilon(p^1, m^1) - \upsilon(p^0, m^0).$$

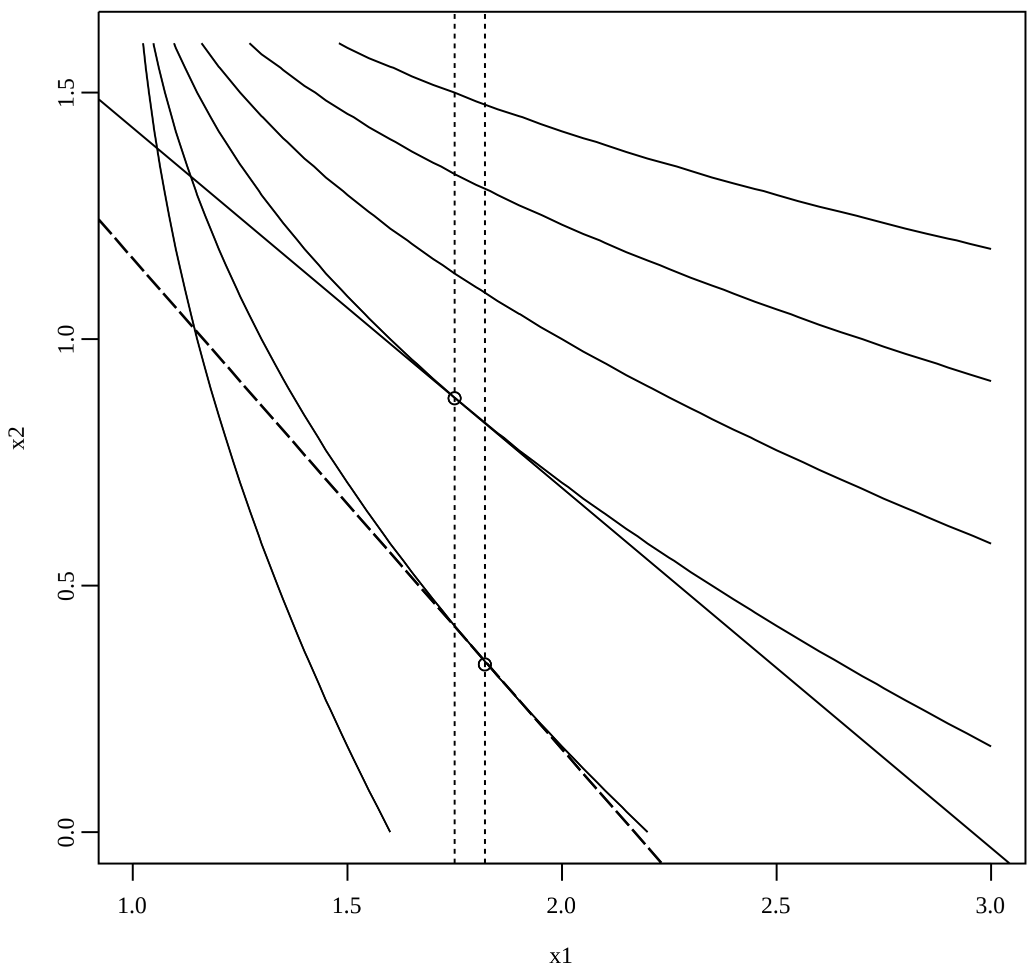


FIGURE 6. The Wold-Jurèen (1953) utility function.

In terms of the function ψ we may define two welfare measures: *equivalent variation* (EV) and *compensating variation* (CV), given by

$$
\begin{aligned}
EV &= \psi(p^0; p^1, m^1) - \psi(p^0; p^0, m^0) = \psi(p^0; p^1, m^1) - m^0 \\
CV &= \psi(p^1; p^1, m^1) - \psi(p^1; p^0, m^0) = m^1 - \psi(p^0; p^1, m^1)
\end{aligned}
$$

EV measures the additional income a consumer would require to accept the new price-income state.

2.6. The Veblen Good

Another violation of the Law of Demand is presented by the *Veblen good*, named after the American sociologist and economist Thorstein Veblen, who first identified *conspicuous consumption* as a mode of status-seeking in *The Theory of the Leisure Class* (1899). Veblen goods are types of luxury goods for which the quantity demanded increases as the price increases, an apparent contradiction of the law of demand, resulting in an upward-sloping demand curve. A higher price may make a product desirable as a status symbol in the practices of conspicuous consumption and *conspicuous leisure*. A product may be a Veblen good because it is a *positional good*, something few others can own.

Conspicuous consumption is the spending of money on and the acquiring of luxury goods and services to publicly display economic power (income and wealth). To the conspicuous consumer, such a public display of discretionary economic power is a means of either attaining or maintaining a given social status. The development of Thorstein Veblen's sociology of conspicuous consumption produced the term *invidious* or *ostentatious consumption*, that is, the consumption of goods that is meant to provoke the *envy* of other people. Related to this is the notion of *conspicuous compassion*, that is, the deliberate use of charitable donations of money or property in order to enhance the social prestige of the donor, with a display of superior socioeconomic status.

The classical formulation of consumer choice presented above *cannot* accommodate Veblen goods, since the only possibility for an upward slopping demand curve there is the Giffen case, which requires that the good in question be inferior. Clearly, a Veblen good must be a luxury, not an inferior, good, and therefore it's Marshallian demand curve with respect to own-price must necessarily be downward slopping.

There is however an ingenuous modification of the classical setting introduced by Ng (1987), yielding results that are in accordance with our intuitive notion of a Veblen good as described above. The utility maximization problem that Ng considers is given by

$$\max_{(q_1,\ldots,q_n)\in\mathbb{R}^n_+} u(p_1q_1/p_n, q_2, \ldots, q_n) \quad \text{such that} \quad \sum_{i=1}^{n} p_iq_i = m, \tag{2.87}$$

where $u : \mathbb{R}^n_+ \to \mathbb{R}$ is a quasiconcave function of its arguments, and $q_i, p_i, i = 1, \ldots, n$ are quantities demanded and prices of the $i = 1, \ldots, n$ goods. The modification pertains to the introduction of the p_1q_1/p_n term in the utility function, where p_1q_1 is the value of the consumption in the first good that will have the Veblen property, and p_n is the price of the nth good that acts as numeraire[18]. Ng (1987) imagines that the Veblen good 1 is diamonds, and the conspicuous consumption term p_1q_1/p_n is the real money value expenditure for diamonds. The consumer derives utility not from the amount of diamonds consumed q_1, but from the real money value expenditure for diamonds. If $u(\cdot)$ is assumed strictly increasing in its arguments, then the more expensive relative to other goods diamonds are, the more utility the conspicuous consumer derives from their consumption.

The Lagrangian of the problem is given by

$$\mathscr{L} = u(p_1q_1/p_n, q_2, ..., q_n) + \lambda(m - \sum_{i=1}^{n} p_iq_i).$$

Assuming an interior solution, the first-order conditions for optimality are

$$\begin{aligned} \frac{\partial \mathscr{L}}{\partial(p_1q_1/p_n)} &= \frac{\partial u}{\partial(p_1q_1/p_n)} - \lambda\frac{\partial p_1q_1}{\partial(p_1q_1/p_n)} \equiv u_1 - \lambda p_n = 0, \qquad (2.88) \\ \frac{\partial \mathscr{L}}{\partial q_i} &= \frac{\partial u}{\partial q_i} - \lambda\frac{\partial p_iq_i}{\partial q_i} \equiv u_i - \lambda p_i = 0, \quad i = 2, ..., n, \\ \frac{\partial \mathscr{L}}{\partial \lambda} &= m - \sum_{i=1}^{n} p_iq_i = 0, \end{aligned}$$

where subscripts of u denote partial differentiation with respect to the corresponding *argument* of u.

3. Empirical Application: The Deadweight Loss of Christmas.

Joel Waldfogel's (1993) American Economic Review paper provides a stylized (and controversial) example of the application of the *Carte Blanche* principle. Waldfogel observes that gift-giving is equivalent to an in-kind transfer and hence should be less efficient for consumer welfare than simply giving cash. In January, 1993, he surveyed approximately 150 Yale undergraduates about their holiday gifts received in 1992:

(i) What were the gifts worth in cash value?
(ii) How much the students be willing to pay for them if they didn't already have them?
(iii) How much would the students be willing to accept in cash in lieu of the gifts. (Usually higher than willingness to pay – an economic anomaly.)

[18]One could replace p_n in p_1q_1/p_n by a price index P of the prices of the $i = 2, ..., n$ goods (the non-Veblen goods) without altering the results of the analysis.

For each gift, Waldfogel calculates the gift's yield $Y_j = V_j/P_j$. As theory (and intuition) would predict, the yield was, on average, well below one. That is, in-kind gift giving 'destroys' value relative to the cost-equivalent cash gift.

The estimated value equation for gifts is (standard errors in parentheses):

$$\begin{aligned} \log(\text{value}_i) &= \underset{(-0.44)}{-0.314} + \underset{(0.08)}{0.964} \log(\text{price}_i) \\ R^2 &= 0.656, \quad n = 86, \quad \hat{\sigma}_u = NA. \end{aligned}$$

The Waldfogel article generated a surprising amount of controversy, even among economists, most of whom probably subscribe to the *Carte Blanche* principle. To many readers, this article seems to exemplify the well-worn gripe about economists, "They know the price of everything and the value of nothing". What is Waldfogel missing?

4. Empirical Application: Was Bread Giffen?

`Science, at bottom, is really anti-intellectual. It always distrusts pure reason and demands the production of objective facts.`

— H. L. Mencken

Perhaps with something similar to the above quotation in mind, Roger Koenker (1977) has called the Giffen good "the Loch Ness monster of economics": it is rumored to exist in the deep, murky lake of theoretical possibilities, but to date no one has ever been able to compellingly document its existence (see also Stigler, 1948). In view, however, of the rationality of Giffen behavior on the part of the consumer under certain circumstances discussed above, we should perhaps be a little less 'severe' and liken it to the Tasmanian tiger instead: the Tasmanian tiger is considered extinct today, but unlike the Loch Ness monster, it is an animal that certainly existed in the recent past and may still be alive in the remote and uninhabited Tasmanian forests. Likewise, Giffen behavior may very well have existed among the medieval peasants or the post-industrial-revolution city poor, and may still be found among the poorest populations on Earth.

Indeed, Jensen and Nolan (2008)

The budget data presented below are taken from Koenker (1977) who considered two studies of the standard of living of English rural laborers conducted during the period 1787-1795 by the Reverend David Davies and Mr. Frederick Morton Eden. These were among the first examples of studies in the long and honorable tradition of econometrically snooping into the private lives of the poor. By the mid 19th century, such studies were being conducted all over Europe by such notables as Ernst Engel, Frederick Engels, Frederick LePlay and others. The novelist Jane Austen (1775-1817) lived during and wrote about the period that Davis and Eden collected the data analyzed here, often referred to as the *Georgian era* of British history (named after the Hanoverian kings George I, George II, George III and George IV) that lasted from 1714 to 1837.

Recall that a good is Giffen if the derivative of its Marshallian demand with respect to its own price is positive. It has been hypothesized that such an anomaly is most likely to occur in situations where the households under study are very poor living at a subsistence level. Then, an increase in the price of an inferior good (like bread) might lead to a shift of funds used to purchase a luxury good (like meat) to the inferior good, so much so, that that the luxury good is almost eliminated and the demand for the inferior good is increased! Or as Marshall himself puts it in his classic textbook,

> a rise in the price of bread makes so large a drain on the resources of the poorer labouring families and raises so much the marginal utility of money to them, that they are forced to curtail their consumption of meat [...]
>
> – Alfred Marshal, *Principles of Economics*, 8th ed., p. 132.

The sample examined by Koenker (1977) is thus ideal for discovering Giffen behavior since the households considered were indeed very poor and bread and meat could very plausibly play the roles of the inferior and luxury goods respectively that theory prescribes.

The price of bread is in old pence per half-peck loaf [19] and the price of meat is in old pence per pound of bacon. Bacon was the lowest quality of meat, consisting essentially of the outer fat and skin of the animal. In the rare cases where the meat consumed was not bacon, an equivalent quantity (in money terms) of bacon was computed. Similarly, in cases where the household purchased flour rather than bread an equivalent quantity of bread was computed. Table 1 gives sample statistics (we will need these to compute elasticities at the mean).

We assume a linear system of *Marshallian demands* for bread (b) and meat (m) given by

$$q_i = \alpha_{i0} + \alpha_{i1}\, s + \gamma_i\, m + \sum_{i,j} \beta_{ij}\, p_j + u_i, \qquad i, j = b, m. \tag{4.1}$$

[19] A half-peck (or gallon) of loaf was made with a gallon of flour, and weighed 8 pounds and 11 ounces, or 8.6875 pounds (3.9406 kg). It was considered that a gallon of bread (a little over a pound a day, 1 pound = 0.4536 kg) was the basic ration for one adult for one week, and it was on this basis that laborer's wages were based.

TABLE 2. Sample Statistics

Variable	*Definition*	*Mean*	*Std.Dev.*
q_b	Quantity demanded of bread (gallons/week)	4.772	1.534
q_m	Quantity demanded of meat (pounds/week)	1.706	1.324
p_b	Price of bread (pence per gallon)	13.629	3.172
p_m	Price of meat (pence per pound)	7.923	1.097
w_b	Budget share of bread (%)	.669	.113
w_m	Budget share of meat (%)	.136	.081
m	Total Expenditure (pence)	95.171	26.796
s	Family size (no. of members)	5.714	1.619

$n = 35$ observations

We also assume that errors are jointly normally distributed with mean zero and covariance matrix

$$\Sigma = \begin{bmatrix} \sigma_{u_b}^2 & \rho\sigma_{u_b}\sigma_{u_m} \\ \rho\sigma_{u_b}\sigma_{u_m} & \sigma_{u_m}^2 \end{bmatrix}.$$

We estimate the Marshallian demands as a system by the method of Seemingly Unrelated Regressions (SUR). Although the two equations have the same set of regressors and therefore SUR here is equivalent to equation-by-equation ordinary least squares (OLS), the joint estimation of the two equations has the advantage that it yields estimates of the covariances of coefficients both within the same equation and across different equations. OLS only provides covariances for coefficients within the same equation but in order to test hypotheses that put restrictions involving parameters from both equations (e.g. to test Slutsky symmetry, see bellow) we also need the across-equations covariances.

The estimated demand equations for bread and meat are (standard errors in parentheses):

$$\begin{aligned} q_b = \;& \underset{(0.958)}{0.3932} + \underset{(0.104)}{0.415}\; s + \underset{(0.0073)}{0.0243}\; m - \underset{(0.058)}{0.354}\; p_b + \underset{(0.150)}{0.571}\; p_m \\ & R^2 = 0.797, \quad n = 35, \quad \chi_4^2 = 137.0, \quad \hat{\sigma}_{u_b} = 0.682, \end{aligned}$$

$$\begin{aligned} q_m = \;& \underset{(0.968)}{7.815} - \underset{(0.106)}{0.437}\; s + \underset{(0.0074)}{0.0411}\; m + \underset{(0.059)}{0.162}\; p_b - \underset{(.151)}{1.229}\; p_m \\ & R^2 = 0.721, \quad n = 35, \quad \chi_4^2 = 90.4, \quad \hat{\sigma}_{u_m} = 0.721, \end{aligned}$$

$$\hat{\rho} = -0.41, \quad \chi_1^2 = 5.829.$$

The fit of the models as measured by the R^2 is quite high, and both equations are significant overall since the χ^2_4 values of 137 for the bread equation and 90.4 for meat equation reject emphatically the null that all slope coefficients in the respective equations are zero. Also, all individual coefficients in both equations are statistically significant at the 5% level, and only the intercept in the bread equation is insignificant at the 1% level.

The uncompensated Marshallian elasticity estimates evaluated at the sample means are as follows:[20]

$$\hat{e}^g_{bb} = -1.01, \quad \hat{e}^g_{bm} = 0.95, \quad \hat{\eta}_b = 0.49,$$
$$\hat{e}^g_{mm} = -5.71, \quad \hat{e}^g_{mb} = 1.29, \quad \hat{\eta}_m = 2.29.$$

We see that bread is a necessity and meat a luxury, but both are normal goods, that is, we do *not* find bread to be Giffen (nay, it is not even inferior). An interesting find is that the coefficient for family size is, positive in the bread equation, but negative in the meat equation. This means that larger families consume more bread and less meat than smaller families.

To compute the Slutsky matrix we use equation (2.81) and evaluate at the sample means (numbers in parentheses are standard errors[21]):

$$\hat{S} \;=\; \begin{bmatrix} \hat{\gamma}_b\bar{q}_b + \hat{\beta}_{bb} & \hat{\gamma}_b\bar{q}_m + \hat{\beta}_{bm} \\ & \\ \hat{\gamma}_m\bar{q}_b + \hat{\beta}_{mb} & \hat{\gamma}_m\bar{q}_m + \hat{\beta}_{mm} \\ & \end{bmatrix} = \begin{bmatrix} -0.238 & 0.612 \\ (0.052) & (0.149) \\ 0.358 & -1.159 \\ (0.052) & (0.150) \end{bmatrix}.$$

As expected, the diagonal elements are negative, since they measure the slope of the compensated Hicksian demands for bread and meat. Since both bread and meat are normal goods these slopes are smaller in absolute value than the corresponding slopes of the Marshallian demands. The positive cross effects in the off-diagonal elements indicate that bread and meat are substitutes. The eigenvalues of this matrix are $(\hat{\lambda}_1, \hat{\lambda}_2) = (-1.36, -0.042)$ and since they are both negative, $\hat{S}$ satisfies the negative-semidefiniteness condition. Symmetry requires that

$$H_0 : \gamma_m\bar{q}_b + \beta_{mb} = \gamma_b\bar{q}_m + \beta_{bm}. \tag{4.2}$$

[20]Price elasticities are computed from the formula $\hat{e}^g_{ij} = \dfrac{d\hat{q}_i}{dp_j}\dfrac{\bar{p}_j}{\bar{q}_i} = \hat{\beta}_{ij}\dfrac{\bar{p}_j}{\bar{q}_i}$, and income elasticities are computed from the formula $\hat{\eta}_i = \dfrac{d\hat{q}_i}{dm}\dfrac{\bar{m}}{\bar{q}_i} = \hat{\gamma}_i\dfrac{\bar{m}}{\bar{q}_i}$, where $i, j = b, m$ for bread and meat, respectively. As is done often in this context, the sample means $\bar{q}_b$, $\bar{q}_m$, $\bar{p}_b$, $\bar{p}_m$ and $\bar{m}$ are treated as constants without variance.

[21]Note that in order to compute standard errors for the elements of the Slutsky matrix we need the covariances between coefficients from different equations (for example, we need $\text{Cov}(\hat{\beta}_{bb}, \hat{\beta}_{mm})$) that are available from SUR estimation but equation-by-equation OLS estimation does not provide. The same is true for testing Slutsky symmetry. The sample means $\bar{q}_b$ and $\bar{q}_m$ are again treated as constants without variance.

Testing for symmetry we obtain a χ^2 statistic on 1 degree of freedom of 3.16 with a p-value of .0761 meaning that we accept symmetry at the 10% level, but we reject it at the 5% level. Thus, although $\hat{s}_{bm} = 0.612$ is almost twice as large as $\hat{s}_{mb} = 0.358$, this disparity can be explained away as the result of noise in the data, at least at the 10% level[22]. In fact, looking at the standard errors of the two estimates, we see that $\hat{s}_{mb} = 0.358$ is much more accurately estimated than $\hat{s}_{bm} = 0.612$, with the standard error of the latter being 3 times that of the former.

We proceed to impose Slutsky symmetry as a constraint in estimation, that is, we re-estimate the SUR model under the linear constraint in (4.2). The constrained SUR estimates are (standard errors in parentheses):

$$\begin{aligned}
q_b = \;& \underset{(0.727)}{1.537} + \underset{(0.108)}{0.429}\, s + \underset{(0.0076)}{0.0246}\, m - \underset{(0.052)}{0.303}\, p_b + \underset{(0.055)}{0.324}\, p_m \\
& R^2 = .78, \quad n = 35, \quad \chi^2_4 = 156.49, \quad \hat{\sigma}_{u_b} = 0.708, \\
q_m = \;& \underset{(0.947)}{7.441} - \underset{(0.106)}{0.442}\, s + \underset{(0.0074)}{0.0415}\, m + \underset{(0.059)}{0.168}\, p_b - \underset{(0.150)}{1.192}\, p_m \\
& R^2 = 0.69, \quad n = 35, \quad \chi^2_4 = 88.17, \quad \hat{\sigma}_{u_m} = 0.692,
\end{aligned}$$

$$\hat{\rho} = -0.38, \quad \chi^2_1 = 5.014.$$

The symmetry-constrained estimated Slutsky matrix is (with s.e.'s in parentheses below the point estimates)

$$\tilde{S} = \begin{bmatrix} -0.185 & \\ (0.043) & \\ 0.366 & -1.121 \\ (0.052) & (0.149) \end{bmatrix},$$

with eigenvalues $(\hat{\lambda}_1, \hat{\lambda}_2) = (-1.25, -0.059)$, and since they are both negative, $\tilde{S}$ also satisfies the negative-semidefiniteness condition. We see that the Slutsky cross coefficient $\tilde{s}_{bm} = \tilde{s}_{mb} = 0.366$ has a value that is very close to the $\hat{s}_{mb} = 0.358$ value of the unconstrained matrix.

The symmetry-constrained elasticity estimates evaluated at the sample means are:

$$\tilde{e}^g_{bb} = -0.87, \quad \tilde{e}^g_{bm} = 0.54, \quad \tilde{\eta}_b = 0.49,$$
$$\tilde{e}^g_{mm} = -5.54, \quad \tilde{e}^g_{mb} = 1.34, \quad \tilde{\eta}_m = 2.32.$$

[22]This is why we should formally and rigorously test hypotheses and not rely on just eyeballing the point estimates that may lead us to unwarranted conclusions. The fact that $\hat{s}_{bm}$ and $\hat{s}_{mb}$ *look* very different should not lead us to conclude that Slutsky symmetry fails here.

We see that the meat elasticities are very little affected by the constraint, but some of the estimated bread elasticities change considerably. In particular, imposing Slutsky symmetry lowers the estimate for e^g_{bb} from -1.01 to -0.87, and also lowers the estimate of e^g_{bm} from 0.95 to 0.54. Thus, if Slutsky symmetry is imposed, bread becomes price-inelastic and much less responsive to changes in meat prices, meaning that bread and meat are not close substitutes, which is not very surprising. This is in agreement with the cross-coefficient of the Slutsky matrix of only 0.366 that is quite low. Income elasticity estimates for both bread and meat are very little affected by the constraint.

In any case, both the unconstrained and the constrained models agree that although bread was a necessity and meat was a luxury, both bread and meat were *normal* goods. So, to answer our original question, we conclude that bread was *not* Giffen.

5. DEMAND SYSTEMS

A popular example of directly specified demand equations is the *double-log demand system*,

$$\log q_i = \alpha_i + \eta_i \log m + \sum_{j=1}^{n} e_{ij} \log p_j, \qquad i = 1, ..., n, \tag{5.1}$$

where q_i is the quantity demanded of good i $(i = 1, ..., n)$; p_j is the price of good j;

$$m = \sum_{k=1}^{n} p_i q_i, \tag{5.2}$$

is total expenditure, which we shall refer to as income for short; and α_i, η_i and e_{ij} are constant coefficients to be estimated from the data. All logarithms here are natural logarithms.

The interpretation of (5.1) is straightforward. The coefficient of income is

$$\eta_i = \frac{\partial(\log q_i)}{\partial(\log m)}, \tag{5.3}$$

which is the income elasticity of demand for good i and answers the question, if income rises by 1 percent with prices constant, what is the percentage change in consumption of i? Commodities with income elasticities less than unity are called *necessities*, while those with income elasticities greater than unity are known as *luxuries*. If the income elasticity is negative, then the good is said to be *inferior* as its consumption falls with increasing income. Similarly, the coefficient e_{ij} is the elasticity of demand for good i with respect to the price of good j,

$$e_{ij} = \frac{\partial(\log q_i)}{\partial(\log p_j)}, \tag{5.4}$$

and gives the percentage change in q_k resulting from a 1 percent change in p_j, income and the other prices held fixed.

Although the double-log demand system is attractive in its simplicity, it does not satisfy the adding-up condition, which means that it not consistent with *any* utility optimization setting. We say that the double-log demand system does not satisfy the 'integrability conditions' (among which is the adding-up condition) required by any 'proper' demand system.

To see how the double-log demand system fails to satisfy the adding-up condition, consider a household expenditure survey in which all participating families pay approximately the same price for each good, so that the k-th equation of (5.1) reduces to

$$\log q_i = \alpha_i + \eta_i \log m, \qquad i = 1, ..., n,$$

where units are chosen such that the price of each good is unity, so that $\sum_{j=1}^{n} e_{ij} \log p_j = 0$. Then the logarithmic change in expenditure on good i is a constant multiple η_i of the change in income, i.e.

$$d(\log p_i q_i) = \eta_i \, d(\log m).$$

Accordingly, if the income elasticity η_i exceeds unity, then expenditure on i increases at a faster rate than does income. If income rises sufficiently, expenditure on good i will eventually exceed income and violate the adding-up constraint (5.2). Therefore, the weakness of the model is that it does not satisfy the adding-up constraint (5.2) for all values of income.

5.1. The Cobb Douglas (CD) Demand System

5.2. The Constant Elasticity of Substitution (CES) Demand System

5.3. The Homothetic Translog Indirect Utility Function

The homothetic translog indirect utility function is given by

$$\log v(\mathfrak{p}) = \alpha_0 + \sum_{i=1}^{n} \alpha_i \log \mathfrak{p}_i + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \beta_{ji} \log \mathfrak{p}_i \log \mathfrak{p}_j \tag{5.5}$$

with the following restrictions imposed

$$\beta_{ij} = \beta_{ji}, \quad \text{for all } i, j = 1, ..., n; \tag{5.6}$$

$$\sum_{i=1}^{n} \beta_{ij} = 0, \qquad \text{for all } j = 1, ..., n; \tag{5.7}$$

$$\sum_{i=1}^{n} \alpha_i = 1. \tag{5.8}$$

Note that the symmetry $\beta_{ij} = \beta_{ji}$ condition is true only for homothetic preferences. This function is a generalization of the Cobb-Douglas function and reduces to it when all β_{ji} are equal to zero. In fact, when all β_{ji} are equal to zero, the homothetic translog reduces to

$$\log v(\mathfrak{p}) = \alpha_0 + \sum_{i=1}^{n} \alpha_i \log \mathfrak{p}_i \tag{5.9}$$

which is the Cobb-Douglas, written in logs.

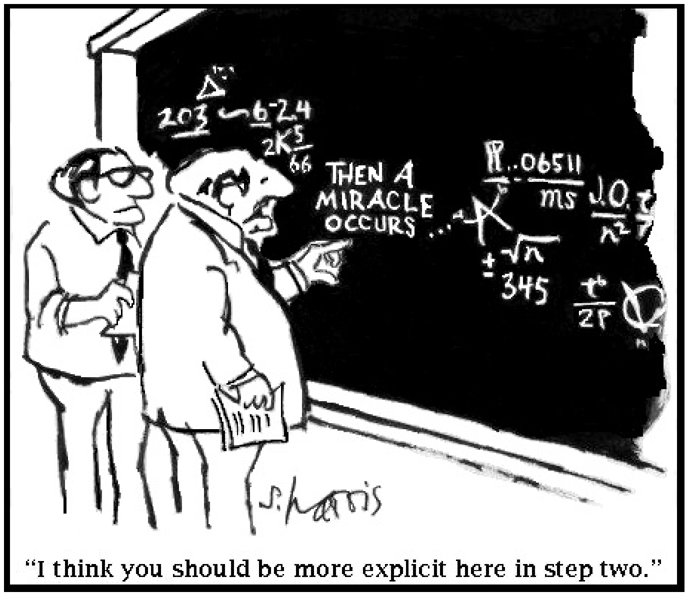
Application of Roy's identity in share form then yields a set of share equations for the homothetic translog

$$w_i = \alpha_i + \sum_{j=1}^{n} \beta_{ij} \log \mathfrak{p}_j, \qquad i = 1, ..., n. \tag{5.10}$$

With n products, the n homothetic translog share equations have $n(n+3)/2$ parameters to be estimated. For example, let us assume that there are only three products ($n = 3$). In this three-product case the homothetic translog share equations become

$$\begin{aligned}
w_1 &= \alpha_1 + \beta_{11} \log \mathfrak{p}_1 + \beta_{12} \log \mathfrak{p}_2 + \beta_{13} \log \mathfrak{p}_3; \\
w_2 &= \alpha_2 + \beta_{12} \log \mathfrak{p}_1 + \beta_{22} \log \mathfrak{p}_2 + \beta_{23} \log \mathfrak{p}_3; \\
w_3 &= \alpha_3 + \beta_{13} \log \mathfrak{p}_1 + \beta_{23} \log \mathfrak{p}_2 + \beta_{33} \log \mathfrak{p}_3,
\end{aligned}$$

and have 9 parameters, α_1, α_2, α_3, β_{11}, β_{12}, β_{13}, β_{22}, β_{23}, and β_{33}.



"I think you should be more explicit here in step two."

FIGURE 7. Cartoon by Sidney Harris.

6. MATHEMATICAL APPENDIX

```
LONG long ago, when wishing still could lead to something,
there lived a king whose daughters all were beautiful...
```

— Grimms' Fairy Tales, No. 1, "The Frog King", opening lines.

Because we don't live "long long ago", and wishing can't lead to anything as the cartoon by Sidney Harris above shows, we need a firm footing in mathematics.

6.1. LINEAR ALGEBRA

6.2. QUASICONVEX AND QUASICONCAVE FUNCTIONS

Some functions that arise in economic models are not convex (concave) but possess a weaker property called quasiconvexity (quasiconcavity).

Definition 2. *Assume that $S \subset \mathbb{R}^n$ is a convex set and $\phi : S \to \mathbb{R}$. We say that ϕ is a quasiconvex function if, for every $\boldsymbol{x}^0, \boldsymbol{x}^1 \in S$ ($\boldsymbol{x}^0 \neq \boldsymbol{x}^1$):*

$$\phi(\alpha\boldsymbol{x}^0 + (1-\alpha)\boldsymbol{x}^1) \leq \max\{\phi(\boldsymbol{x}^0), \phi(\boldsymbol{x}^1)\}, \text{ for every } \alpha \in (0,1).$$

We say that ϕ is a quasiconcave function if $-\phi$ is quasiconvex or, equivalently, if:

$$\phi(\alpha\boldsymbol{x}^0 + (1-\alpha)\boldsymbol{x}^1) \geq \min\{\phi(\boldsymbol{x}^0), \phi(\boldsymbol{x}^1)\}, \text{ for every } \alpha \in (0,1).$$

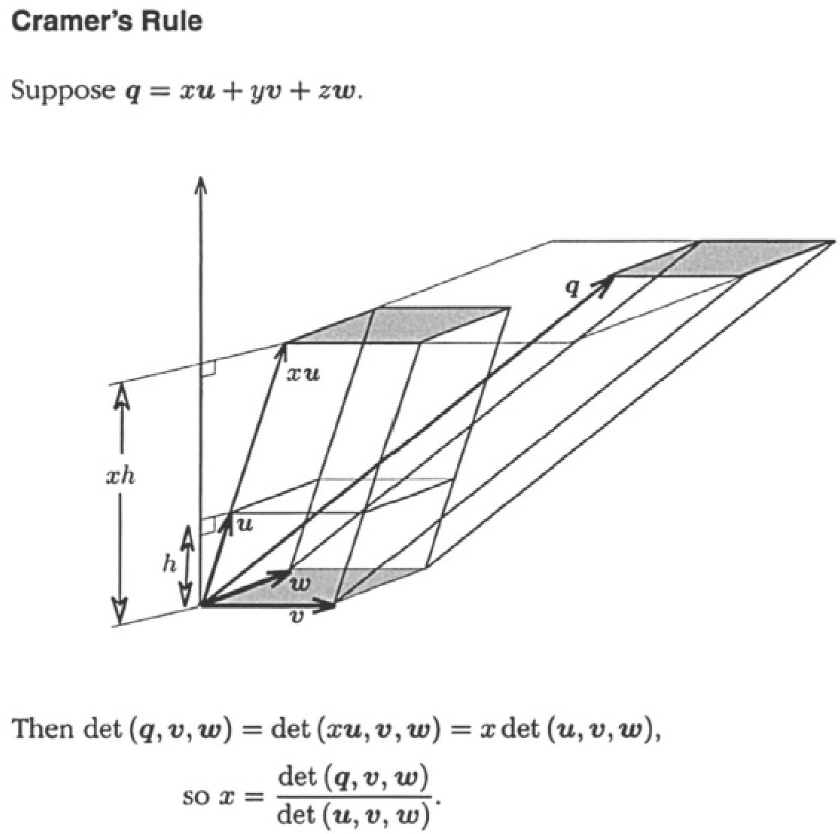


Figure 8. Cramer's Rule, *College Mathematics Journal*, 1997, **28**:2, 118.

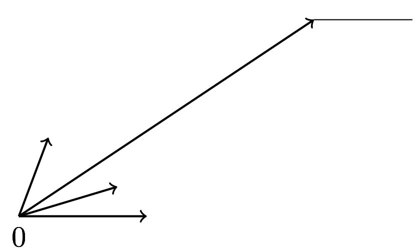


Figure 9. Cramer's Rule

The quasiconvexity/quasiconcavity is strict if the inequality in the above definitions is strict.

6.3. The Implicit Function Theorem

Suppose we have a system of n nonlinear equations depending on n endogenous variables $\boldsymbol{y} = (y_1, ..., y_n)^\top$ and m exogenous variables (also called parameters) $\boldsymbol{x} = (x_1, ..., x_m)^\top$ given by

$$\begin{aligned} f_1(y_1, ..., y_n; x_1, ..., x_m) &= 0 \\ &\vdots \\ f_n(y_1, ..., y_n; x_1, ..., x_m) &= 0, \end{aligned}$$

which we may write compactly as

$$\boldsymbol{f}(\boldsymbol{y}, \boldsymbol{x}) = \mathbf{0}, \tag{6.1}$$

where $\boldsymbol{f} : A \times B \to \mathbb{R}^m$ is a continuously differentiable vector function. We assume that the domain of the endogenous variables $\boldsymbol{y}$ is $A \subseteq \mathbb{R}^n$ and the domain of the parameters $\boldsymbol{x}$ is $B \subseteq \mathbb{R}^m$, where A and B are open sets, so that $A \times B \subseteq \mathbb{R}^{n+m}$.

Suppose that $\boldsymbol{y}^* \in A$ and $\boldsymbol{x}^* \in B$ satisfy the system of equations in (6.1), that is $f_i(\boldsymbol{y}^*, \boldsymbol{x}^*) = 0$ for all $i = 1, ..., n$. We are then interested in the possibility of solving for $\boldsymbol{y}$ as a function of $\boldsymbol{x}$. That is, we are interested in the existence of n uniquely determined "implicit" functions $\boldsymbol{g}(\cdot) = (g_1(\cdot), ..., g_n(\cdot))^\top$ such that

$$f_i(g_1(\boldsymbol{x}), ..., g_n(\boldsymbol{x}); \boldsymbol{x}) = 0 \qquad \text{for all } i = 1, ..., n, \text{ and } \boldsymbol{x} \in B' \tag{6.2}$$

and

$$g_i(\boldsymbol{x}^*) = y_i^* \qquad \text{for all } i = 1, ..., n. \tag{6.3}$$

The implicit function theorem gives a sufficient condition for the existence of such implicit functions and tells us the first-order comparative statics effects of $\boldsymbol{x}$ on $\boldsymbol{y}$ at a solution.

Theorem 13. (Implicit Function Theorem) *Suppose that $\boldsymbol{f} : A \times B \to \mathbb{R}^m$ is continuously differentiable with respect to the $n+m$ variables $(\boldsymbol{y}, \boldsymbol{x})$, and that $\boldsymbol{f}(\boldsymbol{y}^*; \boldsymbol{x}^*) = \mathbf{0}$. If the Jacobian matrix of the system $\boldsymbol{f}(\boldsymbol{y}; \boldsymbol{x}) = \mathbf{0}$ with respect to the endogenous variables $\boldsymbol{y}$ evaluated at $(\boldsymbol{y}^*, \boldsymbol{x}^*)$ is nonsingular, that is, if*

$$J = \begin{vmatrix} \dfrac{\partial f_1(\boldsymbol{y}^*; \boldsymbol{x}^*)}{\partial y_1} & \cdots & \dfrac{\partial f_1(\boldsymbol{y}^*; \boldsymbol{x}^*)}{\partial y_n} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial f_n(\boldsymbol{y}^*); \boldsymbol{x}^*}{\partial y_1} & \cdots & \dfrac{\partial f_n(\boldsymbol{y}^*); \boldsymbol{x}^*}{\partial y_n} \end{vmatrix} \neq \mathbf{0}, \tag{6.4}$$

then there exist implicitly defined functions $g_i : B' \to A'$, $i = 1, ..., n$ that are continuously differentiable and satisfy (6.2) and (6.3), where A' and B' are open subsets of A and B, *respectively. Moreover, the first-order effects of* $\boldsymbol{x}$ *on* $\boldsymbol{y}$ *at* $(\boldsymbol{y}^*, \boldsymbol{x}^*)$ *are given by*

$$\nabla_{\boldsymbol{x}}\, \boldsymbol{g}(\boldsymbol{x}^*)_{n\times m} = -\Big[\nabla_{\boldsymbol{y}} \boldsymbol{f}(\boldsymbol{y}^*; \boldsymbol{x}^*)\Big]^{-1}_{n\times n} \nabla_{\boldsymbol{x}} \boldsymbol{f}(\boldsymbol{y}^*; \boldsymbol{x}^*)_{n\times m}. \tag{6.5}$$

These effects may also be computed by Cramer's rule.

The first part of the theorem regarding the *existence* of the continuously differentiable functions $g_i(\boldsymbol{x}), i = 1, ..., n$ that satisfy (6.2) and (6.3) is the deep part of the implicit function theorem and its proof requires advanced methods. The second part in which we obtain the derivatives of $g_i(\boldsymbol{x}), i = 1, ..., n$ is almost trivial since continuously differentiable functions are locally linear, but because it is so useful in applications we state it explicitly. Indeed, if $\boldsymbol{g}(\boldsymbol{x})$ exists and if both $\boldsymbol{f}(\boldsymbol{y}; \boldsymbol{x})$ and $\boldsymbol{g}(\boldsymbol{x})$ are continuously differentiable around $(\boldsymbol{x}^*, \boldsymbol{y}^*)$ and $\boldsymbol{f}(\boldsymbol{y}^*; \boldsymbol{x}^*) = 0$, then the differential of $\boldsymbol{f}(\boldsymbol{y}; \boldsymbol{x})$ evaluated at $(\boldsymbol{y}^*, \boldsymbol{x}^*)$ is

$$d\boldsymbol{f} = \nabla_{\boldsymbol{y}} \boldsymbol{f}(\boldsymbol{y}^*; \boldsymbol{x}^*)\, d\boldsymbol{y} + \nabla_{\boldsymbol{x}} \boldsymbol{f}(\boldsymbol{y}^*; \boldsymbol{x}^*)\, d\boldsymbol{x} = \boldsymbol{0}, \tag{6.6}$$

where $d\boldsymbol{y} = (dy_1, ..., dy_n)^\mathsf{T}$ is the $n \times 1$ vector of differentials of $\boldsymbol{y}$ and $d\boldsymbol{x} = (dx_1, ..., dx_m)^\mathsf{T}$ is the $m \times 1$ vector of differentials of $\boldsymbol{x}$. Then using (6.2) and (6.3) we get $d\boldsymbol{y} = \nabla_{\boldsymbol{x}} \boldsymbol{g}(\boldsymbol{x}^*) d\boldsymbol{x}$ so that

$$\nabla_{\boldsymbol{y}} \boldsymbol{f}(\boldsymbol{y}^*; \boldsymbol{x}^*)_{n\times n} \nabla_{\boldsymbol{x}} \boldsymbol{g}(\boldsymbol{x}^*)_{n\times m}\, d\boldsymbol{x}_{m\times 1} + \nabla_{\boldsymbol{x}} \boldsymbol{f}(\boldsymbol{y}^*; \boldsymbol{x}^*)_{n\times m}\, d\boldsymbol{x}_{m\times 1} = \boldsymbol{0}_{n\times 1}. \tag{6.7}$$

This is the linearization of system (6.1) around $(\boldsymbol{x}^*, \boldsymbol{y}^*)$ that yields the second conclusion of the implicit function theorem. The condition in (6.4) guarantees that the $n \times n$ matrix $\nabla_{\boldsymbol{y}} \boldsymbol{f}(\boldsymbol{y}^*; \boldsymbol{x}^*)$ is invertible, so that

$$\nabla_{\boldsymbol{x}} \boldsymbol{g}(\boldsymbol{x}^*)\, d\boldsymbol{x} = -\Big[\nabla_{\boldsymbol{y}} \boldsymbol{f}(\boldsymbol{y}^*; \boldsymbol{x}^*)\Big]^{-1} \nabla_{\boldsymbol{x}} \boldsymbol{f}(\boldsymbol{y}^*; \boldsymbol{x}^*)\, d\boldsymbol{x},$$

and the coefficients of $d\boldsymbol{x}$ must be equal. This way of computing the gradient of $\boldsymbol{g}(\boldsymbol{x})$ at x^* is called *implicit differentiation*.

6.4. The Envelope Theorem

Shephard's Lemma and Roy's Identity are special cases (applications) of the Envelope Theorem.

Many properties of optima of functions are fundamentally consequences of the following simple observation:

The maximal value of a constrained maximization problem cannot increase (and the minimal value of a constrained minimization problem cannot decrease) as

(1) existing constraints become more stringent, and/or

(2) new constraints are added.

Given a value function $V(\boldsymbol{a})$, we wish to evaluate the effect of a change in the parameters $\boldsymbol{a}$. If we have a closed-form expression for V, we simply take the derivative of this expression with respect to $\boldsymbol{a}$ and we are done. But we often find ourselves in situations where V is only *implicitly* defined (i.e. we know that such a function exists and, being a value function, satisfies 'certain restrictions' and has 'certain properties') but there is no closed-form expression available to differentiate directly.

THEOREM 14. (ENVELOPE THEOREM)
Consider the problem () when there is just one constraint and suppose the objective function f and the constraint function g are continuously differentiable in $(\boldsymbol{x}, \boldsymbol{a})$ on an open subset $W \times U$ of $\mathbb{R}^n \times A$. For each $\boldsymbol{a} \in U$, suppose that $\boldsymbol{x}(\boldsymbol{a}) \in W$ uniquely solves (A2.35), is continuously differentiable in $\boldsymbol{a}$ on U, and that the constraint $g(\boldsymbol{x}(\boldsymbol{a}), \boldsymbol{a}) \leq 0$ is binding for every $\boldsymbol{a} \in U$. Let $\mathscr{L}(\boldsymbol{x}, \boldsymbol{a}, \lambda)$ be the associated Lagrangian function and let $(\boldsymbol{x}^*(\boldsymbol{a}), \lambda^*(\boldsymbol{a}))$ solve the Kuhn-Tucker conditions in Theorem A2.20. Finally, let $V(\boldsymbol{a})$ be the problem's associated value function. Then, the Envelope theorem states that for every $\boldsymbol{a} = (a_1, ..., a_m)^\mathsf{T} \in U$,

$$\frac{\partial V(\boldsymbol{a})}{\partial a_j} = \left.\frac{\partial \mathscr{L}}{\partial a_j}\right|_{\boldsymbol{x}^*(\boldsymbol{a}), \lambda^*(\boldsymbol{a})}, \qquad j = 1, ..., m. \tag{6.8}$$

where the right-hand side denotes the partial derivative of the Lagrangian function with respect to the parameter a_j evaluated at the point $(\boldsymbol{x}^*(\boldsymbol{a}), \lambda^*(\boldsymbol{a}))$.

The theorem says that the *total* effect on the optimized value of the objective function when a parameter changes (and so, presumably, the whole problem must be reoptimized) can be deduced simply by taking the partial of the problem's Lagrangian with respect to the parameter and then evaluating that derivative at the solution to the original problem's first-order Kuhn-Tucker conditions. Although $\boldsymbol{x}^*(\boldsymbol{a})$ and $\lambda^*(\boldsymbol{a})$ also depend on $\boldsymbol{a}$, the envelope theorem says that we don't need to consider this dependence when we differentiate the Lagrangian at the optimum, and we can thus treat $\boldsymbol{x}^*(\boldsymbol{a})$ and $\lambda^*(\boldsymbol{a})$ as fixed. Put differently, we may quickly and easily differentiate an optimal value function V at $\boldsymbol{a}^*$ without thinking about any implicit changes in the optimal choice, or the multiplier, that may occur in the background - even when the constraint is active - by differentiating the associated Lagrangian function $\mathscr{L}$ exclusively with respect to the parameter $\boldsymbol{\alpha}$ and evaluating that derivative at $(\boldsymbol{x}^*(\boldsymbol{a}), \lambda^*(\boldsymbol{a}))$.

Although we have confined ourselves in the statement of the theorem to the case of a single constraint, the theorem applies regardless of the number of constraints, with the usual proviso that there be fewer constraints than choice variables.

Proof: First, form the Lagrangian for the maximization problem: ■

References

Barten, AP. (1964), "Consumer Demand Functions Under Conditions of Almost Additive Preferences", *Econometrica*, 32: 1-38.

Barten, A., (1969), "Maximum Likelihood Estimation of a Complete System of Demand Equations", *European Economic Review*, 1, pp. 7-73.

Barten A., and V. Böhm (1982). "Consumer Theory", in K. Arrow and M. Intrilligator (eds.), *Handbook of Mathematical Economics*. Amsterdam: North Holland, 2: 382–429.

Christensen, L.R., D.W. Jorgenson, and L.J. Lau, (1975), "Transcendental Logarithmic Utility Functions", *American Economic Review*, 65, pp. 367-383.

Deaton, A.S., and J. Muellbauer (1980), "An Almost Ideal Demand System", *American Economic Review*, 70, pp. 312-326.

Deaton, A.S., and J. Muellbauer (1980), *Economics and Consumer Behaviour*, Cambridge.

Deaton, A.S. (1998), *The Analysis of Economic Surveys*, Johns Hopkins.

Debreu, G. (1954). "Representation of a Preference Ordering by a Numerical Function", in R. M. Thrall et al. (eds.), *Decision Processes*. New York: John Wiley, 159–165.

Debreu, G. (1972). "Smooth Preferences", *Econometrica*, 40: 603–615.

Hicks, J.R. (1956), *A Revision of Demand Theory*, London.

Houthakker, H.S., and L.D. Taylor (2010), *Consumer Demand in the United States: Prices, Income, and Consumption Behavior*, Springer-Verlag, New York.

Hurwicz, L., and H. Uzawa, 1971. On the integrability of demand functions. In: Chipman, J., Hurwicz, L., Richter, M., Sonnenschein, H. (Eds.), Preferences, Utility, and Demand. Harcourt Brace, Jovanovich, New York.

Jensen, Robert T., and Nolan H. Miller (2008) "Giffen Behavior and Subsistence Consumption", *American Economic Review*, 98, pp. 1553-1577.

Koenker, R. (1977). "Was Bread Giffen? The Demand for Food in England circa 1790", *Review of Economics and Statistics*, 59, pp. 225-229.

Klein, L.R. and H. Rubin (1948). "A Constant-Utility Index of the Cost of Living," *Review of Economic Studies*, 15: 84-7.

McKenzie, L.W. (1957), "Demand Theory Without a Utility Index", *Review of Economic Studies*, 24, pp. 185-189.

Ng, Yew-Kwang (1987) – "Diamonds Are a Government's Best Friend: Burden-Free Taxes on Goods Valued for Their Values", *American Economic Review*, 77(1), pp. 186-191.

Samuels, W.J. (1990), *Economics As Discourse: An Analysis of the Language of Economists*, Springer Netherlands.

Stigler, G. (1948), "Notes on the History of the Giffen Paradox", *Journal of Political Economy*, 56, pp. 61-62.

Stone, R.N. (1954), "Linear Expenditure Systems and Demand Analysis: An Application to the Pattern of British Demand", *Economic Journal*, 64, pp. 511-527.

Theil, H. (1965), "The Information Approach to Demand Analysis", *Econometrica*, 33: 67-87.

Theil, H. (1971), *Principles of Econometrics*, New York: John Wiley and Sons.

Theil, H. (1975/76), *Theory and Measurement of Consumer Demand*, 2 vols., Amsterdam: North Holland Publishing Company.

Theil, H. (1980), *The System-wide Approach to Microeconomics*, Chicago: The University of Chicago Press.

Varian, H.R. (1992), *Microeconomic Analysis*, 3rd edition, New York: W.W. Norton & Company.

Wold, H., and L. Jureen (1953), *Demand Analysis: A Study in Econometrics.* New York: Wiley.

Table 3. Weekly Budgets of English Rural Laborers, Koenker(1977).

County	Date	Expenditure for Bread (pence)	Expenditure for Meat (pence)	Family Size	Total Expenditure (pence)	Price of Bread (pence)	Price of Meat (pence)
		79	8	7	99	11.5	8
		68.5	16	7	96.5	11.5	8
		68.5	8	6	88.5	11.5	8
		32	21	5	75	11.5	3.3
Berks	1787	53	12	4	75	11.5	8
		48	20	5	78	11.5	8
		56	18	5	90.5	13.5	8
		98	12	7	124	13.5	8
		50	12	3	81.25	13.5	8
		95.5	0	8	120.25	13.5	8
		49.5	0	7	60.5	13	7.5
		61	12	6	85.5	13	7.5
		37.5	8	5	58	13	7.5
		37	8	4	65.5	13	7.5
		43	12	5	68.5	13	7.5
Dorset	1789	43	8	4	61	13	7.5
		59	30	4	106.25	13	7.5
		59	30	7	108.5	13	7.5
		50	10.5	4	70.75	13	7.5
		41	22.5	4	80.25	13	7.5
		58.5	15	6	88.5	13	7.5
Derby	1788	54	18	6	104	12	7.5
		74	8	6	99	11.5	8
		40	0	4	69.75	11.5	8
		58	8	5	84	11.5	8
		95	4	9	113	11.5	8
Dorset	1789	75	0	8	107.25	11.5	8
		79	4	5	89	11.5	8
		98	8	9	115.25	14	8
		84	24	8	162	14	8
		48	12	5	87	14	8
		59.5	18	4	109	22	10
		87	12	6	113	22	10
Oxford	1795	117	36	8	183.5	22	10
		78	18	4	114	22	10