Economic Evaluation and Transport Modelling: Theory and Practice

John Bates

Resource paper

Moving through nets:
The physical and social dimensions of travel
10th International Conference on Travel Behaviour Research
Lucerne, 10-15. August 2003

Economic Evaluation and Transport Modelling: Theory and Practice

John Bates

Abstract

A central role for Transport Modelling is to allow alternative strategies to be assessed. Most people involved in modelling are aware of the general principles of economic evaluation, but it remains a specialised topic, and tends to take second place in the literature to the more glamorous discussion of models. However, if the assessment tools are not integrated with the models themselves, much of the model sophistication may be wasted.

It was therefore considered appropriate to commission a resource paper for the 10th IATBR Conference. The paper has a number of aims. It begins by discussing the theoretical foundations of evaluation in economics, and, without being excessively academic, attempts to give an understanding of how the theory has developed. It then contrasts this with current practice, examining the validity of the simplifications which are typically made, as well as the developments in this area.

Continuing progress in the theory of discrete choice models, as well as the development of their use in transport modelling, has led to some key theoretical conclusions, which serve usefully to integrate modelling and evaluation. However, there are still some conflicts between the ideal theoretical requirements and what is considered practical.

The paper is primarily aimed at mathematical modellers rather than at economists, most of whom will be familiar with much of the material. It includes the essential theory of the consumer that is available in economic textbooks, but without requiring the reader to become too involved in general questions of economic theory. Particular emphasis is placed on practical issues.

Finally, the paper draws attention to a number of controversial areas, where further work is required, both theoretical and practical.

An admission is due at the outset. Part of my intention for the paper was to effect a synthesis of the theory of modelling and evaluation in a way which I did not believe was available in an accessible form. I had not seen the excellent paper by Jara-Diaz & Farah (1988) which does indeed achieve such a synthesis, although I was aware of other work by Jara-Diaz *et al*. I believe that we are in substantial agreement, though there are some nuances of presentational difference.

There remain some unresolved issues. It is hoped that, both at the conference itself and thereafter, this paper may assist in stimulating their resolution. It is also my intention that the final version of the paper should benefit substantially from advice and criticism received. In particular, appropriate contributions from others involved in this area will be gratefully received, incorporated, and acknowledged! Since the practical issues discussed in Section 3 are rather heavily based on the experience in the UK, examples demonstrating different approaches from other countries would also be most welcome.

1. History of Evaluation

1.1 Consumer Surplus

The current approach has its roots in the pioneering work of the French economist Dupuit (1844), who carefully discussed the value ("utility", as he termed it) of public investment, taking the particular example of a bridge. The essential reasoning is as follows:

> "Suppose that all those similar commodities of which we want to discover the utilities are all subjected to a tax which rises by small steps. Each successive increase will cause a certain quantity of the commodity to disappear from consumption. This quantity, multiplied by the rate of tax, will give its utility expressed in money. By thus letting the tax go up until there are no more consumers, and by adding together all the products of this multiplication process, we will arrive at the total utility of the goods.
>
> Let us illustrate this formula by an example. We want to know the utility of a footbridge which is being used free of charge at the rate of 2,080,000 crossings annually. Suppose that a toll of 0 fr. 01 would reduce the number by 330,000, that a tax of 0 fr. 02 reduces it by 294,000, and so on. We then say that for 330,000 crossings the utility is about 0 fr. 01 and that for the next 294,000 crossings the utility is about 0 fr. 02 and we can then draw up the table [below].

No. of crossings disappearing	as toll rises by 0.01 fr to:	implied utility (francs)
330,000	0.01	3,300
294,000	0.02	5,880
260,000	0.03	7,800
228,000	0.04	9,120
198,000	0.05	9,900
170,000	0.06	10,200
144,000	0.07	10,080
120,000	0.08	9,600
98,000	0.09	8,820
78,000	0.10	7,800
60,000	0.11	6,600
44,000	0.12	5,280
30,000	0.13	3,900
18,000	0.14	2,520
8,000	0.15	1,200
TOTAL		
2,080,000		102,000

Thus 102,000 francs would be the absolute utility to society of the bridge. We can find the relative utility by deducting the costs of maintenance and the interest on the capital expended in construction. If this latter sum were to reach or exceed 102,000 francs the construction would have produced no utility, the difference expressing the loss which would have been made. Such is the calculation to be made in the case where crossing is free of charge. If there is a toll we must take only the figures below that of the charge. Thus for a toll of 0 fr. 05, for example, the absolute utility of the bridge is expressed by the sum of the ten last figures or 66,000 francs; the utility lost, by the sum of the first five, or 36,000 francs; the product of the toll would be 770,000 crossings at 0 fr. 05 or 38,500 francs. With this toll, then, the possible utility of the bridge would be distributed in the following manner:

To the toll collector	38,500
Derived by those crossing the bridge (66,000-38,500)	27,500 "
Loss of utility arising from the 1,310,000 crossings which would have been made but for the toll	36,000 "
-Total	***102,000 "***

This example, set out with textbook clarity, is the direct forerunner of the concept of "Consumer Surplus". Dupuit's article is not merely of interest for its clear exposition, but also because of the way it shows that many of the other suggestions for measuring benefit current at the time were fallacious.

Using Dupuit's figures, we can construct the (Marshallian) demand curve, as in Figure 1 below. Note that in economic theory it is standard to reverse the normal mathematical functionality for demand curves, and plot demand along the X-axis and cost along the Y-axis. If we write P for the price of the toll, and T for the number of crossings, then the Marshallian demand curve can be written:

$$T = f(P) \qquad (1.1)$$

However, given the way the figure is conventionally drawn, we deal with the **inverse** demand curve:

$$P = f^{-1}(T) \qquad (1.2)$$

representing the price at which demand for crossings would be equal to T.

The total value (willingness to pay) is seen as the area under the demand curve. When a non-zero price is charged, the area is only measured up to the actual demand at that price, and from this total willingness to pay **at that price**, we must subtract the amount actually paid, to obtain the user benefit, or Consumer Surplus.

Thus, at a given price P′, the total willingness-to-pay among those choosing to cross is given as:

$$W = \int_0^{T(P')} f^{-1}(T).dT \qquad (1.3)$$

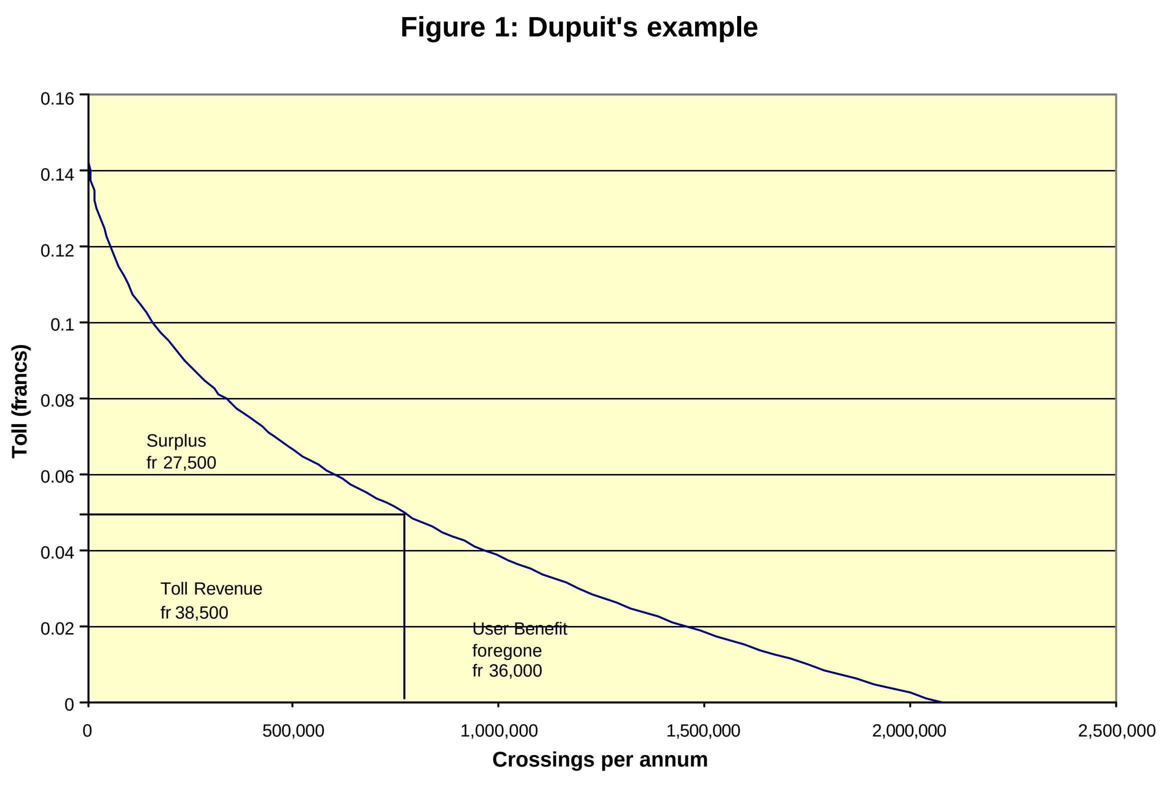
Total expenditure is given as:

$$E = T(P').P' \qquad (1.4)$$

and hence the Consumer Surplus is given as:

$$S = W - E \qquad (1.5)$$

Figure 1 Dupuit's example



Note that the integration can also be performed with respect to cost, by the general property that, if y = g(x) and x = g^{-1}(y), then

$$\int_{X_0}^{X_1} g(x).dx - Y_1.(X_1 - X_0) = \int_{Y_1}^{Y_0} g^{-1}(y).dy - X_0.(Y_0 - Y_1) \qquad (1.6)$$

Applying this rule allows us to express W as

$$W = \int_0^{T(P')} f^{-1}(T).\,dT \;=\; \int_{P'}^{\infty} f(P).\,dP - 0 + T(P').P' \;=\; \int_{P'}^{\infty} f(P).\,dP + E \quad (1.7)$$

Hence, when integrating with respect to price, we have an alternative formula for Consumer Surplus:

$$S = \int_{P'}^{\infty} f(P).\,dP \qquad (1.8)$$

Note that this represents the benefit of moving from a position in which there is no bridge (and so the price may be considered infinite) to a position where crossings are permitted for a payment of P'. More generally, we may consider the benefit of moving from a "base" position represented by a price of P_0 to a "policy" position represented by a price of P_1. By analogy, the benefit may be written as:

$$\Delta S = -\int_{P_0}^{P_1} f(P).\,dP \qquad (1.9)$$

"Consumer Surplus" in this form, as taken up later by Marshall (1920), assumes that the demand curves are essentially aggregate: they relate to the population as a whole, and it is implicit that there will be different "willingness-to-pay" among the population. If a monopolist had perfect information about the distribution of such willingness-to-pay, then he would be able to abstract the consumer surplus using a perfectly discriminatory pricing.

> "... the price which a person pays for a thing can never exceed, and seldom comes up to that which he would be willing to pay rather than go without it: so that the satisfaction which he gets from its purchase generally exceeds that which he gives up in paying away its price; and he thus derives from the purchase a surplus of satisfaction. The excess of the price which he would be willing to pay rather than go without the thing, over that which he actually does pay, is the economic measure of this surplus satisfaction. It may be called consumers' surplus." Marshall (1920) ch. 6

1.2 Subsequent Developments

The classic Marshallian demand response gives the change in demand for a commodity as a result of a price change for that commodity, assuming other prices remain fixed. Clearly, however, it will also be affected by **income**. The separation of demand response into price and income effects was originally set out by Slutzky (1915).

As we will discuss in more detail below, the essential observation was that a change in the price of a commodity would lead to a change in utility which could be viewed as a change in income.

More to the point, it would be possible to consider **compensating** the consumer for this implied change in income. If this were done, then any remaining shift in demand could be explained entirely by the price change. This led to the idea of "compensated demand" curves which are the cornerstone of current consumer theory.

A further important development was the generalisation to multiple commodities first proposed by Hotelling (1935, 1938). He showed that the one-dimensional integral for surplus S given in Eq (1.9) above could be generalised to a line integral allowing for simultaneous price changes.

Rather than discuss these important contributions in their own right, it is more straightforward to consider the "modern" theory of the consumer, as set out by, for example, Deaton & Muellbauer (1980), or Varian (1992), which codifies all this thinking into an elegant mathematical form.

While as noted the Marshallian treatment is essentially aggregate, from the point of view of development the modern standard theory deals with an individual consumer. There is then a further requirement to develop this to consider the welfare of society at large.

2. The standard microeconomic theory of the consumer

2.1 The Key Concepts

Based on a small number of axioms relating to preferences (most importantly, that of **transitivity**[1]), we can define an individual's "utility function", which is assumed to be derived from the commodities that he consumes. For convenience, these functions are usually assumed to be convex. Thus, if **x**, **y** and **z** are alternative vectors ("bundles") of commodities, and both **x** and **y** are "weakly preferred" to **z**, so that:

$$U(\mathbf{x}) \geq U(\mathbf{z}) \quad \text{and } U(\mathbf{y}) \geq U(\mathbf{z})$$

then $$U(\lambda\mathbf{x} + [1-\lambda]\,\mathbf{y}) \geq U(\mathbf{z}) \qquad \forall\ 0 < \lambda < 1$$

Further, if U is **strictly convex**, then the combination ($\lambda\mathbf{x} + [1-\lambda]\,\mathbf{y}$) will be strictly preferred to **z**, so that the "≥" sign becomes a strict inequality ">".

[1] This requires that if an individual prefers A to B, and prefers B to C, then he should prefer A to C

We now characterise the consumer as attempting to maximise available utility from expenditure. If **q** is a vector of commodities, with **p** the vector of prices, then with total available income Y, this can be written as:

$$\text{Max } U(\mathbf{q}) \text{ wrt } \mathbf{q} \text{ subj } \mathbf{p}.\mathbf{q} = Y \qquad (2.1)$$

and this gives:

$$\nabla U(\mathbf{q}^*) = \lambda\, \mathbf{p} \quad \Rightarrow \quad \mathbf{q}^* = \mathbf{g}(Y,\mathbf{p}) \qquad (2.2)$$

Marshallian demand

However, it has become standard to work with the **indirect** utility function which can be written as:

$$\psi(Y,\mathbf{p}) = U(\mathbf{g}(Y,\mathbf{p})) \qquad (2.3)$$

This represents the maximum utility that can be obtained, given income Y and price vector **p**.

Given indirect utility we use Roy's identity to obtain Marshallian demand:

$$\mathbf{g}(Y,\mathbf{p}) = -\nabla_p\psi\, /(\partial\psi/\partial Y) \qquad (2.4)$$

The derivative of ψ with respect to Y, usually denoted as λ, the "marginal utility of income", will be positive, and the second derivative can reasonably be expected to be negative ("declining marginal utility of money")[2].

As we noted, Consumers' surplus (CS) is conventionally defined in terms of an aggregate (Marshallian) demand curve, as the area under the (inverse) demand curve and above the current market price. Since as just shown a corresponding demand curve can be derived for the individual consumer, expressing the amount which the consumer would purchase at different prices, it is clearly possible to discuss CS at the individual level as well. However, as Deaton & Muellbauer (1980: §7.4) point out, the Marshallian concept of CS does not take into account the income (Slutzky) effects of a change in price. This can be seen as follows.

---

[2] However, since utility has no absolute value, these derivatives are not strictly measurable. The general theory of preference allows for any monotonic transformation of utility without affecting the results. In other words, any specific functional forms used are merely conveniences for mathematical tractability. Some economists have therefore argued that the marginal utility of income is essentially a meaningless concept (eg Deaton and Muellbauer, 1980 §5.3).

In a multi-dimensional context, the area under the (multi-commodity) demand curve and above the market price is given by (Hotelling, 1938) the line integral:

$$-\int \mathbf{g}(Y,\mathbf{p})\, d\mathbf{p} \qquad (2.5)$$

where the line integral is defined along a path between two positions $\mathbf{P}_0$ and $\mathbf{P}_1$, say. For this integral to be path independent, Green's theorem tells us that $\nabla\mathbf{g}$, the matrix of partial derivatives of the vector **g** with respect to the vector **p**, must be symmetric - in other words, for any two commodities r and s, we must have:

$$\frac{\partial g_r}{\partial p_s} = \frac{\partial g_s}{\partial p_r} \qquad (2.6)$$

We therefore need to investigate whether this condition (also referred to as the "Integrability Condition") holds for various demand functions. This can usefully be done by taking account of the Slutzky impact of price change on income.

2.2 The expenditure (cost) function and the Slutzky equation

In order to investigate this, we define the Dual problem:

minimise expenditure $E = \mathbf{p}.\mathbf{q}$ wrt $\mathbf{q}$ subj $U(\mathbf{q}) = V$ (2.7)

This gives $\mathbf{q^*} = \mathbf{h}(V,\mathbf{p})$ (2.8)

where **h** is termed the compensated (or Hicksian[3]) demand function. V is the maximum utility available given Y and **p**, and **h** shows how the demand varies with changing prices **p**, while remaining at a **fixed** level of utility.

By substituting **h** back, we get the so-called expenditure (or cost) function:

$$E(V,\mathbf{p}) = \mathbf{p}.\,\mathbf{h}(V,\mathbf{p}) \qquad (2.9)$$

This indicates the minimum expenditure required to maintain a constant utility V, in the face of changing prices **p**.

[3] after Hicks (1956)

E can be shown to be concave in **p**, and the derivative vector $\nabla_p E$ can be shown to be equal to **h**(V,**p**) (as long as the derivatives exist): see eg Deaton & Muellbauer, §2.3 for proof. Since this is so, the matrix

$$\mathbf{S} = \nabla_p \mathbf{h}(V,\mathbf{p}) = \nabla^2{}_p E(V,\mathbf{p}) \qquad (2.10)$$

is symmetric (because the order of partial derivatives is irrelevant), and by the concavity of E, it is also negative semi-definite.

For **given p**, we must have **h**(V,**p**) = **g**(Y,**p**) = **q***: ie, both the Marshallian and the Hicksian demand functions must predict the same quantity vector. Totally differentiating wrt **p** gives

$$\mathbf{S} = \nabla_p \mathbf{h}(V,\mathbf{p}) = \nabla_p \mathbf{g}(Y,\mathbf{p}) + \partial/\partial Y[\mathbf{g}(Y,\mathbf{p})].\ \nabla_p E(V,\mathbf{p})^T$$

$$= \nabla_p \mathbf{g}(Y,\mathbf{p}) + \partial/\partial Y[\mathbf{g}(Y,\mathbf{p})].\ \mathbf{h}^T \qquad (2.11)$$

This is the Slutzky matrix equation, usually re-arranged to give the derivative of the Marshallian demand function **g** as :

$$\nabla_p \mathbf{g}(Y,\mathbf{p}) = \mathbf{S} - \partial/\partial Y[\mathbf{g}(Y,\mathbf{p})].\ \mathbf{h}^T \qquad (2.12)$$

The first term is the "substitution effect", ie the change in demand resulting from the change in prices, **assuming constant utility**. The second term is the change in demand from a change in income, keeping prices constant, **times** the change in income to ensure constant utility when prices change, which, by the differential property of the expenditure function, is equal to the quantity demanded. It is easier to interpret the second term in terms of a change in price **dp**, leading to an income change **h**.**dp**. Then the change in demand **dg** can be expressed as:

$$\mathbf{dg} = \nabla_p \mathbf{g}(Y,\mathbf{p}).\mathbf{dp} = \mathbf{S}.\mathbf{dp} - \partial/\partial Y[\mathbf{g}(Y,\mathbf{p})].\ \mathbf{h}.\mathbf{dp} \qquad (2.13)$$

We know that **S** is symmetric. For the Marshallian integral to be path-independent, we require $\partial/\partial Y[\mathbf{g}(Y,\mathbf{p})].\ \mathbf{h}^T$ also to be symmetric, in other words:

$$\partial/\partial Y[q_r(Y,\mathbf{p})].\ q_s = \partial/\partial Y[q_s(Y,\mathbf{p})].\ q_r \qquad (2.14)$$

In general, this will **not** be the case.

Note in passing the effects of a monotonic transformation of utility. Suppose we redefine utility as W = f(U). Then it can be shown that while this will affect the functional form of the indirect utility and the expenditure functions, it will have no effect on the Marshallian demand function nor on the marginal rate of substitution between commodities.

2.3 An illustration

As a demonstration, which also illustrates the general properties of the theory, we can consider the "Linear Expenditure System" due to Stone (1954), which can be derived from a utility function of the form:

$$U(\mathbf{q}) = \prod_k (q_k - \gamma_k)^{\beta_k} : \quad \sum_k \beta_k = 1 \qquad (2.15)$$

Maximising with respect to **q**, given prices **p** and income Y gives:

$$0 = \nabla U(\mathbf{q}) - \lambda \mathbf{p} \Rightarrow \frac{U^*}{q^*_k - \gamma_k} \beta_k = \lambda\, p_k \;\; so \;\; q^*_k = \gamma_k + \frac{U^*}{\lambda} \frac{\beta_k}{p_k} \quad (2.16)$$

To obtain the Marshallian demand function, we need to make **q** dependent on Y rather than U*. To do this, we substitute **q*** in the budget:

$$\mathrm{Y} = \mathbf{p}.\mathbf{q}^* = \mathbf{p}.\mathrm{g} + \frac{U^*}{\lambda} \sum_k \beta_k \text{, whence } \frac{U^*}{\lambda} = (\mathrm{Y} - \mathbf{p}.\mathrm{g}) \quad (2.17)$$

Hence $g_k(\mathrm{Y}, \mathbf{p}) = \gamma_k + (\mathrm{Y} - \mathbf{p}.\mathrm{g}).\, \frac{\beta_k}{p_k}$ (2.18)

For the indirect utility function, we substitute **g** into the direct utility function, giving

$$\psi(\mathrm{Y}, \mathbf{p}) = \prod_k (g_k - \gamma_k)^{\beta_k} = \prod_k \left[(Y - \mathbf{p}\gamma) \frac{\beta_k}{p_k} \right]^{\beta_k} = (Y - \mathbf{p}\gamma) \frac{\beta_0}{\prod_k (p_k)^{\beta_k}} \qquad (2.19)$$

where for convenience we write β_0 for $\prod_k (\beta_k)^{\beta_k}$

For the current values of income and price, we have utility U* = ψ(Y, **p**). Since from (2.17) $\frac{U^*}{\lambda}$ = (Y – **p**.g), this allows us to solve for $\lambda = \frac{\beta_0}{\prod_k (p_k)^{\beta_k}}$ (2.20)

Hence, using (2.16), we obtain the Hicksian demand function:

$$h_k(U^*,\mathbf{p}) = \gamma_k + U^*.\frac{\Pi_k (p_k)^{\beta_k}}{\beta_0}\ \frac{\beta_k}{p_k} \qquad (2.21)$$

For the expenditure function we multiply **h** by **p**, yielding:

$$E(U^*,\mathbf{p}) = \mathbf{p}.g + U^*.\frac{\Pi_k (p_k)^{\beta_k}}{\beta_0} \qquad (2.22)$$

(since $\sum_k \beta_k = 1$)

Now consider the Slutzky equation (from (2.11) above):

$$\nabla_p \mathbf{g}(Y,\mathbf{p}) = \mathbf{S} - \partial/\partial Y[\mathbf{g}(Y,\mathbf{p})].\ \mathbf{h}^T$$

$\mathbf{S} = \nabla_p \mathbf{h}(U^*,\mathbf{p})$; hence $S_{kh} = U^*.\frac{\partial}{\partial p_h}\left[\frac{\Pi_k (p_k)^{\beta_k}}{\beta_0}\frac{\beta_k}{p_k}\right]$

$$= U^*.\frac{\Pi_k (p_k)^{\beta_k}}{\beta_0}\ \frac{\beta_k}{p_k}.\frac{\beta_h}{p_h}(1-\delta_{hk}/\beta_h) \qquad (2.23)$$

(where δ_{kh} is the 'Kronecker delta", with the property that $\delta_{kh} = 1$ if h = k, 0 otherwise).

For $\nabla_p \mathbf{g}(Y,\mathbf{p})$, we have $\frac{\partial}{\partial p_h}\left[\ \gamma_k + (Y - \mathbf{p}.\gamma).\ \frac{\beta_k}{p_k}\right]$

$$= -\frac{\beta_k}{p_k}\left(\gamma_h\ + \delta_{hk}.\ \frac{(Y - \mathbf{p}.\gamma)}{p_h}\right) \qquad (2.24)$$

We note that this is **not** symmetrical.

For $\partial/\partial Y[\mathbf{g}(Y,\mathbf{p})].\ \mathbf{h}^T$, we have $\frac{\partial}{\partial Y}\left[\ \gamma_k + (Y - \mathbf{p}.\gamma).\ \frac{\beta_k}{p_k}\right].\left(\gamma_h + U^*.\frac{\Pi_k (p_k)^{\beta_k}}{\beta_0}.\frac{\beta_h}{p_h}\right)$

$$= \frac{\beta_k}{p_k}.\left(\gamma_h + U^*.\frac{\Pi_k (p_k)^{\beta_k}}{\beta_0}.\frac{\beta_h}{p_h}\right) \qquad (2.25)$$

Adding the representative "rs" terms for $\nabla_p \mathbf{g}(Y,\mathbf{p})$ and $\partial/\partial Y[\mathbf{g}(Y,\mathbf{p})]$. $\mathbf{h}^T$, and remembering that $(Y - \mathbf{p}.g) = U^*.\frac{\Pi_k (p_k)^{\beta_k}}{\beta_0}$, we have:

$$\frac{\beta_k}{p_k}.\frac{\beta_h}{p_h}.U^*.\frac{\Pi_k (p_k)^{\beta_k}}{\beta_0}\left(1-\frac{\delta_{hk}}{p_h}\right) \qquad (2.26)$$

which is equal to the representative term for the Slutzky matrix **S**, as required.

Thus in this case, the condition for Green's theorem does not hold, and the consumer surplus defined as the integral of Marshallian demand **g** is **not** path-independent.

2.4 Separable functions of income and price

Suppose that we postulate an indirect utility function of the form:

$$\psi(Y,\mathbf{p}) = \lambda Y + f(\mathbf{p}) \qquad (2.27)$$

in other words, a separable function between income and prices, with the property that the derivative $\partial\psi/\partial Y$ is constant (and thus independent of **p**).

Using Roy's identity, we obtain $\mathbf{g} = -\nabla_p\psi/\lambda$.

We can also "invert" the indirect utility function to obtain the expenditure function:

$$E(U^*,\mathbf{p}) = 1/\lambda(U^* - f(\mathbf{p})) \qquad (2.28)$$

According to the theory (see p 7 above), we can differentiate this with respect to **p** to obtain the Hicksian demand function **h**, and this is easily seen to be $-\nabla_p\psi/\lambda$. But this is identical to the Marshallian demand function **g**! Thus we derive the important result that if the indirect utility has the separable form given above, so that the "marginal utility of income" is constant, the Marshallian and Hicksian demand functions coincide.

Since the Hicksian demand function satisfies the condition for Green's theorem, so, in this case, will the Marshallian demand function, so that the line integral for the Consumer Surplus is path-independent. This turns out to be of major importance for transport models, because most of the forms used are compatible with the separable form for indirect utility, as we discuss later.

Note that from Roy's identity given above, it can be seen that if $\lambda = \partial\psi/\partial Y$ is constant, then the CS integral $- \int \mathbf{g}(Y,\mathbf{p})\, d\mathbf{p}$ gives the indirect utility ψ, ie the maximum utility given income and prices. This explains the intuitive appeal of CS.

2.5 Compensating and Equivalent Variation

In terms of the overall theory, however, it is not **in general** reasonable to assume that λ is constant. To deal with this, two alternative measures have been proposed. The Compensating Variation (CV), proposed by Hicks (1956), is the minimum amount of income by which a consumer would have to be compensated after a price change in order to have the same indirect utility as before. In other words, if the prime (′) denotes the "after" position:

$$\psi(Y + CV, \mathbf{p}') = \psi (Y, \mathbf{p}) \qquad (2.28)$$

Note, of course, that CV may be positive or negative: it depends on the direction of the change in price. If prices go down, this will convey a benefit, and the consumer should be willing to give up some income to remain at the previous utility level.

The alternative definition, the Equivalent Variation (EV), is the maximum amount which the consumer, from the standpoint of the "after" position, would be willing to pay to have the price change reversed (assuming that income was unchanged). This implies:

$$\psi(Y - EV, \mathbf{p}) = \psi (Y, \mathbf{p}') \qquad (2.29)$$

Neither of these measures is the same as CS, except in very special circumstances. As Deaton & Muellbauer point out, the difference between CV and EV corresponds exactly to the difference between the Laspeyres and Paasche cost of living indices. In both cases we are dealing with alternative first order approximations, and the true measure is likely to be "somewhere in-between". Essentially, the need to approximate comes from the desire to transform a change in utility to a monetary value when, in general, there is no fixed conversion rate.

To derive the two measures, we formulate the Dual problem as before to obtain the Hicksian (compensated) demand curves. Because the Hicksian demand curves compensate for the impact of price on income, they are steeper (less elastic) than the corresponding Marshallian curve (strictly, this is only true if the commodities in question are “normal goods”, meaning that the elasticity of demand with respect to income is positive).

Figure 2 shows, for a single commodity, the two Hicksian demand curves associated with the before and after positions, as well as the (standard) Marshallian curve. The initial position is D_0

(Q_0 ,P_0) and the new position is D_1 (Q_1 ,P_1). The areas under the curves correspond with the three measures CS, CV and EV. It can be seen that the CS (shaded in the figure) does indeed lie somewhere in-between the other two measures (again, this depends on the qualification about "normal goods").

Note also that because the Hicksian demand curve **h**(V,**p**) is the derivative of the expenditure function with respect to **p**, we can represent the measures CV and EV in terms of the expenditure function:

$$CV = E(V^0, \mathbf{p}^1) - E(V^0, \mathbf{p}^0)$$

$$EV = E(V^1, \mathbf{p}^1) - E(V^1, \mathbf{p}^0) \qquad (2.30)$$

Figure 2 Compensating and Equivalent Variation

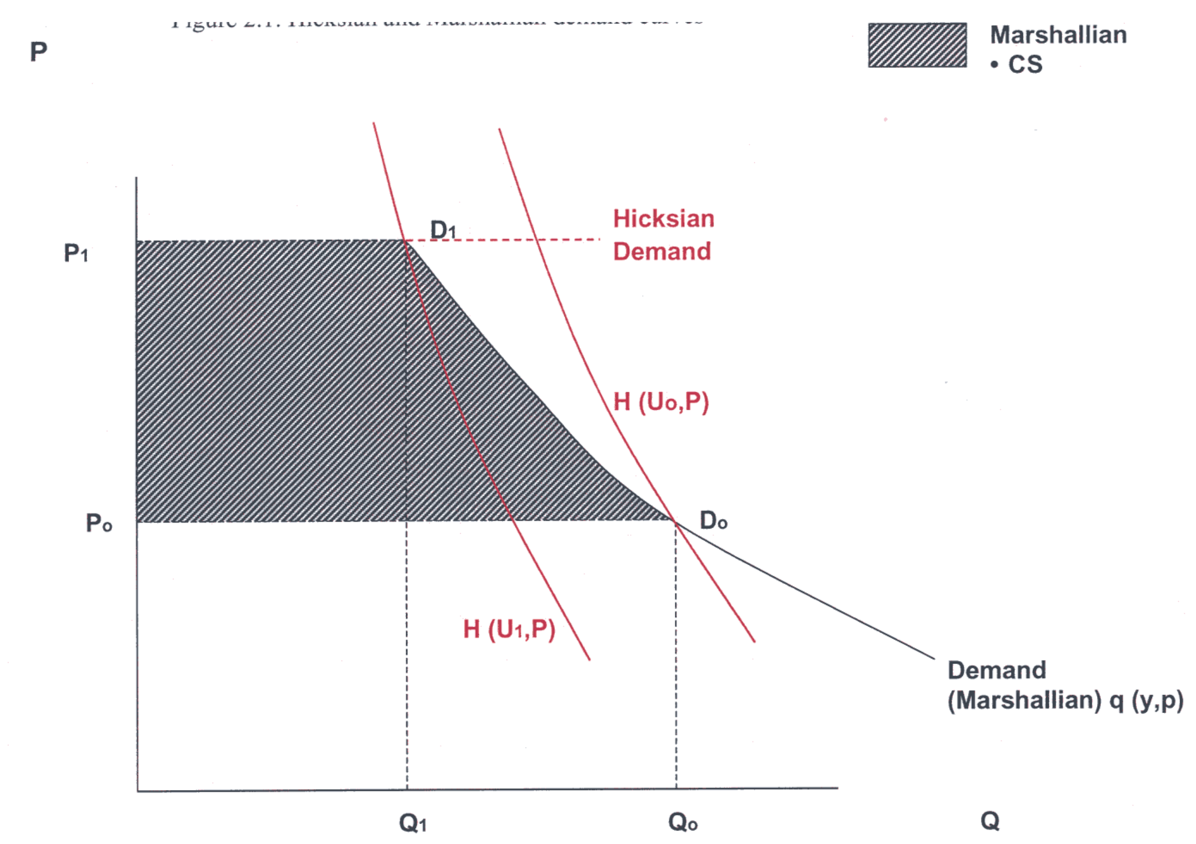


Figure 2 is the standard diagram used for theoretical exposition of the concepts CV and EV. In practical terms, however, it is somewhat misleading, on two separate counts. In the first place, the marked difference in the slopes of the Hicksian and Marshallian demand curves implies that a change in the commodity price will have a substantial “income effect”. But this will generally only be the case if a significant proportion of total expenditure is allocated to the commodity in question: for most instances of assessment in the transport field, this will not occur. Secondly, the shift **between** the two Hicksian demand curves depends on the magnitude of the price change, and once again, as we discuss later, standard assessment will be dealing with relatively marginal changes.

The result of these observations is that both the difference between the two “variation” measures, **and** the difference between them and the CS measure, are likely to be much less important than the Figure implies. Thus, in spite of the accepted theoretical superiority of the variation measures, it is not unreasonable to regard CS as a satisfactory measure. Indeed, some economists have argued more forcefully for it (Willig, 1976). In addition, CV and EV require a more detailed understanding of economic theory, and the Marshallian demand curve is an easier concept for the layman to understand.

In conventional economic terms, the argument is presented in terms of a **price** change. However, it is clear that the analysis can be taken directly across to the case of a quality improvement (or decline), which brings about a change in utility comparable to that caused by changing prices. This is a topic of particular interest for transport assessment, and we return to it later.

2.6 Aggregation and Welfare Measures

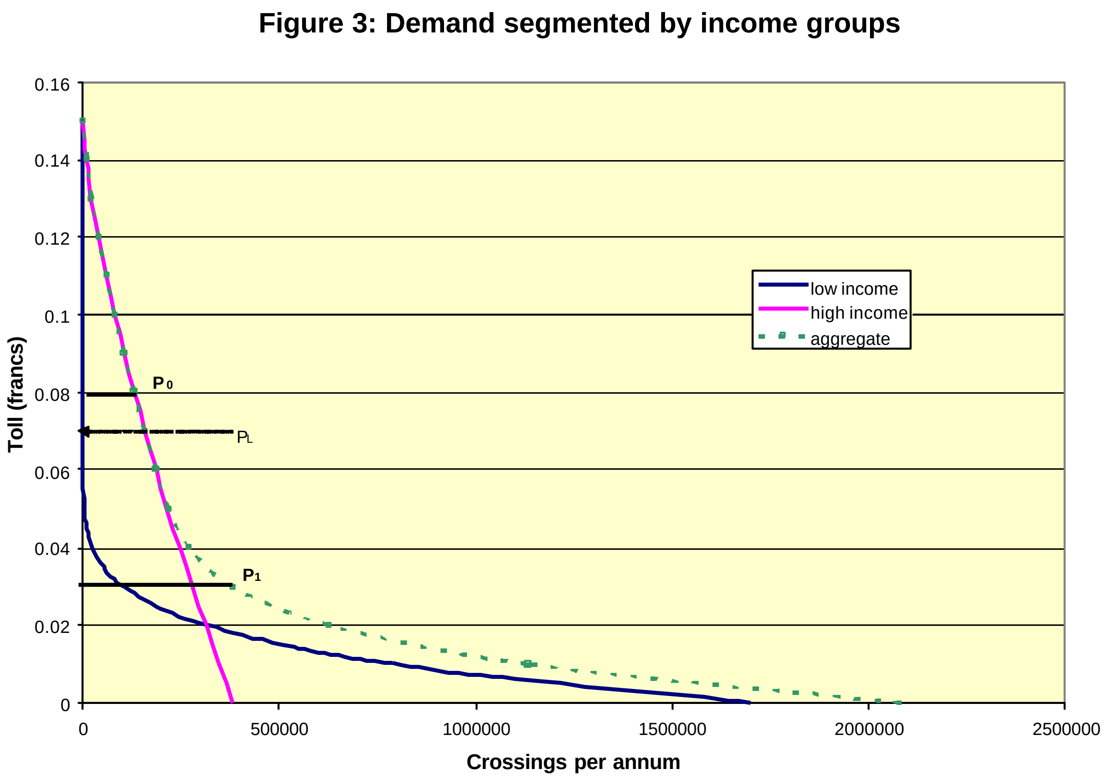
While the Consumer Surplus as proposed by Dupuit and Marshall was essentially conceived as an aggregate measure, the modern neo-classical theory of the consumer has been constructed at the individual level. **Assuming** that individual utility functions are known, this present no problems for demand modelling: the individual demand functions can be straightforwardly “added up”. In practice, of course, we do not know these utility functions, and any data which might allow us to estimate utility functions is likely to be subject to a certain amount of aggregation. Nonetheless, demand models which are compatible with the theory can be developed and tested empirically.

There are far greater problems relating to the construction of meaningful indices of social welfare. Indeed, there is a well-known theorem, due to Arrow (1951), which demonstrates the impossibility of basing social preference on individual preferences. The general ramifications of this are well beyond the territory of this paper. However, a simplified analysis will be presented, based on Varian (1992).

Consider first a modified version of the aggregate demand curve taken from Dupuit. We are interested in the impact of a price change from P_0 to P_1. In reality, different groups of individuals may have very different demand curves. In Figure 3, it is assumed that total demand can be split between those with high income and those with low income. Low income people will not purchase anything if the price rises above their maximum affordable price (P_L, say). Above this price, all variation in demand relates to high income people. Below this price, the total demand curve is obtained by summing the two curves along the Q-axis, and swings away from the high income curve as shown by the dashed line.

In the case illustrated, P_L lies between the original price P_0 and the new price P_1. The result is that the change in Consumer Surplus relates largely to high income people. The consumer surplus can be calculated separately for the two curves in the standard way: it must yield the same result as the area under the total curve since the total curve is obtained by summation.

Figure 3 Demand segmented by income groups



Suppose we postulate a "social welfare function" W, which is in some way related to the separate utilities of all the members of the society. Thus we write $W(U_1, U_2, \ldots.. U_n)$. Since, as already noted, utilities are only determined up to a monotonic transform, we must assume that we can arbitrarily fix the definition - separately for each individual. We assume for each individual q, that an increase in the utility for q will contribute positively to overall social welfare (though this is not entirely uncontroversial!): this implies that $\partial W/\partial U_q > 0 \ \forall$ q.

To simplify the problem, suppose that, at least for the purpose of a marginal analysis, the differentials $\partial W/\partial U_q$ can be approximated by a set of constant weights a_q. Effectively, we linearise the social welfare function so that

$$W(U_1, U_2 . U_n) = \Sigma_q \, a_q.U_q. \qquad (2.31)$$

The weights $\{a_q\}$ can be thought of as the value judgments of a "social planner".

Suppose we are at a market equilibrium (**q**, **p**) and are considering moving to a new allocation **q′**, where **q** ranges over all individuals q as well as over all commodities.

Approximately, $\Delta W \approx \Sigma_q \, a_q.\nabla U_q(\mathbf{q}_q).(\mathbf{q'}_q - \mathbf{q}_q) \qquad (2.32)$

Since the base is in equilibrium, we must have: $\nabla U_q(\mathbf{q}_q) = \lambda_q \, \mathbf{p}$, where, as usual, the Lagrangean multipliers λ_q represent the marginal utility of income for individual q.

Hence, $\Delta W \approx \Sigma_q \, a_q. \, \lambda_q \, \mathbf{p}. \, (\mathbf{q'}_q - \mathbf{q}_q) \qquad (2.33)$

This is equivalent to a weighted change of expenditures.

Further, **if** the original allocation was a welfare optimum, then it can be shown that $a_q = 1/\lambda_q$. In this case, therefore, the change in welfare resulting from a change in allocation is given by the change in total expenditure (calculated at the base prices), which implies correspondingly a change in total income.

To quote Varian, "*This means that if the social planner consistently follows a policy of maximising welfare both with respect to lump sum income distribution and with respect to other policy choices that affect allocations, then the policy choices that affect allocations can be valued independently of the effect on the income distribution*". The trouble with this is that it means precisely what it says! Thus if the condition is not met, then there is **no** implication that the effect on the income distribution can be ignored.

It turns out that there are simplified conditions in which more accessible results can be obtained. These restricted cases will have correspondingly restricted application, but they can be acceptable approximations for a range of practical problems. A particularly fruitful assumption is that of the "quasi-linear" utility function, defined as:

$$\tilde{U}(\mathbf{q}, a) = a + U(\mathbf{q}) \qquad (2.34)$$

where a is the consumption of a particular commodity not included in **q**, with, for convenience, a price of 1.

By the standard procedure we have:

$$\max \tilde{U} + \lambda(Y - a - \mathbf{p}.\mathbf{q}), \text{ giving } \lambda = 1 \text{ (wrt a) and } \nabla U = \mathbf{p}$$

This implies that $\mathbf{q} = \mathbf{g}(\mathbf{p})$, **independent** of Y. The demand for the particular commodity is given from the budget constraint:

$$a = Y - \mathbf{p}.\, \mathbf{g}(\mathbf{p})$$

Hence, substituting into utility function, we obtain

$$\psi(\mathbf{p},Y) = Y - \mathbf{p}.\, \mathbf{g}(\mathbf{p}) + U(\mathbf{g}(\mathbf{p})) \qquad (2.35)$$

Note that we can obtain $U(\mathbf{g}(\mathbf{p}))$ by direct integration as $\int_{0\to\mathbf{g}(\mathbf{p})} \mathbf{p}.d\mathbf{q}$, because of the absence of income effects.

Gorman (1953, 1959) investigated the conditions under which it was possible to prescribe an "aggregate" or "representative" utility function which would lead to an aggregate demand function. He concluded that the only type of indirect utility function which would have the required property was of the form:

$$\psi_q(\mathbf{p},Y_q) = a_q(\mathbf{p}) + b(\mathbf{p}).Y_q \qquad (2.36)$$

for each individual q

Under this assumption, the individual demand function becomes (Roy's identity):

$$\mathbf{q}_q = \mathbf{g}_q(\mathbf{p},Y_q) = -(\nabla a_q + Y_q.\, \nabla b)/b(\mathbf{p}) \qquad (2.37)$$

and the aggregate demand function is therefore

$$\mathbf{q} = \Sigma_q\, \mathbf{q}_q = -\Sigma_q\, (\nabla a_q + Y_q.\nabla b)/b(\mathbf{p})$$

$$= -[\Sigma_q\, \nabla a_q + \nabla b.\, \Sigma_q\, Y_q]/b(\mathbf{p}) = \mathbf{g}(\mathbf{p}, \Sigma_q\, Y_q) \qquad (2.38)$$

It is clear that this aggregate demand function could be generated by a "representative consumer" with an aggregate income of $\Sigma_q Y_q$

It can be seen that the indirect utility function corresponding to the quasi-linear utility is of the Gorman form, in which (adding subscripts for individual q)

$$a_q(\mathbf{p}) = U_q(\mathbf{g}_q(\mathbf{p})) - \mathbf{p}.\, \mathbf{g}_q(\mathbf{p}) \text{ and } b(\mathbf{p}) = 1 \qquad (2.39)$$

It can also be seen that $a_q(\mathbf{p})$ is the Consumer Surplus CS for individual q.

Hence, associated with the quasi-linear utility, there is an aggregate form for the indirect utility if the "representative consumer" from which the aggregate demand curve can be derived. The functional form is:

$$\psi\,(\mathbf{p}, \Sigma_q Y_q) = a(\mathbf{p}) + \Sigma_q Y_q \qquad (2.40)$$

where the term $a(\mathbf{p})$ corresponds with the sum of the individual CS, and this is also the CS that would be derived from integration of the aggregate demand curve. Thus, in the special case of a quasi-linear utility function, the aggregate CS is an appropriate measure of aggregate welfare. It can also readily be shown that in this case CS coincides with the two "variation" measures, since the marginal utility with respect to income is constant..

2.7 Contributions from Discrete Choice Theory

The primary contribution here is from McFadden's path-breaking though relatively inaccessible theoretical analysis (McFadden, 1981), itself a major elaboration of the discussion in Domencich & McFadden (1975, reprinted 1997). Further theoretical points are made in Small & Rosen (1981), while a useful description is provided in Glaister (1981).

McFadden begins by extending the continuous analysis given earlier to allow for discrete alternatives. Suppose that in addition to the consumption of commodities $\mathbf{q}$, there is a set of discrete alternatives $\{i \in \mathbf{I}\}$ with the characteristic that one and only one will be chosen by an individual consumer. The characteristics of alternative i are represented by a vector $\mathbf{x}_i$, and the cost is c_i.

The discrete choice formulation leads to the notion of **conditional** indirect utility depending on the alternative actually consumed. If the individual chooses alternative i, then only $Y - c_i$ is available as the budget for the consumption of commodities $\mathbf{q}$. With this modification, the analysis goes through as before for the derivation of the conditional indirect utility functions

$\psi_i(Y - c_i, \mathbf{p})$. Since the total utility will reflect the additional utility from the consumption of the discrete alternative, we write it as $\psi_i(Y - c_i, \mathbf{p}, \mathbf{x}_i)$.

On this basis, the alternative which will in fact be chosen will be that which affords the maximum indirect utility, ie alternative i such that $\psi_i(Y - c_i, \mathbf{p}, \mathbf{x}_i) > \psi_j(Y - c_j, \mathbf{p}, \mathbf{x}_j) \ \forall \ j \neq i$. Hence the unconditional indirect utility can be written as:

$$\psi^*(Y, \mathbf{c}, \mathbf{p}, \mathbf{X}) = \max_i \psi_i(Y - c_i, \mathbf{p}, \mathbf{x}_i) \qquad (2.41)$$

In order to effect a generalisation to a population rather than a single individual, we introduce a random element to the individual's utility. There are a number of ways in which this can be motivated. In Domencich & McFadden we read (§4.3):

> "In principle, the theory of individual utility maximization provides a complete model of individual choice. However, within the framework of economic rationality and the postulated structure of utility maximization, there will be unobserved characteristics, such as tastes and unmeasured attributes of alternatives, which vary over the population. These variations may induce variations in observed choice among individuals facing the same measured alternatives. A specification of a distribution for the unobserved factors then generates a distribution of choices in the population.
>
> To clarify the conceptual issues involved in this construction, we consider the textbook model of economic consumer behavior. The individual has a utility function $u = U(\mathbf{x}, \mathbf{s}, e)$, representing tastes, where $\mathbf{x}$ is the vector of observed attributes of an alternative, $\mathbf{s}$ is a vector of observed socioeconomic characteristics, such as sex, education, and age, and e is a vector of unobserved characteristics of alternatives and unobserved factors, such as intelligence, experience, childhood training and other variables determining tastes. The utility function is maximized subject to a "budget constraint" $\mathbf{x} \in \mathbf{B}$ at a value $\mathbf{x}$ given by a system of demand functions,
>
> $$\mathbf{x} = \mathbf{h}(\mathbf{B}, \mathbf{s}; e). \qquad (4.1)$$
>
> The econometrician typically observes the budget constraint B_n, socio-economic characteristics s_n and chosen alternative x_n for a cross-section of consumers n = 1, ..., N. He wishes to test hypotheses about the behavioral model (4.1)."

McFadden (1981) expands on this:

> (§5.2) "The idea of taste variation in a population influencing aggregate demand behaviour is an old one. Many of the classical demand studies [refs…] consider this as a nuisance to be eliminated by assumption…. More recently analysis of econometric models with random parameters has been motivated by the presence of unobserved variations among economic agents."

An individual's utility function U is defined on the vectors $\mathbf{q}$ (consumption of non-discrete commodities) and $\mathbf{x}$ (attributes of discrete alternatives). To indicate the individual-specific nature of the resulting indirect utility functions, McFadden (1981) includes U as an argument, thus:

$\psi_i(Y - c_i, \mathbf{p}, \mathbf{x}_i: U)$ [this is a variant of McFadden's eq (5.5)]. He then treats U as a random element, conditional on population (socio-economic) characteristics **s**.

If now we **suppose** ψ_i has the form $\dfrac{Y - c_i - \alpha(\mathbf{p},\mathbf{x}_i : U)}{\beta(\mathbf{p})}$ [this is a variant of McFadden's eq (5.12)], it is readily seen that this is of the Gorman form. McFadden then considers the function

$$\overline{V} = E_{U|\mathbf{s}}[\psi^*] = \frac{Y}{\beta(\mathbf{p})} + E_{U|\mathbf{s}}\left[max_i \frac{-c_i - \alpha(\mathbf{p},\mathbf{x}_i : U)}{\beta(\mathbf{p})} \right] \qquad (2.42)$$

and shows, with rather sparse explanation, that a) this has the characteristics of an indirect utility function, and b) the choice probabilities P_j can be obtained by applying Roy's identity to $\overline{V}$ - ie, P_i is the negative ratio of the partial derivatives of $\overline{V}$ with respect to c_i and Y. As can be seen, $\overline{V}$ is the expectation over the population variation in U of the maximum, over the alternatives, of the conditional indirect utilities. Given a valid social indirect utility function, "*the demand distribution can be analysed as if it were generated by a population with common tastes, with each (representative) consumer having fractional consumption rates for the discrete alternatives…*"

Ignoring the scale factor $\beta(\mathbf{p})$, which merely ensures that function α is scaled in money units, it can be seen, by analogy with the continuous exposition given earlier for the quasi-linear utility, that the term $E_{U|\mathbf{s}}[max_i(-c_i - \alpha(\mathbf{p},\mathbf{x}_i : U))]$, which McFadden writes (allowing for notational changes) as $G(\mathbf{c},\mathbf{p},\mathbf{x},\mathbf{B},\mathbf{s})$, where **B** is the set of alternatives, has the characteristics of a "surplus function". Further, since by assumption the marginal utility of income is constant, P_j is also given by $-\partial G/\partial c_j$. This is the critical property for what McFadden defines as the "AIRUM" (Additive Income Random Utility Maximizing) form.

In more detail, he defines the conditions in which G will be termed a social surplus function, and concludes that "the presence of discrete choice places no new restrictions on the validity of consumer surplus methods". This needs to be seen, however, in the context of the assumed simplification associated with AIRUM. As Small & Rosen (1981) point out, there **are** some additional complications associated with discrete choice, due essentially to local discontinuity of the derivatives of the expenditure (or cost) function, which is the standard way of deriving the compensated demand functions. This issue does not arise in the simpler case to which the Gorman form applies.

A key point is made at the end of §5.7: "*It should be noted that the utility structure (5.12) yields choice probabilities that are independent of current income. However, tastes (the distribution of*

U) may depend on individual characteristics that are correlates of current income such as historical wage rates, income levels, or occupation. Then these variables may enter the PCS [Probabilistic Choice System]".

This is a potential source of confusion, and we will discuss it later in the paper.

Before leaving this section, we will briefly allude to some of the issues faced in the general discrete choice case for defining welfare measures, following some of the arguments of Small & Rosen (1981), McFadden (1999) and Karlström (1998, 2000).

Suppose individual k faces a choice from a set of alternatives $j \in J$, and has a conditional indirect utility function for each alternative of the form:

$$V_{kj}(p_j, \ldots.) + \varepsilon_{kj}$$

To keep the illustration simple, assume that the price of only one alternative changes, and without loss of generality take this as the first: ie $p_1 \to p'_1$. Again for the purposes of illustration, we assume that this is an increase.

It will be seen that the compensating variation will depend on the individual's choice in the before and after situations. If individual k chooses alternative 1 both before and after the price change, then his reduction in utility[4] is given as

$$V_{kj}(p_j, \ldots.) - V_{kj}(p'_j, \ldots.). \qquad (2.43)$$

This then needs to be converted into the equivalent amount of income, which we write as μ_{11}. This is straightforward if the marginal utility of income is constant, but in other cases may be more complex to evaluate (see Section 4.?): nonetheless the relationship will be monotonic.

Correspondingly, if individual k does not choose 1 before the price change (increase), he will certainly not choose it after, and therefore is unaffected by the price change. In this case, the compensating variation is zero.

As Karlström (2000) notes, the difficult case is when the individual chooses alternative 1 before the price change but a different alternative j after the price change. The implication is that:

$$V_{k1}(p_1, \ldots.) + \varepsilon_{k1} \geq V_{kj}(p_j, \ldots.) + \varepsilon_{kj} \geq V_{k1}(p'_1, \ldots.) + \varepsilon_{k1} \qquad (2.44)$$

---

[4] This assumes, as would generally be reasonable, that the random element ε_{k1} is not affected by the price change

so that the utility loss is lower, and correspondingly the compensation will be less than μ_{11}. In this case, however, the reduction in utility is stochastic, because of the elements $\varepsilon_{k1}, \varepsilon_{kj}$: in fact it is given as

$$[V_{k1}(p_1, \ldots.) + \varepsilon_{k1}] - [V_{kj}(p_j, \ldots.) + \varepsilon_{kj}] \qquad (2.45).$$

In addition, the distribution of the difference $(\varepsilon_{k1} - \varepsilon_{kj})$ is truncated, since we know that it cannot be less than $[V_{kj}(p_j, \ldots.) - V_{k1}(p_1, \ldots.)]$, otherwise alternative 1 would not have been chosen before the price increase.

For this case, therefore, we can only calculate the **expected** compensation, based on the appropriate distribution for the error terms.

This argument shows that we can calculate the (expected) compensation, conditional on the choices made in the before and after situations. Overall, we require the unconditional compensation, and this will be in some sense a weighted average of the conditional values. As indicated earlier, this calculation turns out to be straightforward when dealing with an AIRUM form: it is much less straightforward for other cases, as we discuss in Section 4.

2.8 The evaluation of non-price changes

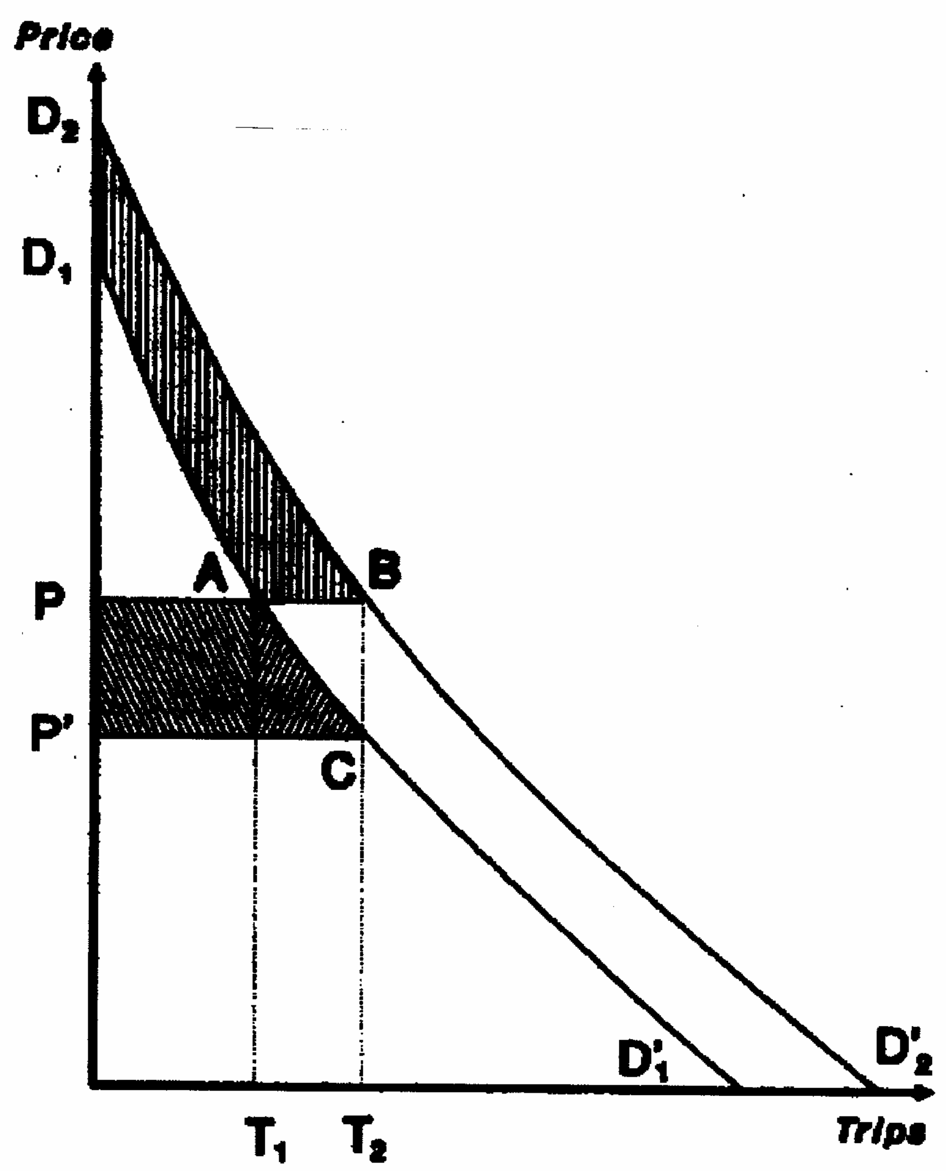
So far, both the continuous and the discrete exposition have been entirely in terms of **price changes**: the concern with the demand response and the welfare implications of changing prices of defined commodities or alternatives. This is, indeed, the standard application of the theory. However, particularly in the field of transport, most policies are concerned with other aspects of the travel experience. Foremost among such aspects is the travel time associated with the journey, but other aspects, relating to comfort, reliability etc. are also of interest.

Figure 4 below, based on the EVA Manual (1991), shows the impact of a quality change. A standard demand curve is given (D^1) together with a price reduction from P to P′ which increases demand and generates consumer surplus in the usual way. To quote from the EVA Manual:

> [The diagram] shows what happens when instead of a price change, we have a quality change, such as a change in the travel time. The demand curve shifts outward (in the case of a time saving), so that the trips at the basic price P (unchanged) increase from T_1 to T_2. The increase in consumer surplus is given by ABD_2D_1. The same effect on the number of trips would have been achieved by reducing price from P to P′, without altering the travel time. The measure of consumer would then be PACP′.
>
> PACP′ is not necessarily equal to ABD_2D_1. (It will be if the demand curve shifts in a way parallel to the original curve.) But it is a reasonable approximation for practical purposes to

make the two equal. In this case, it is possible, and usual, to redefine the demand curve as a function of a linear combination of price and quality variables such as travel time, known as user cost, generalized cost or sometimes disutility. This curve would be drawn in exactly the same way …, but with the vertical axis measuring user cost instead of price.

Figure 4 Change in Demand for Travel through Improved Quality



Reinterpreting this in Discrete Choice terms, if we have a conditional indirect utility function which is **linear** in price, we can effect a direct conversion between quality variables in the utility function and price: effectively this is an extension of the "generalised cost" concept introduced by McIntosh & Quarmby (1970). If price does **not** enter linearly, then, although we can still calculate benefit in units of utility, there will be problems converting these to money units.

3 THE PRACTICE OF EVALUATION

3.1 The practical context

The role of supply and demand are crucial in transport modelling. In line with the theoretical discussion, we may characterise a model of demand as one which estimates what travel would take place, **given** an estimate of travel costs for all possible journeys, where by costs we refer not only to money costs but other components of travel "utility" such as travel time: generalised cost.

However, if, at a given generalised cost, the predicted travel were actually realised, the costs might not stay constant. This is the function of the **supply** model.

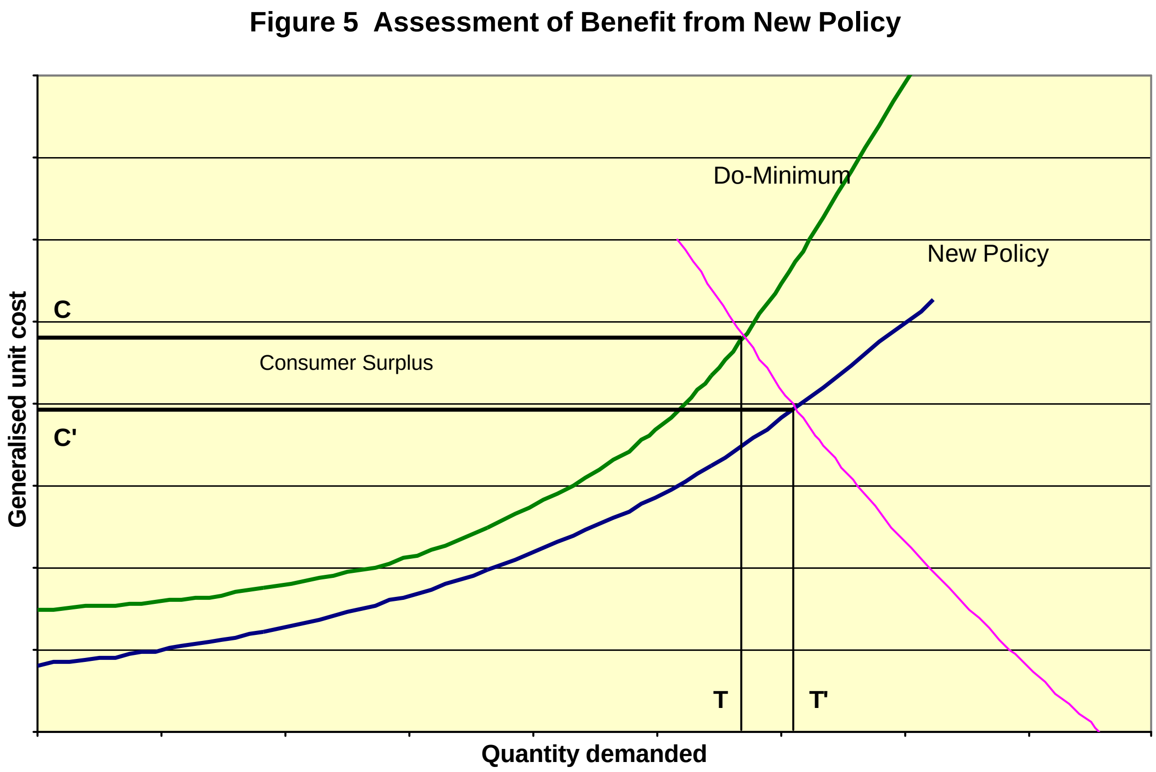
In classical economics the supply curve gives the quantity T which would be produced, given a market cost C. However, in transport it is more convenient to define the inverse relationship, whereby C is the unit generalised cost of meeting a demand T.

Thus, we use the supply curve to answer the question: what would the generalised cost be if the estimated demand were "loaded" on to the system? The best known 'supply' model is the conventional traffic assignment reflecting *inter alia* the deterioration in highway speeds as traffic volumes rise.

Moreover, in terms of transport policy, our most frequent interest is in changing the supply curve (eg by providing new capacity, or modifying prices). In this case, we are interested in the comparison of two or more alternative supply curves, representing different policies. Very often, we wish to compare the effects of a policy with the "Do-Minimum" option.

Since both demand and supply curves relate volume of travel with generalised cost, the actual volume of travel must be where the two curves cross - the 'equilibrium' point. Using the principles of **Consumer Surplus**, we measure the area under the demand curve between the two equilibrium points, (T, C) and (T′, C′): this is illustrated in Figure 5. The consumer surplus is the area bounded above by the line “generalised cost = C”, below by the line “generalised cost = C′ ”, and to the right by the (downward-sloping) demand curve.

Figure 5 Asessment of Benefit from New Policy



3.2 The “Rule of a Half”

By assuming as an approximation that the demand curve is linear between the two equilibrium points, we obtain the well-known “Rule of a half” [RoH] expression of benefit:

$$\text{Benefit} = \Delta S \approx -\tfrac{1}{2}\,(T' + T)(C' - C) \qquad (3.1)$$

(NB: As a general rule, it is highly advisable to adopt a notational convention which clearly distinguishes the points ("before" and "after"), and we will consistently use the prime symbol (′) to denote the "after" position. We will also, as in this formula, maintain the order "after / before", making use of minus signs where necessary, and we will define the difference operator Δ to mean "after minus before".)

This implies that to carry out an appraisal, we require only the equilibrium demand with and without the policy to be tested, and the components of cost compatible with the two demands.

An intuitive rationale for this measure is as follows. Existing travellers (T) obtain the full value of the reduction in C. New travellers, on the other hand, (T'-T), are assumed to get on average only half the benefit, since it is argued that while some of them were, in the 'before' situation, on the verge of travelling and therefore get almost the full benefit, others in the 'after' situation are on the verge of not travelling, and therefore get almost zero benefit.

However, this straightforward rationale loses its simplicity as soon as we consider the demand for more than one type of journey. Within general transport evaluation we cannot confine ourselves to journeys between a single pair of zones, by a single mode, etc., because of the interconnectedness of the transport system. For instance, if we introduce a new link into a road network, the demand for travel on some links may fall, while for others it increases. There may be a redistribution of travel between origins and destinations, and new road traffic may be attracted (either from other modes or 'pure' generated traffic). The graphical representation of this interconnectedness quickly becomes impossibly complex.

In addition, the simple example suggests that we can distinguish between 'existing' and 'new' travellers, thereby introducing an element of asymmetry between the before and after situation. As soon as we consider the demand for more than one type of journey, this can quickly lead to confusion. It also implies that the benefits can be unambiguously allocated to different groups of travellers: this is in fact far from being the case, and while we may from time to time wish to make such an attempt, it is important to bear in mind that the ultimate attribution of benefits remains controversial. We return to this point below.

Hence, although the figure treats transport as a one-dimensional commodity, when we consider the problem in a general transport context, we are not interested in single elements of demand, but rather a matrix of elements: the domain of the demand model is essentially the **i-j pair** ie between an origin and destination.

Additionally, the transport problem is complicated by the supply domain being that of a network of **links**, while the demand for travel relates to the inherent value of being at j, given a current location at i, and not to the particular paths used to reach j. Changes to the generalised cost of a single network link will typically impact on a number of i-j pairs.

This means that the one-dimensional formula needs to be generalised to the case where the costs of all transport alternatives can change simultaneously. Assuming base costs **C** and changed costs **C′**, the resulting formula is, as discussed earlier, the (Hotelling) line integral

$$\Delta S = -\oint_{\mathbf{C}\to\mathbf{C}'} \mathbf{T}.\mathrm{d}\mathbf{C} \qquad (3.2)$$

Provided that the demand function **T**(**C**) derives from a model which is consistent with AIRUM, this line integral will be path independent.

Although in most cases of transport appraisal there will be an explicit demand function **T**(**C**), the problems of calculating the multi-dimensional integral are potentially serious. Fortunately, however, the linear approximation (RoH) form is also appropriate in this case, along the lines of:

$$\text{Benefit} = \Delta S \approx -\tfrac{1}{2}\,\Sigma_\xi\,(T'_\xi + T_\xi)(C'_\xi - C_\xi) \qquad (3.3)$$

where ξ indicates members of the set of transport choices. For the purposes of illustration, we shall consider a model which allows ξ to range over origin i, destination j, mode m and time of day t, in which case the benefit formula becomes:

$$\Delta S \approx -\tfrac{1}{2}\,\Sigma_i\,\Sigma_j\,\Sigma_m\,\Sigma_t\,(T'_{ijmt} + T_{ijmt})(C'_{ijmt} - C_{ijmt}) \qquad (3.4)$$

3.3 The accuracy of the RoH Approximation

It can be shown that the RoH is a very good approximation to the true surplus provided the change in cost can be regarded as "marginal". Suppose we have a choice model for the proportion p_i choosing alternative i of the logit type:

$$p_j = \exp(-?\,C_j) \,/\, S_k \exp(-?\,C_k) \qquad (3.5)$$

with a total demand given by T.

We wish to consider the change in consumer surplus from a policy which changes the values of $\{C_k\}$ to $\{C'_k\}$. As is well known, with a logit model, there is a closed form solution for the integral under the demand curve, and in this case the formula for benefit becomes

$$?S = -T\,(C^{*\prime} - C^*) \qquad (3.6)$$

where C* is the so-called "composite cost" (the negative of the maximum expected utility) defined as the "logsum":

$$C^* = -1/?\ \ln S_k \exp(-?\,C_k) \qquad (3.7)$$

We wish to compare the true results using (3.6) with the approximation formula in (3.8):

$$\Delta S \approx -\tfrac{1}{2}\, S_k\, (T'_k + T_k).(C'_k - C_k) \qquad (3.8)$$

A simple example will suffice. Suppose we have five choices k. Total demand T is 1000, and we assume ? = .02, which is a typical value for choices ranging over mode and destination when cost is measured in **minutes**. The results are set out in the Table below:

Table 1

	Base		Strategy		ΔS		
Option k	costs C_k	demand T_k	costs C'_k	demand T'_k	RoH calculations	approx	True
1	20	225	15	237	–½(237+225).(15–20)	1155	
2	25	204	22	206	–½(206+204).(22 –25)	615	
3	45	137	35	159	–½(159+137).(35– 45)	1480	
4	15	249	18	223	–½(223+249).(18 –15)	–708	
5	30	185	30	175	–½(175+185).(30 - 30)	0	
* (Σ)	–54.48	1000	–57.02	1000	1000. (–54.48 + 57.02)	*2542*	2539

There are a number of things to note about this example. In the first place, the RoH approximation involves summing over what may appear to be elements of benefit calculated separately for each alternative. Note that cells in the RoH formula where the costs do not change (eg option 5 above) do not contribute to total benefit, even though the demand has changed. There is no equivalent calculation for the true integral result: the benefit integral is not decomposable.

Secondly, it will be seen that the composite costs are **negative** (in this instance). This causes some presentational difficulty, and some practitioners have been reluctant to accept the measure for this reason[5]. However, it is easily shown that it does not influence the evaluation outcome: it is only the **difference** between the before and after situations which is material.

Thirdly, according to the exact formula, consumer surplus has increased by 2539, whereas the RoH formula gives 2542, which differs from the true value by 0.1%. Note that, given the fundamental convexity of the demand curve, the RoH will always give an **over**estimate of benefit: as this example shows, however, the error is (normally) very small, with the exception of some pathological cases which we shall investigate below.

As a further example, suppose now further that total demand is elastic, and responds to the change in composite cost. Specifically, write

$$T = T_0 \exp(-aC^*) \text{ where } a = 0.01$$

Then, with the same changes in cost, because these result in a change in C* of -2.539, T will increase to 1025.72, and hence the values of the individual cells after the change are increased (the changed cells are indicated in **bold)**:

[5] Various proposals have been made to normalise the measure. It can be shown that if all the individual costs are positive, the addition of γ/λ to the composite cost, where γ is Euler's constant (= 0.577216…), will generally correct for this, but not in all cases. For further discussion see eg Williams (1977)

Table 2

	Base		Strategy				
Option k	costs C_k	demand T_k	costs C'_k	demand T'_k	ΔS	approx	integral
				elastic			
1	20	225	15	**243**	**–½(243+225). (15–20)**	**1170**	
2	25	204	22	**211**	**–½(211+204). (22 –25)**	**622**	
3	45	137	35	**163**	**–½(163+137). (35– 45)**	**1500**	
4	15	249	18	**229**	**–½(229+249). (18 –15)**	**–717**	
5	30	185	30	**180**	**–½(180+185). (30 - 30)**	**0**	
* (Σ)	–54.48	1000	–57.02	**1026**	**(integral - see below)**	***2576***	**2572**

With such a simple total demand curve, we can easily derive the total change in consumer surplus analytically - it is

$$- \int_{C^*}^{C'^*} T_0 \, e^{-aC} \, dC \; = \; 1/a \, . [T_0 \, e^{-aC}]_{C^*}^{C'^*} \qquad (3.9)$$

Substituting, we obtain an increase in consumer surplus of 2572, whereas the rule of a half approximation gives 2576, which differs from the true value by 0.2%.

Quite generally, the "composite costs" obtained from a (possibly hierarchical) logit model, appropriately scaled, represent a measure of (negative) consumer surplus per trip, and the change in composite cost is an indicator of benefit. If the total demand at some level is fixed at T, then the benefit is given by $\Delta S = -T.(C^{*\prime} - C^*) = -T.\Delta C^*$, where C* is the composite cost calculated at the level of T.

Where demand is not constant at **any** level within the model, the simple RoH approximation can be used: $B \approx -\frac{1}{2} (T' + T).\ \Delta C^*$. In practice, it is usually acceptable to apply RoH to the **composite** cost at a lower level, and we discuss this further in the following section.

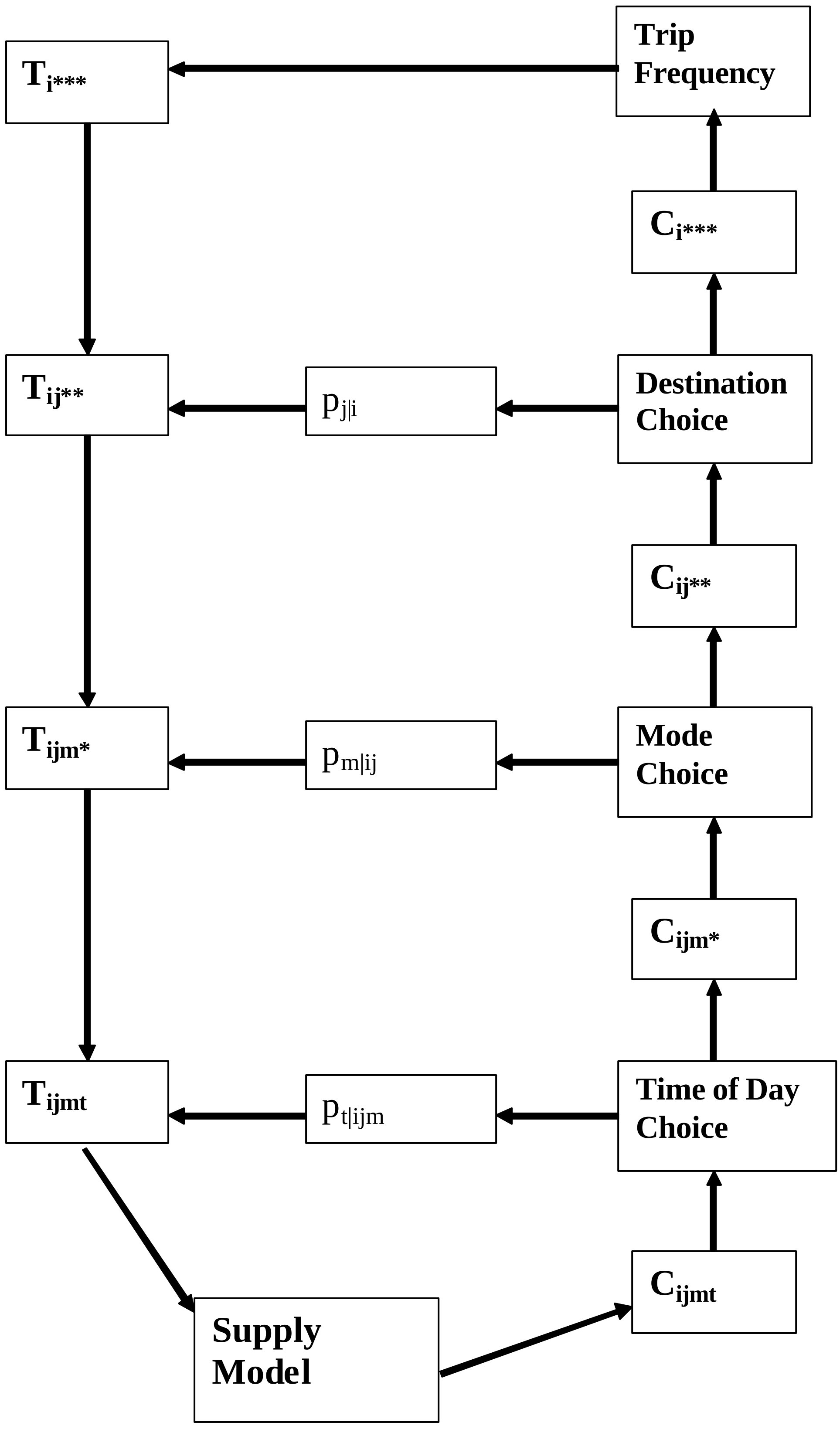
Although the "composite cost" methodology is specific to a demand formulation relying on hierarchical logit, this is generally in line with most modelling practice, though the burgeoning interest in more complex error structures (in particular, the "mixed logit" formulation) is likely to change this in the foreseeable future. It may be noted, however, that closed form solutions are available for all members of the Generalised Extreme Value (GEV) family of random utility models.

3.3 Applying the RoH at different levels in the hierarchy

In the case of a hierarchical logit model, whether we **can** appropriately apply the RoH at any particular level relates essentially to the magnitude of the term $\lambda.\Delta C$, where ΔC is the change in cost, and λ is the "scaling parameter" for the given level. If it is "large" (which for practical purposes we might define as > 3), then the RoH will start to present problems. However, because the parameters λ decline in absolute magnitude as we go to higher levels, the problems of inaccuracy in the RoH can generally be avoided by carrying out the approximation at a sufficiently high level. We shall produce some specific examples below.

It is useful to take a hierarchical logit choice model as a reasonably general example. For the purposes of illustration only, we will assume a structure in which time of day is conditional on mode, mode is conditional on destination, and destination is conditional on origin, as illustrated in Figure 6. Although this is an intuitively reasonable structure, the results will be indicative of **any** such structure, and indeed of more complex assumptions about the error structure of the various choices.

Figure 6 Illustrative Hierarchy of Discrete Choices



At the bottom level we have a demand model of the form:

$$p_{t|ijm} = \exp(-?^{T} C_{ijmt}) \Sigma_{t'} \exp(-?^{T} C_{ijmt'}) \qquad (3.10)$$

with an associated composite cost over t, which we write as C_{ijm*}, given, as usual, by the formula:

$$\exp(-?^{T} C_{ijm*}) = \Sigma_{t'} \exp(-?^{T} C_{ijmt'}) \qquad (3.11)$$

Corresponding relationships apply to the higher level choice models, with a set of scaling parameters which we write as $?^{M}$, $?^{D}$, $?^{O}$, where the superscripts M D and O refer, respectively, to mode, destination and origin choices. Note that with the hierarchy illustrated, it would be a structural requirement that $?^{T} \geq ?^{M} \geq ?^{D} \geq ?^{O}$.

Then, the following set of calculations will generally all give approximately equivalent results for ΔS:

$$-\tfrac{1}{2} S_{ijmt} (T'_{ijmt} + T_{ijmt}).(C'_{ijmt} - C_{ijmt}) \qquad (3.12a)$$

$$-\tfrac{1}{2} S_{ijm} (T'_{ijm*} + T_{ijm*}).(C'_{ijm*} - C_{ijm*}) \qquad (3.12b)$$

$$-\tfrac{1}{2} S_{ij} (T'_{ij**} + T_{ij**}).(C'_{ij**} - C_{ij**}) \qquad (3.12c)$$

$$-\tfrac{1}{2} S_{i} (T'_{i***} + T_{i***}).(C'_{i***} - C_{i***}) \qquad (3.12d)$$

Furthermore, if total demand T_{****} is fixed, then all these estimates will be a good approximation to the true integral result: $T_{****} (C'_{****} - C_{****})$.

3.4 Link-based formulae

Although the theory is perfectly general, in practice it has most often been applied in the context of highway appraisal. As indicated earlier, this leads to some interface issues with the link-based nature of the supply network. In this section, we discuss some of the relevant issues.

Suppose we have a demand matrix T_{ij}, a set of links {l}, and appropriate formulae for deducing the components of link cost, based on various fixed items (eg link capacity). In the more general case, certain components, typically link travel time, will depend on the flow on the link V_l, (and possibly flows on other links).

An assignment model will require a criterion for choosing the best path(s) between i and j: a standard criterion is to minimise "generalised cost" . This implies a formula for combining the components of link cost into a value C_l, which might be, for example $C_l = f(d_l, t_l, ...)$ where t and d are respectively the time and distance on the link.

We write the proportion of those choosing a particular path p as $\pi_{p|ij}$. If we define a "link-path incidence matrix" (see eg Bell & Iida (1997)) as δ_{pl} with the value 1 if link l is on path p, 0 otherwise, then we can combine these to create:

$$e_{ijl} = \Sigma_p \pi_{p|ij} . \delta_{pl} \qquad (3.13)$$

and we can work on a link basis by means of the set $\{e_{ijl}\}$ interpreted as "the **proportion** of the total travel between i and j T_{ij} which uses link l". [6]

As is well understood, the quantity e_{ijl} is used to produce the link "loads" Q_l, using the formula

$$Q_l = S_i\, S_j\, T_{ij}\, e_{ijl} \qquad (3.14)$$

The same quantity will also provide the matrix of average generalised cost between zones i and j: this can be written as

$$C_{ij} = S_l\, e_{ijl}\, C_l \qquad (3.15)$$

If a Wardrop equilibrium has been assumed for the assignment, **and** the process has adequately converged, then it should be the case that this average generalised cost will be the same as the minimum, since all allocated paths should have the same cost according to the equilibrium conditions.

Applying the RoH at the O-D level, we have

$$\Delta S = -\tfrac{1}{2}\, S_{ij}\, (T'_{ij} + T_{ij}).(C'_{ij} - C_{ij}) \qquad (3.16)$$

Substituting for C_{ij}, this becomes:

$$\Delta S = -\tfrac{1}{2}\, S_{ij}\, (T'_{ij} + T_{ij}).(\, S_l\, e'_{ijl}\, C'_l - S_l\, e_{ijl}\, C_l\,)$$

$$= -\tfrac{1}{2}\, S_l\, [\, S_{ij}\, (T'_{ij} + T_{ij}).(\, e'_{ijl}\, C'_l - e_{ijl}\, C_l\,)] \qquad (3.17)$$

As is well-known, if the matrix is **fixed**, then the benefit can be obtained by calculating the change in the product of the flow and cost on each **link**, including any new links brought about as a result of the scheme, and summing over all links. This approach has generally been favoured in the UK, because the information can be directly derived from the assignment model.

Since in this case $T'_{ij} \equiv T_{ij}$, the formula becomes:

[6] In the simpler case of a single best path ("all or nothing" assignment), e_{ijl} = 1 if link l lies on the best path between i and j, and = 0 otherwise

$$\Delta S = - S_l \, [\, S_{ij}\, (T_{ij}).(\, e'_{ijl}\, C'_l - e_{ijl}\, C_l\,)] = - S_l \, [\; Q'_l\, C'_l - Q_l\, C_l\,] \qquad (3.18)$$

However, a link-based benefit calculation is not legitimate with "variable" matrices, since there are "cross-product" terms involving the "before" matrix assigned to the "after" paths etc.. Although some assignment routines may be able to carry out such procedures, there is no inherent advantage in proceeding on a link basis.

Note also that even if demand is constant at a "higher" level (eg when summing over time of day, or modes, or destinations), cross-product terms will still occur. Hence, the benefits can only be calculated on a link basis when there is no change in demand at the level at which the paths are determined.

3.5 Some difficulties with the RoH

As noted, in transport evaluation we are generally concerned with the changes in generalised cost, and their impact on demand. However, the form of the composite cost implicitly involves other components of utility, and in particular the alternative-specific constants which are often estimated as the "intrinsic" utility associated with particular alternatives.

The RoH approximation is commensurate with the assumption that the benefit of switching between alternatives is related only to the cost changes associated with the alternatives, and can ignore the **underlying** attractiveness of the alternatives, since this does not change. However, the validity is critically dependent on the scale of the cost change relative to the random process assumed to underlie the choice process.

It is of interest to discuss this in relation to time-period choice, where we may argue that the implicit attractiveness of travelling at one period rather than another remains constant, but variations in relative generalised cost could bring about a shift in demand.

Suppose we have two periods 1 and 2, with respective generalised costs C_1, C_2. We assume a simple binary logit model, and translate these into utility units by means of the scale parameter $-\lambda$. As is well known, λ (≥ 0) is inversely related to the standard deviation of the random component of utility.

In addition, we may postulate that there are "schedule utilities", which connote the inherent advantage of travelling in the given periods. Assume that their average values, scaled to cost units, are α_1, α_2. Hence travellers facing a choice between periods 1 and 2 have average utilities:

$$V_1 = \lambda\, (\alpha_1 - C_1) : \qquad V_2 = \lambda\, (\alpha_2 - C_2) \qquad (3.19)$$

Now consider a transport improvement whereby the generalised cost in period 2 reduces from C^0_2 to C'_2. We do not expect any changes in the average schedule utilities α_1, α_2. As a result of the change $V_2 \rightarrow V'_2 > V^0_2$.

As usual, we can measure the benefit of this change as the scaled increase in composite utility (ie the expected maximum utility from the choice set). Using the logit function, this gives

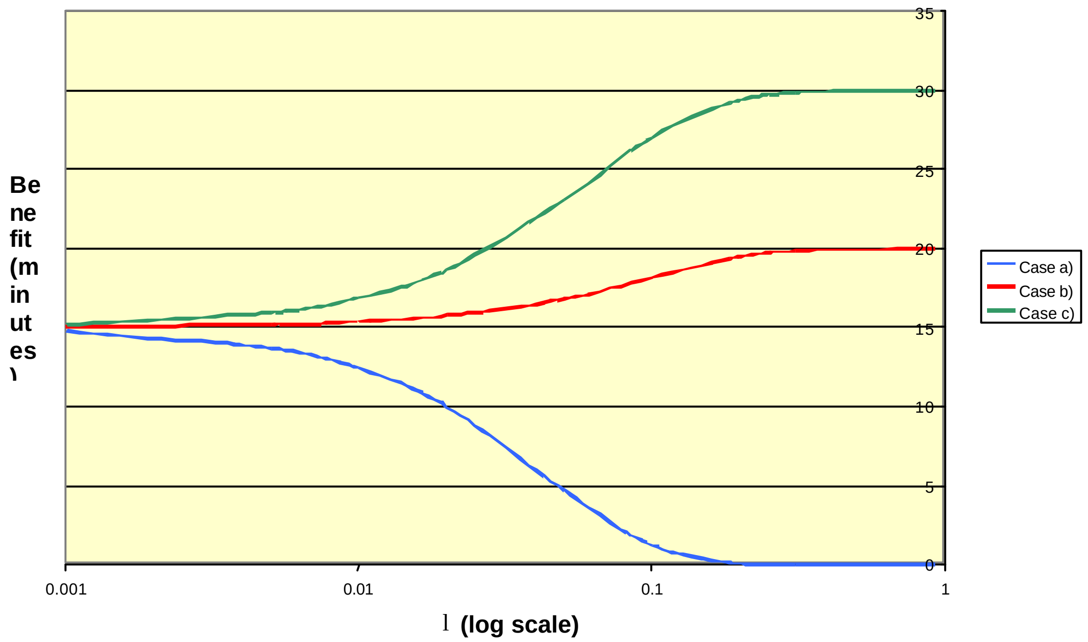
$$B = 1/\lambda\ [\ln(\exp V_1 + \exp V'_2) - \ln(\exp V_l + \exp V^0_2)] \qquad (3.20)$$

The conditions under which the RoH approximation will hold are essentially dependent on the size of $[(V_1 - V'_2) - (V_l - V^0_2)]$, ie $(V^0_2 - V'_2)$, relative to the standard deviation of the random element in utility. If $V^0_2 - V'_2$ is relatively small, then the approximation will hold, and it is generally possible to measure the benefit from a time shift without knowing the schedule disutility.

However, it is reasonable in this case to expect that the randomness will be related to the size of the time period, particularly if it relates largely to the schedule utility. As the benefit approximation breaks down, either because the cost change is too large, or because the random variance is too small, we are no longer able to avoid taking account of schedule utility, either in the evaluation or in the modelling.

We illustrate this in Figure 7, using the example given. Measure cost in units of minutes, and assume that the improvement in period 2 is equivalent to 30 minutes (ie $C^0_2 - C'_2 = 30$). Consider three base cases, where it is sufficient merely to define the difference in utilities between the two time periods: a) $V_l - V^0_2 = 50\lambda$; b) $V_l - V^0_2 = 10\lambda$; c) $V_l - V^0_2 = -10\lambda$. The Figure shows the estimate of benefit for different levels of randomness, measured by λ.

Figure 7 Effect of scale parameter on Benefit



It can be seen that for all three base cases, the benefits tend to 15 (ie half the improvement) for small values of λ, but for larger values the outcome depends strongly on the base case. In case a), the improvement is insufficient to overcome the superiority of period 1, so the benefit tends to zero. In case b), the improvement converts period 2 to the superior alternative: hence the benefit tends to the value of $V'_2 - V_1$, ie 30 –10 = 20. In case c), period 2 was the superior alternative in any case, and the benefit tends to the full value of the improvement: 30 minutes.

The generalisation of the result for high λ is

$$\lim_{\lambda \to \infty} \Delta C_* = \min_{i=1,2} (C'_i - a_i) - \min_{i=1,2} (C_i^0 - a_i) \qquad (3.20)$$

To re-state the conclusions, the Rule of a Half (RoH) approximation implies that the benefit of switching between alternatives is related only to the cost changes associated with the alternatives. Where the RoH is not appropriate, we cannot ignore the constant components associated with the implicit attractiveness of specific alternatives (in this example, the scheduling benefits brought about by travelling at different times).

It is of some interest to relate these general conclusions to two well known circumstances in which the RoH is inappropriate: it will be seen that they are both manifestations of the conditions just described.

The first case is where the set of available choices is different between the Do-Minimum and Do-Something, and the most commonly encountered example within general urban transport appraisal is the problem of **new modes**, where the "before" cost of the new mode is effectively infinite. Thus this is an example where the change in cost is too large, relative to the random component, for the RoH to be valid. We discuss this further in Section 4.2

The second case is where the choice between two options can be considered to be made entirely on deterministic grounds, as is commonly assumed for the choice of **route**. Suppose, as in Section 3.4, we have a demand matrix T_{ij}, a set of links {l}, and appropriate formulae for deducing the components of link cost. Assume that there is a finite set of possible paths $\{P_{ij}\}$ between each origin and destination, and that the proportion of those choosing a particular path p as $\pi_{p|ij}$. Combining the "link-path incidence matrix" δ_{pl} with the link costs, we obtain the path costs:

$$C_{ijp} = \Sigma_l C_l . \delta_{pl} \qquad (3.21)$$

Because the choice of paths is deterministic, we cannot calculate the RoH at this level. In other words,

$$\Delta S\ ^1 - ½\ S_{ij}\ S_p\ [(\pi'_{p|ij}.T'_{ij} + \pi_{p|ij}.T_{ij}).\ S_l\ \delta_{pl}\,(C'_l - C_l\,)] \qquad (3.22)$$

We may reasonably ask why the RoH is not valid in this case, given that it **is** typically legitimate when a total travel matrix is segmented by **mode**. To give a simple example, if paths 1 and 2 have costs 20 and 30 respectively in the base, and 18 and 15 respectively in the "after" situation, then the benefit **cannot** be calculated as

$$½\ [\ Q\ (20\text{-}18) + Q'\ (30\text{-}15)\ ]$$

where Q is total flow on the best path in the before situation, and Q′ is total flow on the best path in the after situation. The cost of path 2 in the base is **irrelevant**, as is the cost of path 1 in the after situation. Effectively, the evaluation has to be done at a higher level, so that the correct application of RoH is given as

$$½\ [(Q^0 + Q')(20\text{-}15)] \qquad (3.23)$$

This second case has come about because the choice between alternatives is assumed to be **non-random** (corresponding to the case of a logit model with scale parameter equal to ∞). In this case, the costs of inferior alternatives are irrelevant. It may be noted that while All-or-Nothing assignment is an obvious case, the conclusion applies to (Deterministic) Equilibrium assignment as well, in which all paths actually used must have the same minimum cost (for a given "user group").

3.6 Partitioning the benefit

Having established the principle that the “true” surplus based on the demand curve is (generally) acceptably approximated by the RoH, it then becomes attractive to “decompose” the surplus, in various ways. The most informative decomposition is by the components of “utility” or generalised cost.

As an illustration, assume generalized cost (C) is expressed as a linear combination of money cost (c) plus time required (t). Thus

$$C = a.c + ß.t \qquad (3.24)$$

In this form, the units are arbitrary, and it is conventional to set one of the coefficients a or ß to unity, while maintaining the ratio v = ß/a, where v is the value of time. If a is set to one, the generalized cost is in money units, while if ß is set to one, it is in time units. Of course, within the demand model itself, this distinction is neutralized by the choice of scaling parameter (though there is an issue for **forecasting** - see Gunn (1983))

The outcome is that the typical terms $(T'_? + T_?).(C'_? - C_?) = (T'_? + T_?).\ \Delta C_?$ in the RoH formula decompose to:

$$a.\ (T'_? + T_?).\ \Delta c_? +\ ß.\ (T'_? + T_?).\ \Delta t_?$$

so that we can assess how much of the benefit accrues in money savings and how much in time savings. This decomposition into generalised cost components is not possible when using the exact (composite cost) formulation. Hence, for general reasons of clarity and disaggregation, there is an interest in carrying out the calculation at the **lowest** level in the hierarchy where the generalised costs are **defined** and hence, of course, the cost is not composite. From what was said in the previous section it will be appreciated that it is, nonetheless, necessary to ensure that the level. used is compatible with the validity of the RoH approximation.

The process of making the components of surplus explicit also opens up a further possibility, that of weighting the components in a way which is believed to represent a **social** evaluation. This has become a widespread practice, particularly in relation to values of time. In other words, regardless of the way in which that the modeller has combined the components in the **demand** model, for evaluation purposes there is a presumption that a different, typically standard, weight should be used. The implications are wide, and will be discussed in Section 3.8.

In considering what further disaggregations of the surplus are appropriate and useful, we need to reflect on the scope of an Economic Evaluation, and the circumstances under which Cost-Benefit Analysis (CBA) is normally applied. The rationale for CBA is to assist in decision making where market forces do not apply or are inappropriate. Corporate bodies that can be assumed to follow market rules (eg oil companies) can be omitted from the CBA, since they are assumed to be able to adjust to any change in demand etc. through normal market processes. If, for example, bus operators are considered to be acting under normal market conditions, then they are outside the scope of CBA, and all costs **and** benefits for this mode must be excluded.

There is room for changes in definition here, but it is probably necessary to recognize the following key "sectors": on the one hand, the "users" of the transport system (the people and goods that move around), and on the other those parties who are involved in the supply, regulation and financing of the system (essentially this consists of the transport operators, the parking (and possibly a toll) authority, and finally Government). We use the term Government in the widest sense, in its potential role as provider of highways, health services, subsidy, etc.. In traditional terminology, these other parties are referred to as "non-travellers".

According to the scope of the Economic Evaluation, the key output is the **net** value of all costs and benefits of the relevant parties. The question then arises as to what should be done about so-called "transfer payments" between parties within the framework. Should these be explicitly recognised, or should they be ignored, on the basis that they will net out in the final evaluation?

Within the field of UK urban transport evaluation in the 1980s, a practice developed whereby travellers' money benefits tended to be ignored in the cost-benefit analysis of public transport schemes. In particular, the result of a change in fares was argued to be that travellers incur a loss, or gain, which is exactly compensated for by an increase, or reduction, in operators' revenue. In such a case, it appears unnecessary to take this element of benefit into account. A similar argument has been raised in respect of the tax elements in, for example, fuel prices.

However, this simplification is in fact only valid in restricted cases. In the general case it can be shown that there are money benefits to travellers who change their behaviour which do **not** cancel out "on both sides". There is in any case a more important principle involved: the benefits to travellers should represent **all** the benefits associated with a given transport proposal. If a large amount of the benefits are subsequently cancelled out by corresponding elements on the cost side (non-travellers), this should of course be reflected in the final balance. Since fares changes should be treated on a consistent basis with any other transport changes, it is essential to define benefits in a way that does not prevent this. And we reiterate the point made earlier, that such practice is in line with the true (integral) benefit formula, which does not indicate how money benefits should be distinguished.

This approach also applies to the treatment of taxation, where the "transfer payments" are between travellers and the Government. A pound saved in petrol represents a pound saved to the traveller, regardless of the fact that much of the cost of petrol represents fuel tax. However, in the final cost-benefit calculus, the corresponding loss to the Government needs to be offset against the money benefit enjoyed by the traveller. The recommendation is therefore that both elements should be distinguished and reported explicitly, rather than netted out from the start.

Strictly speaking, it is also necessary to make allowance for the different **incidence** of elements of indirect taxation, such as VAT, fuel duty etc.: it is also conventional to make different assumptions according to whether the travel is considered to be “leisure” (including the journey to work) or “employers’ business”). In practice, this is a marginal correction, and likely to be well inside the error margin of the calculations. Nevertheless, it has become conventional practice to make the correction.

A correct treatment of taxation requires a clear definition of units and an analysis of the flow of resources between sectors. A reasonable approach is to measure all benefits and costs **net** of any indirect taxation which applies **outside** the scope of the evaluation (ie, in the remaining part of the normal market economy). For non-private sector travellers (ie travel made on behalf of corporate bodies) the correction can be ignored, since (most) indirect taxes can be reclaimed. It is therefore only necessary to adjust the benefits accruing to non-business travel to take account of the average level of indirect taxation in the rest of the economy.

In passing, we should note that, inasfar as values of time are usually based directly on willingness to pay calculations, these will be expressed in terms of "normal market economy” units of currency, and therefore will also need to be adjusted, if the convention just set out is adopted[7].

A final form of partitioning is to return to the original arguments in Section 1 whereby S can be decomposed into the elements W (total willingness to pay) and E (actual expenditure). Hence, using the RoH approximation, the change in surplus ΔS can be written as:

[7] In this respect, the UK Department for Transport has recently changed its practice. Previously benefits were calculated in “factor costs”, with indirect taxation netted out: this is consistent with the convention outlined here. However, since 1999, all benefits are given in market prices.

$$\Delta S \approx + \tfrac{1}{2}\,(T' - T)(C' + C) - [T'.C' \; - T\,.C] \qquad (3.25)$$

$$(\Delta W) \qquad - \qquad (\Delta E)$$

Note that only the first term is an approximation: the change in expenditure can be exactly calculated from the demand model outputs.

In setting out the overall incidence of money flows, this form of partitioning can be useful to make explicit the nature of "transfer payments". In the context of current UK practice, it has an additional advantage in relation to "cost misperception". It has been traditional to argue that car travellers do not "perceive" the non-fuel costs of car operation (eg tyres, maintenance, wear and tear etc), and therefore these costs should not enter the "willingness to pay" calculations. Nevertheless, a change in expenditure will be experienced, and may also have tax implications, so that it cannot be ignored. The current UK methodology is equivalent to omitting the non-fuel costs from the ΔW calculation but including them in the ΔE calculation.

3.7 The attribution of benefits

There are numerous cautions within the literature about the disaggregation of the overall consumer surplus measure. Thus, for example, Jones (1977) says: It may be noted that for presentational purposes the over-all benefit measure can be disaggregated by types of benefit, although these types of benefit should not be taken as a final measure of incidence". The general principle is that while disaggregations of the total benefit may be indicative, only the total is theoretically unambiguous.

The question about 'final incidence' referred to in the quotation from Jones is one of the reasons for this caution. Transport investment may bring about a number of changes -for example, in property prices -which are not directly related to travellers. In the first UK Value of Time report (MVA et al, 1987), it is suggested that the reduction of overall transport costs may to a considerable extent represent windfall gains to existing owners of land, given the relative inflexibility of supply and demand in the urban housing market (para 4.13.15).

Clearly, the more a given change in transport cost can be confined, the more reasonable it is to attribute benefits to specific groups. For instance, if a particular road was restricted in its access to certain groups, then it would be reasonable to attribute the benefit of an improvement to those groups. Correspondingly, the more we can clearly identify the users of a improved facility after the improvement with those before the improvement, the more confident we can be in attributing the benefits.

The problems emerge when an improvement results in a substantial change in travel patterns. As soon as we have more than two travel options, whether discussing mode, destination, route or whatever, the logical attribution of benefit to 'changed' travellers depends on their previous behaviour. A simple example of this is given in the Appendix, which demonstrates the basic point that while the overall benefit calculation is independent of the details by which individuals change between the before and after situation, the attribution of benefits to particular groups of 'changers' is not.

To this must be added the observation that even where current models provide a mechanism for considering what kind of changes take place, very little credibility can be attached to such estimates: the models provide estimates of T_{ijm} and T'_{ijm} but they do not attempt to describe the details by which individuals or groups move between cells of these two matrices. Indeed, given the evidence from panel surveys about the high degree of day-to-day variability in people's travel behaviour, it would not be reasonable for them to do so.

A strong caveat must therefore be made against the disaggregation of benefits whenever there is a significant proportion of travellers changing behaviour. What appears **less** objectionable is to disaggregate the benefit by the **source** of the saving. Thus, a reduction in generalized cost on a particular i-j-m link gives rise to an identifiable amount of benefit, using the standard RoH formula. This precise amount of benefit does not, however, necessarily accrue to the final users of the link. It clearly also fails to take account of benefits generated elsewhere due, for example, to the relief of congestion on other links.

Despite all the above, it is accepted that there will be many instances when attempts will be made to ‘trace the beneficiaries'. In particular, there will be a presumption that the benefits in respect of a particular link or set of links do accrue to the users of those links. We can only repeat the caveat that the extent to which this is acceptable depends crucially on the amount of changing between different parts of the system.

3.8 “Re-engineering” the weights of the components of utility

As noted, in the practice of evaluation it is often decided to pre-specify the weights of components, regardless of the assumptions made in the demand model. While there may be compelling “political” reasons for doing this, it does lead to potential incompatibilities. In particular, changing the weights interferes with the composite cost calculation. Since, more or less by definition, the composite cost is compatible with, and encapsulates, the demand changes among the alternatives to which it relates, re-calculating the composite cost *post hoc* with the revised weights is **not** an option.

A particular case relates to the value of time, which is one of the most "visible" items in the evaluation procedure. The adoption of "standard value" of non-working time[8] for cost-benefit analysis has been a feature of UK appraisal practice since the 1960s and a similar approach is found in a number of other countries. Two arguments have traditionally been advanced for using a standard value.

- in principle, the same values for non-working time savings on all locations and modes should be applied, irrespective of the willingness to pay of the particular group of consumers who get the benefits;

- using a single standard value is a practical procedure to follow given the difficulty of acquiring relevant market information (incomes etc.) on which case-specific values would need to be based.

This practice has been widely criticised. For example, Sugden (1999) called for an end to the use of the standard value of non-working time on the grounds that it is "incompatible with the logic of CBA". It is useful to set out the argument in a formal way, drawing on work by Mackie *et al* (2001).

Return to the social welfare function $W = W(U_1, U_2 \ldots U_q)$ discussed in Section 2.6. Now consider a change in a particular travel opportunity whereby both time and cost change, by Δt and Δc. This results in a change in utilities ΔU_q for each q and hence a change in overall welfare ΔW, given by

$$\Delta W \approx \Sigma_q \frac{\partial W}{\partial U_q} \Delta U_q = \Sigma_q \Omega_q \Delta U_q \qquad (3.26)$$

where Ω_q are the relative weights attached to the utility of the different groups q.

For small changes, it is acceptable to linearise the utility function so that, with an implied value of time $V_q = \alpha_q/\lambda_q$:

$$\Delta U_q = \alpha_q \Delta t + \lambda_q \Delta c \qquad (3.27)$$

Hence combining (3.26) and (3.27),

$$\Delta W = \Sigma_q \Omega_q (\alpha_q \Delta t + \lambda_q \Delta c) \qquad (3.28)$$

[8] for time savings in **working time**, a different approach is normally used, based on the value to the employer as deduced from marginal productivity theory. This value is sometimes modified to account for other aspects of working conditions while travelling, following the arguments of Hensher (1977).

Note in passing that this implies that the values of time V_q are a separate matter from the set of social weights to use, Ω_q. Although in practice these have often been run together, there is no reason in principle to do so. Moreover, the choice of welfare weights should come as a matter of cross-sectoral Government policy, whereas the value of travel time savings will be a transport specific matter. There are therefore attractions in keeping them separate.

Now, let us consider some interesting cases. If we assume that $\Omega_q = 1/\lambda_q$, we imply that a unit change in income bears equally on all q. In this case, the social benefit is given as the sum of individual willingness to pay for benefits.

$$\Delta W = \Sigma_q (V_q \Delta t + \Delta c) \tag{3.29}$$

This is the Harberger approach to cost-benefit analysis – unweighted adding up of willingness to pay (wtp). Arguments for this are of the following kind:

- It is what happens with normal market commodities in a commercial appraisal context, and in particular, it is how revenues and costs are typically treated in transport appraisal;
- If the existing income distribution is considered optimal, it is the optimal social weighting scheme;
- Even if the existing income distribution is not considered optimal, it is not the business of transport policy to put it right.

These are the arguments of those who see cost-benefit analysis as an analogue to commercial appraisal, but accounting for external effects and consumer surplus as well as producer surplus. But there are some difficulties concerning the treatment of safety and environmental impact within such a framework. The third argument is particularly weak, since if the income distribution is sub-optimal, it is possible for public policy to take account of this at sector level without explicitly trying to correct the income distribution. Policy dimensions such as 'social exclusion' make sense in this context.

In the wtp approach, $\Omega_q = 1/\lambda_q$ where λ_q is the marginal utility of income for group q. Since we know that this declines with income, it follows that wtp weights in favour of the richer q. Reflections such as this have led to the exposition by Galvez and Jara-Diaz (1998). This argues that the most attractive option is to set the Ω_q factors equal to each other (e.g. unity) so that individuals' utility is weighted equally. Relative to the willingness to pay approach this rescales the benefits towards the lower income groups.

One possible way of implementing the Galvez and Jara-Diaz model is to standardise on time rather than income. In other words, assume that a small change in travel time bears equally heavily on utility terms on all groups q. Then the benefit is given as the sum of individual time equivalences and is in time units.

$$\Delta W = \Sigma_q (\Delta t + 1/V_q \Delta c) \tag{3.30}$$

To convert this to money units for the CBA, we require a single value of time V which can be considered equivalent to the standard value. So then, in money terms,

$$\Delta W = \Sigma_q \left(V \Delta t + \frac{V}{V_q} \Delta c\right) \tag{3.31}$$

Here we are effectively saying that time savings/losses are equally weighted among the different q but that costs are differentially weighted by the ratio of the standard value to the individual or group value V_q.

Suppose for a moment that the cost term Δc is zero. This may be roughly considered to be the case under which the standard value (originally termed “equity value” in the UK) was conceived – to appraise time savings from road investment without direct payment. The individual values of time V_q do not enter the evaluation formula (except indirectly since V is a weighted average of V_q), and the equity argument is directly reliant on the assumption that time savings are equally weighted for all q.

But this could also easily be a poor assumption, even if perhaps preferable to assuming that cost savings are equally weighted for all q. Tastes could easily vary across q. People on higher incomes might tend to work more hours so that their marginal utility of non-work time might be higher. The old argument that “we all have twenty four hours a day available” is too general to provide a rigorous defence of the single standard value of time.

Also, there is another difficulty. Appraisal practice in Britain and elsewhere has been to use neither (3.29) nor (3.31), but a mixture,

$$\Delta W = \Sigma_q (V \Delta t + \Delta c) \tag{3.32}$$

So, comparing with (3.29), time savings are rescaled by the ratio of V/V_q, but cost savings are not rescaled. This is inconsistent and has led to criticism. As Pearce and Nash (1981) point out,

“This inconsistency could lead to misallocation of resources; for example a scheme which gives the poor time savings at an increased money cost of travel could be selected in circumstances in which they would rather forgo the time savings for the sake of cheaper travel" [p 182]. A similar example, but from the opposite end of the income spectrum is given by Sugden (1999), para 7.2.

From the perspective of principle, therefore, we may conclude that:

- The standard value of non-working time is an incomplete approach to social weighting and introduces problems of inconsistency between time and costs;
- Specifically, it leads to the relativities between time and costs being different in modelling and evaluation, and this introduces problems where users are paying for benefits through fares or charges;
- The standard value relies on the strong assumption of equal marginal utility of time across groups;

Ideally, appraisal should:

- Discover the willingness to pay for all the costs and benefits accruing to all relevant social groups q;
- Use those values consistently in modelling and evaluation;
- Re-weight the costs and benefits according to some social weighting scheme which is common across sectors.

The weighting scheme should apply consistently across all impacts (time, money, safety risk, environment…). There is no particular reason to expect that the outcome would be a social value of time which is equal for all q. We therefore conclude that the argument of principle for the standard value of time falls.

In spite of this, on practical grounds it must be conceded that a full distributive weighting approach to appraisal is very ambitious for most transport applications. We can mention the following difficulties:

- Obtaining the relevant data on the pattern of usage by income and social group q at the scheme level;
- Defining the final incidence of costs and benefits to groups q – especially difficult for working time and revenue effects;
- Treating the non-monetised elements in the appraisal consistently with the monetised ones within the social weighting scheme;
- Agreeing the set of social weights.

3.9 Practical Problems

In the previous section we developed a strong case for using the utility weights in the demand model for assessing welfare changes for particular segments, and then facing up to the distributional consequences of that - effectively a political re-weighting. Nevertheless, we suspect that on practical grounds this will often be resisted, and that as a "second best" solution, we may be forced to live with externally imposed weights on the components of utility. The purpose of this section is to discuss the practical implications of the second best solution. A particular issue is in the specification of generalised costs derived from networks.

As before, we have a demand matrix T_{ij}, a set of links {l}, and appropriate formulae for deducing the components of link cost. After carrying out the route choice calculations, we derive the matrix of minimum (strictly, average) generalised cost between zones i and j:

$$C_{ij} = S_l\, e_{ijl}\, C_l \qquad (3.33)$$

Typically, this matrix C_{ij}, based on the route choice criterion underlying the formula for C_l will be used for carrying out further ("higher") demand calculations. However, if it is decided to re-weight the components for the purpose of **evaluation**, we will in general have an alternative "evaluation" version of the generalised cost, which we can write as C^u_l. [9]

It follows logically that the appropriate cost matrix for evaluation is given as

$$C^u_{ij} = S_l\, e_{ijl}\, C^u_l \qquad (3.34)$$

where the paths are, as before, decided on the basis of a "behavioural" formula C_l, but the link costs are subsequently calculated according to an "evaluation" formula.

Now **if** the formula for C^u_l is linear in the link cost components, for example

$$C^u_l = a.t_l + b.d_l \qquad (3.35)$$

then the evaluation matrix C^u_{ij} may be constructed on the basis of the appropriate component matrices: in this case, t_{ij} and d_{ij}. In other words, we can calculate

[9] A similar approach might be required for the demand model if the route choice criterion was for some reason different from the general definition of generalised cost used in the demand model. Although this is not theoretically "respectable", it is often done in practice.

$$t_{ij} = S_l\, e_{ijl}\, t_l \qquad (3.36)$$

and correspondingly d_{ij}, and then compute

$$C^u_{ij} = a.t_{ij} + b.d_{ij} \qquad (3.37)$$

In practice, however, this is not usually done, largely, it appears, because of restrictions imposed by assignment software. When there is only a single path between any origin and destination, the practical problems are slight. However, greater difficulty is encountered in cases of multiple routes. The calculations presented in the equations above, which are based on the a**verage** times, distances etc, by means of the e_{ijl} terms, can only be calculated if all the allocated paths are known. But most assignment programs discard the paths once they have been used to assign (a proportion of) the demand.

Hence, rather than calculate the generalised cost matrix as an average (Eq 3.33), it is taken as the cost along the (current) minimum cost route. If we are using an "equilibrium" assignment and it has adequately converged, the resulting matrix should be effectively identical. However, if the component matrices t_{ij} and d_{ij} are taken, correspondingly, as the matrices of time and distance along the minimum cost route, they are likely to have quite different values from those using the formula in (3.36).

In any case, with a non-linear formula for C^u_l this "component matrices" approach cannot be used at all: the evaluation matrix **must** be calculated on the basis of Eq (3.34): in other words, skimming the evaluation cost formula along the actual paths used. A good example of this is the standard vehicle-operating cost formula

$X_l = (a + b/v_l + c.v_l^2\,).\, d_l$ where v is the link speed[10]

The standard solution to equilibrium assignment of a fixed matrix T_{ij} takes the results of successive "all-or-nothing" assignments and combines them along the following lines. Define $e^{(n)}_{ijl}$ as representing the minimum cost paths found at iteration n. At the start (iteration 0) we set the link flows $Q^{(0)}_l$ to zero, and choose an appropriate set of link cost components (typically free-flow), yielding link generalised costs $C^{(0)}_l$. At each iteration we assign T_{ij} to the minimum cost paths to give "auxiliary" flows

---

[10] Note that although v_l is given by d_l/t_l, it is not generally appropriate to write the formula in matrix terms treating V_{ij} as if it can be derived as d_{ij}/t_{ij}, though this is by no means always respected in practice.

$$F^{(n)}_l = S_i\, S_j\, T_{ij}\, e^{(n)}_{ijl} \qquad (3.38)$$

We then calculate "average" flows $Q^{(n)}_l$ using the formula

$$Q^{(n)}_l = (1 - f^{(n)}).Q^{(n-1)}_l + f^{(n)}.F^{(n)}_l \qquad (3.39)$$

where $0 \le f^{(n)} \le 1$, and $f^{(1)} = 1$

The link costs are then updated based on $Q^{(n)}_l$ to produce new costs $C^{(n)}_l$, and the process is repeated until convergence is achieved on either $C^{(n)}_l$ or $Q^{(n)}_l$.

Since $Q^{(n)}_l$ is of the form of a recurrence relation, it can be shown that

$$Q^{(n)}_l = S_r\, ?^{(r)}\, F^{(r)}_l \quad r = 1....n \qquad (3.40)$$

where $\lambda^{(r)} = \phi^{(r)} . \prod_{s=1}^{s=n-r} (1 - \phi^{(n-s+1)})$

so that $Q^{(n)}_l$ is a weighted average of the set of auxiliary flows $\{F^{(n)}_l\}$.

This allows us to write the formula for the averaged flows at iteration n as:

$$Q^{(n)}_l = S_r\, ?^{(r)}\, S_i\, S_j\, T_{ij}\, e^{(r)}_{ijl} \quad r = 1....n$$

or, changing the order of the summations,

$$Q^{(n)}_l = S_i\, S_j\, T_{ij}\, S_r\, ?^{(r)}\, e^{(r)}_{ijl} \quad r = 1....n \qquad (3.41)$$

From this it is clear that each cell of the matrix is distributed among the separate paths identified in each iteration r according to the fractions $?^{(r)}$. In line with the notation in Section 3.4, we write

$$e_{ijl}^{n} = S_r\, ?^{(r)}\, e^{(r)}_{ijl} \quad r = 1....n \qquad (3.42)$$

where e_{ijl}^{n} is "the proportion of the total travel between i and j T_{ij} which uses link l", **as estimated after n iterations**.

Now, by assumption, the costs $C^{(n)}_l$ associated with these flows are dependent entirely on the levels of $Q^{(n)}_l$. Hence, if we **assumed** the multi-routeing pattern implied by e_{ijl}^{n}, and loaded the matrix T_{ij} accordingly, we would obtain the same set of link costs, since the same flows $Q^{(n)}_l$ would result.

Provided convergence is satisfactory, it should be the case that the average generalised cost over the routes actually used, which we may write as

$$\overline{C}*^{(n)}_{ij} = S_r\, ?^{(r)}\, S_l\, C^{(n)}_l\, e^{(r)}_{ijl} = S_l\, C^{(n)}_l\, \bar{\varepsilon}_{ijl}^{n} \qquad (3.43)$$

will be equal to $C^{*(n)}_{ij}$, the current minimum cost at iteration n, thus satisfying the Wardrop criterion. For general demand modelling, the Wardrop criterion ensures that the generalised cost matrix is independent of the actual routes used and, as noted above, for ease of calculation we would normally use $C^{*(n)}_{ij}$. However, if the generalised cost criterion C_l is to be re-weighted for evaluation purposes, ie using C^u_l instead, then it will not generally be the case that the same evaluation cost will be found along each path.

Despite the computational attraction of obtaining the evaluation cost matrix C^u_{ij} merely by "skimming" the evaluation generalised cost function C^u_l along the final set of paths $e^{(n)}_{ijl}$, this is unlikely to be a satisfactory approach, for two main reasons.

Firstly, if the matrix T_{ij} was actually assigned to the paths $e^{(n)}_{ijl}$, it would **not** produce the equilibrium costs and flows: these require the averaging process taking explicit account of multi-routeing. Secondly, the very process of equilibrium assignment aims to produce a number of alternative paths through the network with the same minimum cost. As the process converges, there will be more and more candidate paths, and the actual paths chosen by the all-or-nothing assignment for iteration n will be an arbitrary selection, typically based on the details of the minimum path algorithm. Inasfar as alternative paths imply very different combinations of the link cost **components**, using the set from the final iteration could be quite misleading.

For these reasons, therefore, it would seem far more sensible to use the paths **actually predicted** by the algorithm, represented by the array $\bar{\varepsilon}_{ijl}{}^{n}$. These <u>are</u> compatible with the assignment of the matrix T_{ij}, in that they generate the equilibrium link flows and costs. Though the paths are not guaranteed to be unique, even at convergence, the approach of using the paths actually generated to calculate the evaluation matrix C^u_{ij} must be preferred to any other straightforward proposal. While the extent to which this can be achieved will be software-dependent, the current trend towards path-based algorithms will facilitate the approach being recommended here.

Stochastic assignment methods

It seems likely that rather more care is needed when multi-routeing is generated on **stochastic** grounds. The feature of stochastic assignment is that not all travellers use the apparent minimum cost route(s), and in line with discrete choice theory, the general assumption is that there is variation in the perceived generalised costs.

In principle it would seem wrong to make use of **either** the minimum cost paths as delivered by the algorithm (based on, say, the mean generalised cost function) **or** the actual routes predicted by the stochastic algorithm. From a demand point of view, it will be appropriate to make use of a composite cost formulation, effectively incorporating the stochastic effects. It is much less clear what to do about the evaluation matrix, assuming that a different C^u_l function is considered appropriate.

In practice, using the average cost over the routes actually used (as with the non-stochastic equilibrium assignment above) may be acceptable in many cases. However, it is likely to lead to problems if there are major shifts in route choice, and *a fortiori* in the case of a new link leading to new routeing possibilities.

3.10 Other practical requirements

We have concentrated entirely on user benefits, which is certainly the item which relates most directly to modelling, as well as involving the greatest level of theoretical complexity. Nevertheless, it is appropriate to note that other considerations apply to carrying out a cost-benefit analysis, and although these are largely uncontroversial, they may give rise to practical difficulties.

In addition to user benefits, we are also interested in:

- • value of total resources consumed,
- • change in financial position of non-traveller groups.

It is worth noting that the estimates associated with the capital costs of infrastructure schemes are subject to considerable uncertainty. In addition, it has to be recognised that there are items that cannot easily be included in the CBA, which are no less important for that reason: this applies in particular to environmental effects. It is therefore important to avoid forms of presentation which increase the chances that the non-quantifiable elements are neglected.

Economic evaluation has mainly been applied to highway investment, usually with the implicit assumption that no direct user charge will be levied. However, for public transport, an essential element in the 'scheme' is the level of fare which it is proposed to charge. The costs and benefits will of course vary according to the level of fare proposed. An identical argument applies to the case of tolled roads. In addition, for public transport, the proposed level of service has an important effect on demand, and on operating costs.

As general practice, therefore, it is important to relate the evaluation of a proposed 'scheme' to the full set of policy assumptions associated with it. A new rail scheme, for example, cannot be evaluated without explicit reference to the proposed frequency of service and the fares to be charged. It should also be appreciated that nothing restricts the application of the CBA to **capital** schemes: it is equally appropriate for the testing of fares policy or changes in the level of service frequency etc.

There are a number of questions relating to units which do not affect the **nature** of the presentation of the results but are essential to interpreting them correctly. The main questions are:

- • the year to which the price levels relate
- • the period to which the demand forecasts relate (eg whole day, peak hour etc.)
- • the year to which the demand forecasts relate
- • whether the results have been factored up to a larger period (for example, a full year)
- • whether the results for several years have been combined, using appropriate growth and discount rates

All these assumptions need to be made clear.

Provision of output on an annual basis for a forecast year means that a single year evaluation can be carried out for that year on the basis of the information provided and a conversion of the capital cost into annual terms using a standard amortization formula. However, it is more usual to provide all quantities as Present Values, using Discounted Cash Flow (DCF) techniques over a time horizon of, say, 30 years. To carry out proper Present Value calculations requires a number of further assumptions which we briefly note.

Difficulty chiefly arises in estimating the stream of benefits, which will be affected by:

- • changes in total levels of travel (demographic factors etc.)
- • changes in values of time and operating costs (rising incomes, fuel prices)
- • changes in modal propensities (car ownership levels)

In the urban strategic context it is largely this last item which causes problems: with increasing car ownership, we can expect the modal choice to become more weighted against public transport even if the quality of service remains unchanged. In practice, of course, declining patronage is likely to lead to reduced services for financial reasons.

4 Current Topics in Evaluation

4.1 Introduction

In this Chapter we introduce a number of topics where the resolution is currently uncertain: these are generally issues which are either ignored in practice, or where the existing practice is in need of substantial development.

The first topic we deal with, in Section 4.2, is that of “new modes”, or, more generally, a change in the set of available travel opportunities. This is a particular instance where the RoH breaks down, while at the same time, partly because of the inherent demand modelling problems, there is a reluctance among practitioners to adopt the composite cost approach.

In Section 4.3, we turn our attention to Land-use changes, and the problems which they cause for evaluation. The standard application of the RoH dealing with generalised cost changes implicitly assumes that land-use is fixed.

Although much of the methodology for land-use evaluation has centred on the “gravity model”, the particular case of the “doubly constrained” model raises further issues about potential “shadow prices”, and some discussion is provided in Section 4.4.

In Section 4.5, we turn to the vexed question of Income effects, which has been a dominant topic in the work of Jara-Diaz, and has recently been taken up in contributions by McFadden and Karlström. This leads naturally on to questions of aggregation, and some further remarks are made on this in Section 4.6.

4.2 Dealing with New Travel Opportunities

The introduction of ‘new modes’ (or, more generally, new alternatives within a discrete choice framework) remains a difficult area where little progress has been made. A review of the general issues is given in Bates (1992): the problems relate both to demand modelling and evaluation. For evaluation, the problems occur when an option which is available in the before case is not available in the after case, or vice versa.

For the sake of illustration, we can revert to the simple logit example used in Section 3.2, where we had five choices k, Total demand T of 1000, and we assume ? = .02. Reverting, for simplicity, to the non-elastic example, suppose that option 5 is removed in the after case. The results are then:

Table 3

	Base		Strategy				
Option k	costs C_k	demand T_k	costs C'_k	demand T'_k	ΔS	approx	integral
1	20	225	15	287	–½(287+225). (15–20)	1280	
2	25	204	22	250	–½(250+204). (22 –25)	681	
3	45	137	35	193	–½(193+137). (35– 45)	1650	
4	15	249	18	271	–½(271+249). (18 –15)	–780	
5	30	185	∞	0	–½(0+185). (∞ - 30)	?	
* (Σ)	–54.48	1000	–47.37	1000	1000. (–54.48 + 47.37)	*2831...*	–7108

Thus, according to the exact formula, consumer surplus has **de**creased by 7108. As we shall see, this change has been largely brought about by the removal of option 5.

For the first four options, the RoH formula gives a total of 2831, but the term for the fifth option cannot be evaluated. Although we could assume a finite value of the cost for option 5, the resulting estimate of consumer surplus is far too sensitive to the value assumed. Neither can we ignore the contribution of option 5, since the composite cost calculation shows that the true value for benefit is a large negative number.

The conclusion is, as before, that the RoH cannot be used in those instances where **large** changes in cost take place, and this is always the case when the availability of options changes[11]. The problem can, as usual, be circumvented by returning to the exact formula for the total benefit at some level above. The benefit associated with the new mode can then in principle be calculated by subtracting the benefits for all other modes, calculated by the standard RoH formula. It must

[11] unless the options which are removed capture a negligible market share when they are in fact available

be noted, however, that, as is always the case with the composite cost approach, this estimate of new mode benefit will not be broken down into components (for example, time and money). In the example, the benefit associated with the removal of option 5 can be calculated as the difference between the total benefit (–7108) and the existing mode benefit (+2831), giving –9939.

In terms of **demand** modelling, the first question to be addressed is: under what circumstances should an alternative be recognised as a 'new' option? IN the case of mode, this relates both to the similarity of and key differences between alternatives **and** the identification of modal constants within the utility function. A further question is the position of a new mode within the model hierarchy: even if we are confident about measuring utility for the new mode, there is considerable variation in the market share prediction depending on the hierarchical assumptions. But in reality, we cannot be in any way confident about the utility of the new mode because a major element in the utility of **existing** modes is the modal constant. We therefore need to examine the possibilities of estimating such a constant for the new mode.

There are three methods for consideration:

- revealed preference studies in situations where the new mode has actually been introduced;
- attempting to 'decompose' existing modal constants into attribute effects and thereby to deduce the likely value for the new mode; and
- stated preference approaches which describe the new mode to existing travellers.

All these methods involve considerable difficulty. In principle, the revealed preference approach offers the most reliable way forward, since, as well as providing a direct estimate of the modal constant for the new mode, it also provides evidence about the appropriate hierarchical structure, as discussed earlier. However, there are major impediments. In the first place the new mode may not actually have been introduced anywhere, or at least not in a form sufficiently close to the proposed introduction to offer reasonable guidance. Even if it has been introduced, the expense of surveying in a different area may be prohibitive. In addition, there are well known requirements for successful revealed preference studies of discrete choice (in particular those relating to the proportion of 'genuine' choices relative to cases where one alternative is clearly dominant) which require careful design and are not easy to satisfy. Finally, even after overcoming these problems, the fundamental issue remains of the interpretation of the modal constant and its scope for transferability.

The safest approach is to define as a 'reference' mode that existing mode considered to be 'most similar' to the new mode. As a default, the constant for the new mode is assumed to be the same as that for the reference mode - any departures from this would need to be specifically justified such as a 5 minute advantage on grounds of comfort, reliability, image and so on.

Nonetheless, the fact that a typical outcome is that a substantial proportion of the benefits is associated with the constant term (more strictly, the provision of a new alternative which is conceptually distinct from other existing alternatives) has made practitioners reluctant to accept the theoretical outcome, especially in cases where the new mode does not in fact offer substantial improvements in terms of time and money benefits.

4.3 Land-use changes[12]

As noted earlier, the development of the logsum formula makes it clear that **all** the changes in "cost" which affect demand must be taken into account in the approximation formula. In particular, if there is an implied change in the "destination" utility, then a failure to include this to include this will invalidate the approximation. It is because such changes are typically associated with land-use effects that the convention has developed that the RoH approximation is only valid when the land-use is constant. More strictly, however, we can deduce that a rule of a half benefit calculation **based only on transport costs** is only valid when the land-use is constant.

In this section we seek to go beyond this restriction, to allow for land-use changes. This is an area which has been relatively little explored in recent years, though a notable exception is the work of Martinez (see, for example, Martinez & Araya, 1998). However, the work reported here has been developed independently of Martinez' contribution: at the time of writing, it is not clear whether the two approaches can be reconciled.

Although there may be practical difficulties, as we discuss later, the rule of a half will in general remain a valid approximation, under the usual conditions relating to linearisation of the demand curve, even when there are land-use changes, **provided** that the costs associated with the land-use change are included. The key point is that

- if **only** transport costs are changing, with no land-use changes, then changes in C_{ijmt} account for all the changes in T_{ijmt} and there are no other transport users' benefits to consider;

[12] Much of the work in this section is the result of a recent collaboration with David Simmonds, and his contribution is hereby gratefully acknowledged

- if on the other hand land-use changes **are** occurring, whether induced by the transport cost or due to other effects, then factors other than the changes in the C_{ijmt} **may be** affecting the T_{ijmt} and the influence of those other factors on user benefits must be considered.

The rest of this section deals with the latter case, where land-uses are changing. To simplify the exposition, we omit any discussion of mode or time of day choice. Although in practice the demand T_{ij} and costs C_{ij} may need to be "composited" over lower level of choice, we ignore this in the notation.

We assume for convenience that we are dealing only with home-based travel, with all travel being "produced" at home and being "attracted" somewhere else. If we first consider the choice of "attraction" zone (conventionally "destination choice") by means of the standard logit choice model:

$$p_{j|i} = \frac{W_j . exp\left(-\lambda^D . C_{ij}\right)}{\sum_j W_j . exp\left(-\lambda^D . C_{ij}\right)} \quad (4.1)$$

where W_j is the measure of the (relative) attractiveness of zone j, then the composite cost or expected utility of all trips from production zone i is given by

$$C_{i*} = \frac{-1}{\lambda^D} ln \sum_j W_j . exp\left(-\lambda^D . C_{ij}\right) \quad (4.2)$$

The values of W_j are related to the relative utility (net of any terminal costs) of reaching or visiting destination j. They can be transformed into units of generalised cost by defining

$$w_j = \frac{1}{l^D} \ln\left(W_j\right) \quad \text{or} \quad W_j = \exp\left(l^D . w_j\right) \qquad (4.3)$$

Substituting, we obtain

$$C_{i*} = \frac{1}{-\lambda^D} ln \sum exp\left(\lambda^D\left[w_j - C_{ij}\right]\right) \quad (4.4)$$

The benefit of changes in transport cost **and** zonal attraction can now be evaluated -exactly by the change in composite cost if the origin totals are unchanged, or more generally by the Rule of a Half approximation. However, because we have now allowed for a change in the destination utility, we have gone beyond the "pure transport" case, and we can write, for the RoH version:

$$\Delta S \approx -\frac{1}{2}\sum_i \sum_j \left(T'_{ij} + T_{ij}\right)\left(c'_{ij} - w'_j - c_{ij} + w_j\right)$$

$$= -\frac{1}{2}\sum_i \sum_j \left(T'_{ij} + T_{ij}\right)\left(c'_{ij} - c_{ij}\right) + \frac{1}{2}\sum_i \sum_j \left(T'_{ij} + T_{ij}\right)\left(w'_j - w_j\right) \qquad (4.5)$$

This suggests that we can estimate the benefits which arise in transport and in relative attractiveness by

converting relative attractiveness into units of generalised cost;

carrying out a rule of a half calculation using the changes in converted attractiveness and the numbers of trips attracted;

adding the result of step 2 to that of conventional rule of a half calculation on changes in transport generalised cost and the number of trips.

For convenience we define the change in surplus associated specifically with changes in attraction as $\Delta S(Att)$ and the corresponding term based on **transport** generalised cost as $\Delta S(Tp)$, so that $\Delta S = \Delta S(Tp) + \Delta S(Att)$ (4.6)

It is critical to note, however, that the calculations have (once again) been partitioned for convenience: it is not the case that the $\Delta S(Tp)$ term represents that part of the benefit "due to transport" nor that $\Delta S(Att)$ represents the benefit due to the destination land-use. The total transport demand reflects both changes.

As our terminology tries to emphasise, it is important to consider all this in terms of production – attraction matrices rather than origin – destination matrices. For conformity the C_{ij} terms should also be defined on this basis. Although the transport term $\Delta S(T)$ will still be valid if calculated on an O-D basis, the land-use term $\Delta S(Att)$ will not, and confusion is likely to arise if the conventions are different.

Further thought needs to be given to the treatment of trips not modelled on a production – attraction basis, and, related to this, the implications of modelling more complex tours.

We must also consider the possible benefit associated with changes in (residential) location, i.e. travellers changing the zone in which they produce trips. Suppose again that we use a logit model to predict residential location, so that

$$p_i = \frac{Z_i . exp\left(-\lambda^L .C_{i*}\right)}{\sum_i Z_i . exp\left(-\lambda^L .C_{i*}\right)} \qquad (4.7)$$

where Z_i is related to the utility of locating in zone i, exclusive of the expected (dis-) utility of travel, and, again, net of "terminal costs" (eg rent)[13]$_i$, and l^L is the coefficient for production zone (eg residential) choice. The expected utility of location is given analogously by

$$C_{**} = \frac{1}{\lambda^L} ln \sum_i Z_i . exp\left(-\lambda^L .C_{i*}\right) \qquad (4.8)$$

Following the same procedure as for attractions, we define

$$z_i = \frac{1}{\lambda^L} \ln\left(Z_i\right) \qquad (4.9)$$

obtaining

$$C_{**} = \frac{1}{-\lambda^L} ln \sum_i exp\left(\lambda^L\left[z_i - C_{i*}\right]\right) \qquad (4.10)$$

By defining analogously the change in "location surplus" as

$$\Delta S(\text{ Prd }) = \frac{1}{2} \sum_i \left(T'_{i*} + T_{i*}\right)\left(z'_i - z_i\right)$$

the overall benefit of change in transport, attraction *and* production-zone location is now given by

$$\Delta S = \Delta S(Tp) + \Delta S(Att) + \Delta S(Prd) \qquad (4.11)$$

The proviso about not interpreting the partitioning into components too literally applies in this case as well.

This suggests that we can in principle obtain a complete evaluation of user benefits, including those that occur from changes in location, by deducing the net change in origin (location) utility, over and above that due to transport and attraction changes, which explains the predicted

[13] Note that if one tried to implement this, one would have to address the problem of expressing Z_i in appropriate units, probably per trip.

changes in location, and carrying out an additional set of rule-of-one-half calculations. It is not necessary that the location choice process should be of the form assumed in (18); providing the probability of locating at i is in equilibrium with C_{i*}, and we can assign an appropriate value to $-l^L$, then in theory we can deduce a set of Z_i.

One approach to evaluation of user benefits in land-use/transport planning is therefore to add to the conventional transport benefits

an evaluation of changes in attraction, and

an evaluation of changes in location,

both measured in terms of transport generalised cost and, if necessary, deduced from the otherwise unexplained changes in travel patterns. The attraction and location benefits will be measures of **net** benefit to users.

4.4 Constraints in evaluation – shadow prices

The exposition in the previous section assumed that the "costs" faced by travellers, including the attraction and production components, were the appropriate values to use. However, in some cases there will be constraints present, and these are often treated as shadow prices. It therefore needs to be discussed how these impact on the evaluation.

The most discussed case in the transport modelling field is that of the doubly constrained distribution model, and there are various alternative interpretations of this. We will begin with a particular interpretation which builds on the discussion so far, and then attempt to show how the same conclusions apply to the standard interpretation.

In line with the earlier discussion, the singly-constrained distribution model can be interpreted as a destination choice model in which the total trips from each origin i were fixed at T_{i*}., with C_{ij} as the (possibly composite) transport cost of reaching j from i and w_j as the inherent destination utility (in cost units). Such a model can also be embedded in a higher level model for the choice of origin (residence), postulating an inherent origin utility of z_i (again, in cost units).

The product of p_i and the conditional destination choice $p_{j|i}$ can then be regarded as a distribution model assuming a fixed total of trips T_{**}.

This structure is hierarchical (assuming $l^L < l^D$). However, transport models usually assume (implicitly) that the two parameters are equal. Under these circumstances we obtain the **unconstrained** "gravity" model with negative exponential "deterrence function":

$$T_{ij} = T_{**}\ p_{ij} = T_{**}\ \frac{exp\left(\lambda^D .(\, z_i + w_j - C_{ij}\,\right)}{\sum_{ij} exp\left(\lambda^D .(\, z_i + w_j - C_{ij}\,\right)} \qquad (4.12)$$

The gravity model is more usually written in the form:

$$T_{ij} = A_i\, O_i\, B_j\, D_j \exp\, (-\, l^D C_{ij}) \qquad (4.13)$$

essentially consisting of an origin factor, a destination factor, and an "interaction" term.

If we now allow for constraints on the number of origins and destinations in each zone, it is well-known that this can be represented mathematically by modifying the values of z_i, w_j. However, since we wish to keep open the possibility of changing the intrinsic utilities, we will not do this, but follow the general approach of Neuburger (1971) in dealing with shadow prices. We define ρ_i as the origin shadow price and σ_j as the destination shadow price. In the unconstrained case, these are both identically equal to zero.

In the constrained case, however, the doubly constrained "gravity" model with negative exponential "deterrence function" can be represented as:

$$T_{ij} = T_{**}\ \tilde{p}_{ij}\ = T_{**}\ \frac{exp\left(\lambda^D .(\, z_i + w_j + \rho_i + \sigma_j - C_{ij}\,\right)}{\sum_{ij} exp\left(\lambda^D .(\, z_i + w_j + \rho_i + \sigma_j - C_{ij}\,\right)} \qquad (4.14)$$

where the symbol $\tilde{p}_{ij}$ is used to denote the constrained model.

If we now define H_{ij} as:

$$H_{ij} = \ l^D\, [z_i + w_j + \ r_i + s_j - C_{ij}] \qquad (4.15)$$

then $T_{ij} = T_{**} .\ \tilde{p}_{ij} \ = \ T_{**} .\ \exp\, (H_{ij})/\Sigma_{rs} \exp\, (H_{rs})$ (4.16)

Defining for convenience the "composite" H_{**} by the usual "logsum" formulation:

$$\exp\, (H_{**}) = \Sigma_{rs} \exp\, (H_{rs}) \qquad (4.17)$$

allows us to write

$\tilde{p}_{ij}$ = exp (H_{ij}–H_{**}) (4.18)

For this to be equivalent to Eq (4.13) we must have

$$\exp (H_{ij}) = [\exp (H_{**})/T_{**}]A_i\ O_i\ B_j\ D_j \exp (-\lambda^D C_{ij}) \qquad (4.19)$$

Taking logs, this implies that

$$\lambda^D [z_i + w_j + \rho_i + \sigma_j] = H_{**} + \ln [(1/T_{**})\ A_i\ O_i\ B_j\ D_j] \qquad (4.20)$$

Although, as is well-known, there is an element of indeterminacy here, it is clear that values of z_i and w_j, ρ_i and σ_j can be chosen to make the two formulations equivalent. Suppose we set:

$$\lambda^D [z_i + \rho_i] = \ln [A_i\ O_i] \ (4.21a)$$

$$\lambda^D [w_j + \sigma_j] = \ln [B_j\ D_j] \ (4.21b)$$

Then exp H_{ij} = A_i O_i B_j D_j exp ($-\lambda^D C_{ij}$) = T_{ij} so that exp H_{**} = T_{**}.

One level of indeterminacy relates to the balancing factors A_i and B_j. Multiplying each value of A_i by an arbitrary x and dividing each B_j by the same x will have no effect on the outcome. This means that each term [$z_i + \rho_i$] may have an arbitrary constant ω added, and each term [$w_j + \sigma_j$] the same constant subtracted. However, provided that these terms always appear additively, the effect will cancel out.

A second, associated, level of indeterminacy relates to the overall H_{**}. Suppose now we merely add ω to each term [$z_i + \rho_i$] as defined, but do **not** subtract it from the corresponding terms [$w_j + \sigma_j$]. Then we will have:

$$\exp H_{ij} = \exp (\omega)\ A_i\ O_i\ B_j\ D_j \exp (-\lambda^D C_{ij}) = \exp (\omega)\ T_{ij}$$

Hence exp H_{**} = exp (ω) T_{**}, and H_{**} = ln (T_{**}) + ω

The upshot is that the term [λ^D ($z_i + w_j + \rho_i + \sigma_j$) – H_{**}] is uniquely defined.

For convenience, we write $G_{ij} = z_i + w_j - C_{ij}$ (4.22)

as the generalised utility attributable to the origin and destination utility and the cost between them. We can re-write

$$H_{ij} = \lambda^D [G_{ij} + \rho_i + \sigma_j] \qquad (4.23)$$

Taking now the differential of H_{**}, we have

dH_{**}

$= \Sigma_{ij} \lambda^D (1/\Sigma_{rs} \exp [\lambda^D (G_{rs} + \rho_r + \sigma_s)]). \exp [\lambda^D (G_{ij} + \rho_i + \sigma_j)]. (dG_{ij} + d\rho_i + d\sigma_j)$

$= \Sigma_{ij} \lambda^D p_{ij}. (dG_{ij} + d\rho_i + d\sigma_j)$ 4.24)

Provided Green's theorem relating to the symmetry of the Jacobian is satisfied (which applies straightforwardly in this case), we can therefore evaluate H_{**} as the line integral with respect to the elements of the matrix **G**:

$$H_{**} = \lambda^D \int_{-\infty}^{\mathbf{G}} \tilde{\mathbf{p}}^T . \frac{\partial[\rho + \sigma]}{\partial \mathbf{G}} \mathbf{dG} \quad + \lambda^D \int_{-\infty}^{\mathbf{G}} \tilde{\mathbf{p}}^T .\mathbf{dG} \quad +\text{constant.} \ (4.25)$$

Note that the shadow price Jacobians are non-zero, since the values of the shadow prices will be affected if the costs change or origin and destination utilities change. There is no general functional relationship between the elements **c**, **z** and **w**, so that the integral of $\tilde{\mathbf{p}}$ with respect to **G** can be evaluated as the sum of the integrals with respect to the component elements[14].

Hence, in the general constrained case, the value of H_{**} can be written:

$$H_{**} = \lambda^D \int_{-\infty}^{\mathbf{v}} \tilde{\mathbf{p}}^T .\mathbf{dz} \ + \lambda^D \int_{-\infty}^{\mathbf{w}} \tilde{\mathbf{p}}^T .\mathbf{dw} \ - \lambda^D \int_{-\infty}^{\mathbf{c}} \tilde{\mathbf{p}}^T .\mathbf{dc} \ + \lambda^D \int_{-\infty}^{\mathbf{G}} \tilde{\mathbf{p}}^T . \frac{\partial[\rho + \sigma]}{\partial \mathbf{G}} \mathbf{dG} \ +\text{constant} \qquad (4.26)$$

The overall Consumer Surplus per trip for a "representative traveller" CSRT at a given level of the v_i, w_j and c_{ij} elements can be written as

$$-\lambda^D \int_{-\infty}^{\mathbf{c}} \tilde{\mathbf{p}}^T .\mathbf{dc} \ + \lambda^D \int_{-\infty}^{\mathbf{v}} \tilde{\mathbf{p}}^T .\mathbf{dz} \ + \lambda^D \int_{-\infty}^{\mathbf{w}} \tilde{\mathbf{p}}^T .\mathbf{dw}$$

and thus, expressed in cost units, is given (up to an indeterminate constant) by

$$1/\lambda^D H_{**} - \int_{-\infty}^{\mathbf{G}} \tilde{\mathbf{p}}^T . \frac{\partial[\rho + \sigma]}{\partial \mathbf{G}} \mathbf{dG} \qquad (4.27)$$

---

[14] though the incorporation of explicit transport impacts on land-use may lead to a more complex treatment

Note therefore that, in general, $1/\lambda^D$ H_{**} does **not** represent the consumer surplus – it must be adjusted for the contribution of the constraints. This is consistent with the conclusion of Williams and Senior (1978) and Neuburger & Wilcox (1976). If the constraints are not present, then r and s are zero, and no adjustment is required.

We can simplify the Jacobian integral $\int_{-\infty}^{\mathbf{G}} \tilde{\mathbf{p}}^T . \frac{\partial[\rho+\sigma]}{\partial \mathbf{G}} \mathbf{dG}$. The integrand is

$$\sum_{ij} \tilde{p}_{ij} \sum_{rs} \frac{\partial[\rho_i + \sigma_j]}{\partial G_{rs}} dG_{rs} = \sum_{ij} \tilde{p}_{ij} \sum_{rs} \left(\frac{\partial \rho_i}{\partial G_{rs}} + \frac{\partial \sigma_j}{\partial G_{rs}} \right) dG_{rs}$$

which can be written as:

$$\sum_{ij} \tilde{p}_{ij} \sum_{rs} \left(\frac{\partial \rho_i}{\partial G_{rs}} \right) dG_{rs} + \sum_{ij} \tilde{p}_{ij} \sum_{rs} \left(\frac{\partial \sigma_j}{\partial G_{rs}} \right) dG_{rs}$$

$$= \sum_{i} \tilde{p}_{i*} \sum_{rs} \left(\frac{\partial \rho_i}{\partial G_{rs}} \right) dG_{rs} + \sum_{j} \tilde{p}_{*j} \sum_{rs} \left(\frac{\partial \sigma_j}{\partial G_{rs}} \right) dG_{rs} \quad (4.28)$$

Hence, the summations over i and j can be taken outside the integrals. Note that for the doubly constrained model, ρ_i and σ_j are dependent on the **entire** matrix **c** and the vectors **v** and **w**, not merely the related rows or columns.

Hence we obtain:

$$\int_{-\infty}^{\mathbf{G}} \tilde{\mathbf{p}}^T . \frac{\partial[\rho+\sigma]}{\partial \mathbf{G}} \mathbf{dG} = \sum_{i} \tilde{p}_{i*} \int_{-\infty}^{\mathbf{G}} \frac{\partial \rho_i}{\partial \mathbf{G}} \mathbf{dG} + \sum_{j} \tilde{p}_{*j} \int_{-\infty}^{\mathbf{G}} \frac{\partial \sigma_j}{\partial \mathbf{G}} \mathbf{dG} = \sum_{i} \tilde{p}_{i*} \int_{-\infty}^{\rho} d\rho_i + \sum_{j} \tilde{p}_{*j} \int_{-\infty}^{\sigma} d\sigma_j$$

$$= \Sigma_i \, p_{i*} \, \rho_i + \Sigma_j \, p_{j*} \, \sigma_j \qquad (4.29)$$

Note that this can be written as $\Sigma_{ij} \, p_{ij} \, [\rho_i + \sigma_j]$, and can be viewed as the “shadow expenditure” per trip.

Thus the consumer surplus for a “representative traveller” may be written

$$\text{CSRT} = 1/\lambda^D \, H_{**} - \int_{-\infty}^{\mathbf{G}} \tilde{\mathbf{p}}^T . \frac{\partial[\rho+\sigma]}{\partial \mathbf{G}} \mathbf{dG} = 1/\lambda^D \, H_{**} - \Sigma_{ij} \, \tilde{p}_{ij} \, [\rho_i + \sigma_j] \qquad (4.30)$$

and the total Consumer surplus, given a constant T_{**}, is

$$S = T_{**}.CSRT = 1/\lambda^D \Sigma_{ij} T_{ij} H_{**} - \Sigma_{ij} T_{ij} [\rho_i + \sigma_j]$$

We can substitute for H_{**} (from Eq 4.18) noting that

$$H_{**} - H_{ij} = H_{**} - \lambda^D (G_{ij} + [\rho_i + \sigma_j]) = - \ln(\tilde{p}_{ij}),$$

so that

$$S = \Sigma_{ij} T_{ij} (G_{ij} - 1/\lambda^D \ln(\tilde{p}_{ij})) \qquad (4.31)$$

We noted earlier that the total area under the demand curve, W, is the sum of S and an "expenditure" or "consumption" term E. It can be shown that the term $\Sigma_{ij} T_{ij} G_{ij}$ (which is negative in transport cost terms) is the (negative) consumption term, while the term $-1/\lambda^D \Sigma_{ij} T_{ij} \ln(\tilde{p}_{ij})$ represents W. Hence, as usual, when we are comparing two situations, the measure of benefit, ΔS, can be calculated as $\Delta W - \Delta E$.

Now consider a general change brought about by $\mathbf{G} \to \mathbf{G}'$, involving a change either or both of transport costs ($\mathbf{C} \to \mathbf{C}'$) and land-use effects ($\mathbf{z} \to \mathbf{z}'$ and/or $\mathbf{w} \to \mathbf{w}'$). Assume that constraints continue to apply, but not necessarily with the same values. As a result we obtain a new demand T'_{ij} and a new set of shadow prices $\rho'_i + \sigma'_j$.

The term ΔE is straightforwardly calculated as:

$$\Delta E = (\mathbf{T}'.\mathbf{G}' - \mathbf{T}.\mathbf{G})$$

$$= \Sigma_{ij} (T'_{ij}.(c'_{ij} - [z'_i + w'_j]) - T_{ij}.(c_{ij} - [z_i + w_j])) \quad (4.32)$$

NB This **excludes** the contribution from the shadow prices.

When T_{**} does not change (ie $T'_{**} = T_{**}$), then ΔW is given exactly by:

$$\Delta W = - 1/\lambda^D [\Sigma_{ij} T'_{ij} \ln(\tilde{p}'_{ij}) - \Sigma_{ij} T_{ij} \ln(\tilde{p}_{ij})] \qquad (4.33)$$

However, this formula is not valid if $T'_{**} \neq T_{**}$, because we need to take account of the shape of the (overall) demand curve between T_{**} and T'_{**}.

In this more general case, we can, as before, use the RoH approach to approximate the change in W, giving:

$$\Delta W \approx ½ \Sigma_{ij} (T'_{ij} - T_{ij}).(c'_{ij} - [z'_i + w'_j + \rho'_i + \sigma'_j] + c_{ij} - [z_i + w_j + \rho_i + \sigma_j]) \quad (4.34)$$

From this we deduct the change in "expenditure" ΔE as in Eq (4.32). Quite generally, therefore, provided the linear approximation is appropriate, we can calculate the overall change in transport **and** land-use benefits as:

$$\Delta S \approx \tfrac{1}{2}\,\Sigma_{ij}\,(T'_{ij} - T_{ij}).(\,c'_{ij} - [z'_i + w'_j + \rho'_i + \sigma'_j\,] + c_{ij} - [z_i + w_j + \rho_i + \sigma_j\,])$$

$$-\,\Sigma_{ij}\,(T'_{ij}.\,(\,c'_{ij} - [z'_i + w'_j\,]\,) - T_{ij}.\,(\,c_{ij} - [z_i + w_j])) \qquad (4.35)$$

By the usual algebraic re-arrangement, we can write this in two alternative forms:

a) $-\tfrac{1}{2}\,\Sigma_{ij}\,(T'_{ij} + T_{ij}).\{(\,c'_{ij} - [z'_i + w'_j + \rho'_i + \sigma'_j\,]) - (c_{ij} - [z_i + w_j + \rho_i + \sigma_j\,])\}$

$$-\,\Sigma_{ij}\,(T'_{ij}.\,(\rho'_i + \sigma'_j) - T_{ij}.\,(\rho_i + \sigma_j)) \qquad (4.36a)$$

b) $-\tfrac{1}{2}\,\Sigma_{ij}\,(T'_{ij} + T_{ij}).\{(\,c'_{ij} - [z'_i + w'_j\,]) - (c_{ij} - [z_i + w_j])\}$

$$-\,\tfrac{1}{2}\,\Sigma_{ij}\,(T'_{ij} - T_{ij}).(\,\rho'_i + \sigma'_j + \rho_i + \sigma_j\,) \qquad (4.36b)$$

Form a) applies the RoH formula to all cost/utility elements **including** the shadow costs, but then has to correct the result for the "shadow expenditure": this is commensurate with the earlier discussion about H_{**}. Form b) applies the rule of a half to the "true" cost elements (the components of **G**), and has to correct with a ΔW RoH formula involving only the shadow prices.

We can now note the following:

a If there are no changes in the origin and destination utilities, then there is no contribution to the benefits from these terms;

b If the origin and destination constraints are unchanged, then from formula b) it can be seen that the shadow price contributions vanish. This is because they only occur in the second (ΔW) term, and the ρ_i terms are constant over j, while the σ_j are constant over i. Hence this term can be written:

$$-\,\tfrac{1}{2}\,\Sigma_i\,(T'_{i*} - T_{i*}).(\,\rho'_i + \rho_i) \;-\tfrac{1}{2}\,\Sigma_j\,(T'_{*j} - T_{*j}).(\,\sigma'_j + \sigma_j\,) = 0$$

since $(T'_{i*} = T_{i*}),\ (T'_{*j} = T_{*j})\ \forall\ i,j$

A variant of this approach is set out by Neuburger (1971), which concludes that because of the constraints **and** the assumption of no change in origin and destination utilities, ΔS is given by the standard rule of a half.

Alternatively, in the simpler case where only the transport costs change, we can return to the earlier general formula:

$$S = \Sigma_{ij}\, T_{ij}\, (G_{ij} - 1/\lambda^D \ln (p_{ij})) \qquad (4.37)$$

and substitute for $\ln (p_{ij}) = \ln [(1/T_{**})\, A_i\, O_i\, B_j\, D_j] - \lambda^D G_{ij}$, so that

$$S = -\, 1/\lambda^D\, \Sigma_{ij}\, T_{ij}\, (\ln [(1/T_{**})\, A_i\, O_i\, B_j\, D_j]) \quad (4.38)$$

If the constraints continue to apply, the quantities O_i, D_j and T_{**} will not change, and this can be used to give:

$$\Delta S = -1/\lambda^D\, [\Sigma_i\, O_i \ln (A'_i/A_i) + \Sigma_j\, D_j \ln (B'_j/B_j)]$$

which corresponds with the result by Williams & Senior (1978): the change in user benefit can be obtained from the changes in the "balancing" factors. Note, however, that if the constraints **change**, so that, while still constrained, the values of O_i, D_j do not necessarily remain constant, then a different result is obtained. The Williams & Senior result assumes that the constraints remain the same.

In the more general case, we can always evaluate the first term in formula (4.36b) as we know **c**, **v**, and **w** in the two alternative scenarios. In the case where the constraints are not equal, we have to deal explicitly with the terms $[\rho_i + \sigma_j]$ in the two scenarios. From the earlier equations (4.21a) and (4.21b), we have:

$$\lambda^D\, [z_i + \rho_i + w_j + \sigma_j] = \ln [A_i\, O_i\, B_j\, D_j]$$

Since we know **v** and **w**, we can estimate the required values for $[\rho_i + \sigma_j]$ making use of the balancing factors in the doubly constrained procedure.

The general approach is well summarised by Neuburger & Wilcox (1976):

either

The surplus function .. for the unconstrained case [is].. used to evaluate surplus, but with the shadow prices added to the actual prices... the change in shadow price revenue [is] added to arrive at user benefits

or

..the shadow prices are added to the [costs in the rule of a half] and the shadow price revenue is added....[s] ince [it] is not actual expenditure

In the case of the doubly constrained model, therefore, there is no need to correct for the shadow prices if the rule of a half is used, provided the constraints do not change and the origin and destination utilities are kept constant. However, this is a special case of a wider evaluation result.

Note also that there is no **mathematical** way to distinguish between the shadow prices and the intrinsic utilities. If the constraints remain the same, then a change in the intrinsic utilities **v** and **w** will have no effect on the final constrained trip pattern.

4.5 Income effects

We have emphasised throughout that the theory applying to welfare measures is greatly simplified by the assumption of a constant marginal utility of income, and that the "linear in income" formulation for the indirect utilities used in most transport models is compatible with this assumption. In this section, we consider some of the implications when this assumption is not considered appropriate.

McFadden (1999) has investigated the computation of the average Compensating Variation when indirect utility is non-linear in income. As well as developing a theoretical treatment, which involves considerable complexity, he shows with a simple example that the impact of non-linearity can be severe, especially (as might be expected) when price or quality changes are large. He also provides computable bounds on the average CV which in some cases may deliver sufficient accuracy to be directly usable. He concludes that "an analytical solution is generally unavailable", and outlines a simulation approach for the practical calculation.

In an impressive contribution, Karlström (2001) clarifies some of the underlying difficulties, as well as illustrating how an **exact** solution can be obtained when the random utility model is of the GEV ("Generalised Extreme Value") type, based on an earlier paper (Karlström (1998)). He notes that the existence of the problem has been recognised for some time, and that various approximations have been proposed (eg Jara-Diaz & Videla (1990)). Effectively, the problem stems from the fact that "the compensation … itself will affect the choice probabilities when the marginal utility of money is not constant".

Karlström is able to show that the "compensated choice probability" $p^c_i(\mu)$ can be calculated in terms of the model assumptions, and implies "the probability that an individual will choose i before and after being compensated" with income μ. Hence he derives an exact formula for the expected income required to restore the base level of utility, given as:

$$\sum_i \left\{ \mu_{ii} p_i^c(\mu_{ii}) - \int_{\tilde{\mu}}^{\mu_{ii}} \mu \frac{\partial p_i^c(\mu)}{\partial \mu} d\mu \right\}$$

where $\tilde{\mu} = \mathrm{Min}_j (\mu_{jj})$.

Being a one-dimensional integral, this is quite tractable for cases where the compensated probability can be directly evaluated (close form), as applies to the GEV model.

Thus, for the class of GEV discrete choice models, a exact solution for the CV is available even when the AIRUM condition does not apply.

4.6 Aggregation (again)

Although the treatment outlined in the previous section is perfectly general, it is important to distinguish two separate effects, which are in danger of being confounded in the theoretical literature. The issue relates to the nature of the non-constancy of the marginal utility of income.

It has generally been argued that **for a given individual**, the impact of any practical change in transport conditions is unlikely to have a significant impact on income (see eg Glaister, 1981). In this case, the straightforward (Marshallian) consumer surplus calculation will be acceptable at an individual level, and the problem becomes essentially an aggregation issue. Against this, the considerable corpus of work by Jara-Diaz and associates (eg Jara-Diaz & Videla, 1989) has made a strong case that the assumption of constant marginal utility of income is not appropriate in Third World conditions, where a substantial proportion of expenditure may be required for the journey to work. In such cases, the Karlström formulation has direct application.

In terms of interpersonal comparisons, however, we should expect, as a **default** presupposition, to find income variation in terms of willingness to pay for transport improvements. An obvious case is the so-called "value of time" - ie the willingness to pay for a unit change in travel time, where there is now a wealth of evidence relating to a positive relationship between the value of time and income. Indeed, it would be reasonable to conjecture that where income effects are **not** found, it is probably because the sample size is insufficiently large to reveal it.

This brings us back to the question of the "representative consumer". It is of some passing interest that questions have been raised as to whether Gorman, in setting out this concept, was trying to suggest a theoretically acceptable simplification or, on the contrary, making it clear that the conditions were so severe that they were unlikely to be satisfied! If there **is** no representative consumer, then we are **forced** to confront the questions relating to the social welfare function -

explicitly, the weights to be allocated to different members or groups of the population - discussed in Sections 2.6 and 3.8. On practical grounds, it is likely that this will be done on the basis of relatively coarse assumptions.

It was noted in Section 2.7, while recalling McFadden's views on taste variation in the utility function, that as long as the AIRUM form was maintained, additional dummy variables could be added into the indirect utility functions to allow for variations between persons. In practical terms, this has normally been taken as a licence to estimate separate demand models for different segments, particularly in the case of journey purpose, without needing to consider the "weighting" consequences when aggregating up to an overall measure of benefit. While this is probably acceptable in most cases as a practical procedure, it might in fact be quite difficult to validate it within a general model estimation context.

However, of all aspects which may cause the utility function to vary between consumers/travellers, it is the income variation which is critical, because this underlies the basis in which changes in utility can be converted into money terms. Moreover, with increasing attention to **pricing** as a major topic in transport policy, variations in willingness to pay (as for example with proposals for tolled roads) are critical for demand forecasting. The upshot is that model estimation is going to be more directly focussed on income variation than was generally the case in the past. Against this, however, the fact that income **data** is usually collected in a relatively aggregate manner (eg, using a small number of income "bands") means that the opportunities for detailed formulation will still be restricted.

Hence, aside from the particular case addressed by Jara-Diaz *et al* in which the assumption of constant marginal utility of income for a particular **individual** (or group of individuals) is unreasonable, there would seem to be a practical solution along the lines discussed in Section 3.8. In other words, we should build demand models which allow explicitly for income variation in different population segments, but nonetheless maintain the AIRUM form: this appears to be what McFadden (1981) implied in the quotation cited at the end of Section 2.7 of this paper. This allows the benefit for each segment to be expressed in terms of **that segment**'s willingness to pay.

What is then required, over and above this, is an explicit grappling with the distributional consequences of variations in willingness to pay, principally because of variations in income. The practical treatment can range from, on the one hand, the "laissez-faire" assumption that the existing distribution is equitable, so that there is no need to apply any kind of weighting, to, on the other, some kind of re-distributive calculus which attempts to re-weight in favour of persons or groups with a higher marginal utility of income.

Since these are essentially political judgments, there would seem to be considerable virtue in making them explicit, rather than having them internalised in the mathematics of benefit calculations.

5 Summary and Conclusions

This paper has tried to cover a large amount of ground, and some topics haves perhaps been better covered than others. However, it has tried to link an impressive theoretical basis, which has been constructed over a period of more than 100 years, with the practical requirements of evaluating transport policies and schemes with the aid of (relatively conventional) transport models.

As implied at the outset, there is a danger in the “leading edge” departing too far from conventional practice. Although the Rule of a Half approximation, and the basic concept of Consumer Surplus, are simplifications of the underlying theory, they do retain substantial appeal, in terms of their relative ease of application, reasonable interpretatability, and acceptable general accuracy (certainly in the face of the general uncertainty which surrounds all forecasts!).

That said, there are certainly cases where the RoH is inappropriate, and there are others where there is a danger of it being misapplied. Both policy requirements and modelling ability is moving towards greater complexity, and this will bring further problems for evaluation in its wake. It is therefore timely to consolidate existing understanding, and it is hoped that this paper has made an initial contribution in this respect.

Bibliography

Arrow K J (1951), Social Choice and Individual Values, Wiley, New York

Bates J J (1992), Developing Choice Models to Deal with New Transport Alternatives, 6th World Conference on Transport Research, Lyons, France.

Bell M G H & Iida Y (1997), Transportation Network Analysis, John Wiley & Sons

Deaton, A and Muellbauer J (1980) Economics and Consumer Behaviour, Cambridge University Press, London.

Domencich T A and McFadden D (1975), Urban Travel Demand: a Behavioural Analysis, North Holland [reprinted 1997].

Dupuit J (1844) De la mesure de l'utilité des travaux publics, Annales des Ponts et Chaussées, 8 [English translation by RH Barback (1952) as 'On the Measurement of the Utility of Public Works', International Economic Papers, 2, pp83-110: also reprinted as "Public Works and the Consumer" in "Transport", ed D Munby, *Penguin Economics Series*, Harmondsworth.

EVA Consortium (1991), Evaluation Process for Road Transport Informatics, EVA-Manual, Project V1036, *DRIVE Programme*, Commission of the European Communities, **DGXIII**.

Galvez, T. and Jara-Diaz, S. (1998). On the social valuation of travel time savings. *International Journal of Transport Economics*, **25**, pp205-219.

Glaister S (1981), Fundamentals of Transport Economics, Basil Blackwell, Oxford

Gorman W M (1953) Community Preference Fields, *Econometrica*, **21**, pp 63-80

Gorman W M (1959) Separable Utility and Aggregation, *Econometrica*, **27**, pp 469-81

Gunn H F (1983) Generalised Cost or Generalised Time*?, Journal of Transport Economics and Policy*, **Vol xvii no. 1**.

Hensher, D.A. (1977) The Value of Business Travel Time. Pergamon Press, Oxford.

Hicks JR (1956) A revision of demand theory, Oxford University Press.

Hotelling H S (1935) Demand Functions with Limited Budgets, *Econometrica*, **3**, pp 66-78.

Hotelling H S (1938) The general welfare in relation to problems of taxation and of railway and utility rates, *Econometrica*, **6**, pp 242-269.

Jara-Diaz S R & Farah M (1988), Valuation of Users' Benefits in Transport Systems, *Transport Reviews*, **8**, pp 197-218

Jara-Diaz S R & Videla J I (1990), On the Role of Income in the Evaluation of Users' Benefits from Mode Choice Models, in Gérardin B (ed), *Travel Behaviour Research*, Gower, London

Jones I S (1977), Urban Transport Appraisal, Macmillan

Karlström A (2001), Welfare Evaluations in Non-Linear Random Utility Models with Income Effects, Chapter 22 in Hensher D A (ed), *Travel Behaviour Research – The Leading Edge*, Pergamon, 2001 (originally paper given at Ninth International Conference on Travel Behaviour, Gold Coast, Queensland, July 2000)

Karlström A (1998), Hicksian Welfare Measures in a Non-Linear Random Utility Framework, Department of Infrastructure and Planning, Royal Institute of Technology, Stockholm, Sweden

Mackie P J, Jara-Diaz S and Fowkes A S (2001), The Value of Travel Time Savings in Evaluation, *Transportation Research*, Part **E 37**, pp 91-106

Marshall A (1920) Principles of Economics, 8th edition, Macmillan

Martinez F J & Araya C (1998), Land-use impacts of transport projects: user benefits, rents and externalities, Paper presented to World Conference on Transport Research, Antwerp, July 1998

McFadden D (1981) Econometric Models of Probabilistic Choice, Chapter 5 in Manski C & McFadden D, eds, *Structural Analysis of Discrete Data with Econometric Applications*, MIT Press.

McFadden D (1999), Computing Willingness-to-pay in Random Utility Models, Chapter 15 in Moore J, Riezman R, & Melvin J (eds), *Trade, Theory and Econometrics: Essays in Honour of John S Chipman*, Routledge.

McIntosh P T & Quarmby D A (1970), Generalized Costs and the Estimation of Movement Costs and Benefits in *Transport Planning*, Department of Transport MAU **Note 179**, London.

MVA Consultancy *et al* (1987), The Value of Travel Time Savings, *Policy Journals*, Newbury, Berks, UK

Neuberger, H L I (1971): User benefit in the evaluation of transport and land-use plans. *Journal of Transport Economics and Policy*, **vol 5 no 1**.

Neuburger H & Wilcox J (1976), The Economic Appraisal of Land-Use Plans, *Journal of Transport Economics and Policy* vol. X no. **3**, pp 227–236.

Pearce D W and Nash C A (1981), The Social Appraisal of Projects, Macmillan.

Slutzky EE (1915) Sulla teoria del bilancio del consomatore, *Giornale degli*

Economist,i Vol **51**, pp 1-26 [English translation in Readings in Price Theory, GJ].

Small K A and Rosen H S (1981), Applied Welfare Economics with Discrete Choice Models, *Econometrica*, **49:1** pp 105-30.

Stone JRN (1954), Linear Expenditure Systems and Demand Analysis: an Application to the Pattern of British Demand, *Economic Journal*, Vol **64** pp511-27.

Sugden R (1999), Developing a Consistent Cost-Benefit Framework for Multi-Modal Transport Appraisal, *Report to UK Department for Transport*, University of East Anglia, Norwich.

Varian H R (1992), Microeconomic Analysis, 3rd Edition, Norton.

Williams H C W L (1977) On the Formation of Travel Demand Models and Economic Evaluation Measures of User Benefits, *Environment & Planning A*, Vol **9**, pp 289-344.

Williams H C W L & Senior M L (1978), Accessibility, Spatial Interaction and the Spatial Benefit Analysis of Land-Use Transportation Plans, in A Karlqvist, L Lundqvist, F

Snickars & J W Weibull (Eds): *Spatial Interaction Theory and Planning Models*, North Holland.

Willig R (1976), Consumer's Surplus without Apology, *American Economic Review*, **66**, pp 589-97

Appendix

A simple example about the Attribution of Benefit to specific alternatives

In this section we provide a simple example demonstrating the problems caused by attempting to attribute benefit in cases where there has been a change in travel demand. We consider a simplified situation where there are three choices facing travellers: these could relate equally to modes, destinations or routes. For the sake of simplicity we will assume that no new traffic is generated.

Suppose that in the before situation the generalized cost for choices 1, 2 and 3 are as the second column of the table below, while the demands are as in the third column. With these costs, there is no demand for choice no. 3.

Now we assume that an improvement is made to choice 3, which reduces generalized cost from 80 to 30: because this now attracts demand away from the other choices, we assume that, in the case of choice 2 only, there is also a reduction in generalized cost (as with congestion relief, for example). These new costs, and the corresponding demand, are shown in columns 4 and 5.

	Before		After			
Option k	costs C_k	demand T_k	costs C'_k	demand T'_k	ΔS	approx
1	30	70	30	25	–½(70+25). (30–30)	0
2	40	50	20	35	–½(50+35). (20 –40)	850
3	80	0	30	60	–½(0+60). (30– 80)	1500
* (Σ)		120		120		2350

First, we calculate the benefit components purely in terms of the source of the savings, using the basic consumer surplus formula $\Delta S = -0.5* (T' + T)(G' - G)$. For choice 1, there is no change in generalized cost, and hence no contribution to benefit. For choices 2 and 3, we obtain a contribution of 850 and 1500, respectively, as shown in the last column, giving a total benefit of 2350.

Suppose now that we felt able to assume that the 60 'new' travellers on choice 3 had been derived as follows: 45 from choice 1, and 15 from choice 2. This implies that the travellers remaining on choices 1 and 2 in the after situation have not changed. We then calculate the benefits as follows:

a. Remaining travellers on choice 1	: no benefit
b. Remaining travellers on choice 2	: 35x20 = 700
c. Travellers on choice 3:	
original	: none
changers from choice 1	: 0.5x45x[0+50]=1125
changers from choice 2	: 0.5x15x[20+50]=525
thus, for all choice 3	: 1650

giving a total of 2350, as before.

Note that changers get half the benefit appropriate to their original choice, and half the benefit for their final choice.

The implication is thus that of the total benefits of 2350, 700 accrue to final travellers on choice 2, and the remainder to choice 3.

Now suppose that we were somehow able to ascertain that what had in fact happened is that of the 45 travellers leaving choice 1, only 30 had in fact moved to choice 3, and the remaining 15 had moved to choice 2. For those originally on choice 2, 30 had moved to choice 3. Thus, while the final positions are the same (and hence the overall benefit is not affected), the 'paths' are different. This affects the attribution of benefits, in the following way.

a. Remaining travellers on choice 1 : no benefit

b. Travellers on choice 2:

original	: 20x20 = 400
changers from choice 1	: 0.5x15x[0+20]=150
thus, for all choice 2	: 550

c. Travellers on choice 3:

original	: none
changers from choice 1	: 0.5x30x[0+50]=750
changers from choice 2	: 0.5x30x[20+50]=1050
thus, for all choice 3	: 1800

giving a total of 2350, as before.

As a result of this new information about the more detailed movements of the various groups, the benefit associated with final users of choice 2 is reduced from 700 to 550, while that for choice 3 is increased correspondingly from 1650 to 1800.

But, in general modelling terms, there is no way in which this detailed information about changes can be made available. The conclusion is thus that the disaggregation of the overall benefit figure to the users of specific 'choices' is in general indeterminate.