,ranspn Res. Vol. 12, pp. 349–353
© Pergamon Press Ltd., 1978. Printed in Great Britain

0041-1647/78/1001-0349/$02.00/0

# THE GOODS/LEISURE TRADEOFF AND DISAGGREGATE WORK TRIP MODE CHOICE MODELS†

KENNETH TRAIN
Cambridge Systematics, Inc./West, 2161 Shattuck Avenue, Berkeley, CA 94704, U.S.A.

and

DANIEL MCFADDEN
University of California, Berkeley; Cambridge Systematics, Inc./West

(*Received* 24 *September* 1977; *in revised form* 22 *February* 1978)

Abstract—In disaggregate work trip mode choice models, the wage of the worker often enters as an explanatory variable. In some models, the cost of travel is divided by the worker's wage, while in other models the travel times are multiplied by the wage. This paper analyzes the use of wage in mode choice models and shows how different assumptions about the worker's indifference mapping between goods and leisure lead to different methods of entering wage. Using a Cobb–Douglas utility function $AG^{1-\beta}L^{\beta}$, where $G$ is goods and $L$ is leisure, it is shown that when $\beta = 0$ time should be multiplied by wage in the mode choice model and when $\beta = 1$ cost should be divided by wage. For values of $\beta$ between 0 and 1 a more general model obtains in which the value of $\beta$ is a parameter subject to estimation. Preliminary evidence indicates that the value of $\beta$ is between 0.7 and 1.0.

## 1. INTRODUCTION

In disaggregate work trip mode choice models, the wage of the worker often enters as an explanatory variable. In some cases (Train, 1976, for example) the cost of travel is divided by the worker's wage to reflect the presumption that a worker with a high wage is less concerned about cost than a worker with a low wage. In other cases (McFadden, 1974, for example) travel time is multiplied by the worker's wage to reflect the presumption that a worker with a high wage is more concerned with lost time than a worker with a low wage.

Two problems arise concerning the use of wage as an explanatory variable. First, it is not immediately evident that wage should be allowed to enter as an explanatory variable. McFadden (1973) has shown that the parameters being estimated in a disaggregate qualitative choice model can be interpreted as parameters of an expected utility function (called "representative" utility). However, in neoclassical theory of consumer behavior, a person's wage enters his constraint on utility maximization, not the utility function itself. Consequently, entering wage into "representative" utility requires justification.

Second, if wage is allowed to enter the model, one must decide how to enter it. The two approaches described above (dividing cost by wage and multiplying time by wage) are not equivalent in that they result in different estimated values of time and (as shown below) entail different assumptions about the shape of the consumer's utility function.

In the present paper, work trip mode choice models are derived from the neoclassical theory of the trade-off between goods and leisure. It is shown that allowing wage to enter "representative" utility is justified. Furthermore, it is shown that the shape of the consumer's indifference mapping for goods and leisure. determines the manner in which wage should enter "representative" utility. Finally, a general model is presented in which the shape of the indifference mapping is represented by a parameter which can be estimated rather than specified *a priori*.

## 2. THEORY

The object is to derive particular functional forms of "representative" utility from particular shapes of the consumer's indifference mapping for goods and leisure.

The general procedure is as follows. Choose some specific functional form for the utility function, $U(G, L)$, where $G$ is goods and $L$ is leisure. (That is, choose some specific shape of the consumer's indifference mapping for goods and leisure.) Assuming the price index to be constant and normalized to one, the following identities must hold:

$$G = V + w \cdot W - c$$

$$L = T - W - t$$

where $V$ = unearned income (given), $w$ = wage rate (given), $T$ = total amount of time (given), $W$ = number of hours worked (a continuous variable, non-negative), $c$ = cost of transportation to and from work (a discrete variable, which can assume values $c_1, \ldots, c_n$), $t$ = time of transportation to and from work (a discrete variable, which can assume values $t_1, \ldots, t_n$), $n$ = number of work-trip modes.

†This research was supported in part by National Science Foundation grant APR74-20392, Research Applied to National Needs program, National Science Foundation, and the Alfred P. Sloan Foundation grant 74-12-8, to the University of California, Berkeley.

The worker chooses the work-trip mode (and hence $c$ and $t$) and the number of hours worked so as to maximize $U$, subject to the identities.

Define:

$$G_i = V + w \cdot W - c_i$$
$$L_i = T - W - t_i$$
$$U_i = U(G_i, L_i).$$

Thus $U_i$ is a function of only the variable $W$. Since $W$ is continuous, $U_i$ can be maximized in the normal way, setting

$$\frac{\partial U_i}{\partial W} = U_{i1} w - U_{i2} = 0.$$

Or, as usual:

$$w = \frac{U_{i2}}{U_{i1}}. \tag{1}$$

Solve this for $W$ and call the solution $W_i^\dagger$. Substitute $W_i^\dagger$ into $U_i$ and call the resulting value $U_i^\dagger$.

Mode $i$ is chosen by the worker if and only if $U_i^\dagger > U_j^\dagger$ for all $j = 1, \ldots, n$; $j \neq i$. Let the expected value of $U_i^\dagger$ be labelled $Y_i$ (the source of variation in $U_i^\dagger$ is explained in Example E below). Furthermore, let $Y$ be the function (of $c$ and $t$) which assumes values $Y_i$, $i = 1, \ldots, n$. This $Y$ is the function used in mode choice models and called "representative" utility. For example, if the mode choice model is multinominal logit, then the probability of the worker choosing mode $i$ is

$$P_i = \frac{e^{Y_i}}{\sum_{j=1}^{n} e^{Y_j}}.$$

In this manner, one can determine what shapes of the indifference mapping for goods and leisure (that is, functional forms of $U(G, L)$) produce particular forms of $Y$.

An equivalent approach which is less heuristic but easier computationally is the following. Choose a specific $U(G, L)$ and derive the corresponding expenditure function:

$$E = E(U, w) \tag{2}$$

where $E$ is expenditures and prices are normalized to one. The following identity holds for expenditures when utility is maximized:

$$E = V - c + w(T - t). \tag{3}$$

Substitute (3) into (2) and solve for $U$. The solution is $U^*$, and its expectation is $Y$. The $Y$ obtained in this manner is the same or a monotonic transformation of the $Y$ obtained in the first approach.

Some specific examples follow.

Example A. Let $U = \alpha_1 \log G + \alpha_2 L$. With this function the derivative of the utility-maximizing $G$ with respect to income is zero: all extra income is absorbed in leisure. The resulting $Y$ is:†

$$Y = \alpha_1 \log\left(\frac{\alpha_1}{\alpha_2} w\right) + \alpha_2 T - \alpha_1 + \alpha_2 \frac{V}{w} - \alpha_2\left(\frac{c}{w} + t\right).$$

When comparing $Y_i$ and $Y_j$ ($i \neq j$), all terms which do not contain either $c$ or $t$ drop out. Thus, operationally, $Y$ is

$$Y = -\alpha_2\left(\frac{c}{w} + t\right). \tag{4}$$

In this case, cost is divided by wage rather than time being multiplied by wage. There is only one parameter since $c/w$ is in units of time, and the individual values work time and transportation time the same on the margin. The two variables can have different coefficients by specifying a model analogous to that in Example D in the appendix.

Example B. Let $U = \alpha_1 G + \alpha_2 \log L$. With this function the derivative of the utility-maximizing $L$ with respect to income is zero and all extra income is absorbed in goods. The resulting $Y$ is:

$$Y = \alpha_1\left(V + Tw - tw - \frac{\alpha_2}{\alpha_1} - c\right) + \alpha_2 \log \frac{\alpha_2}{\alpha_1} w.$$

Operationally,

$$Y = -\alpha_1(tw + c). \tag{5}$$

In this case, time is multiplied by wage rather than cost being divided by wage. Different parameters can be given to the two terms in a manner analogous to Example D in the appendix.

Example C. Let $U$ be a Cobb–Douglas utility function:

$$U = AG^{1-\beta}L^{\beta}, \qquad 0 < \beta < 1.$$

The expenditure function (with prices normalized to one) is:

$$E = Uk^{-1}w^{\beta}$$

where $k$ is a constant. Recall that

$$E = V - c + w(T - t).$$

Thus,

$$Uk^{-1}w^{\beta} = V - c + w(T - t).$$

and

$$Y = k(w^{-\beta}V - w^{-\beta}c + w^{1-\beta}T - w^{1-\beta}t).$$

†Note that in this example, and the next two examples, $Y_i$ equals $U_i^\dagger$ exactly since no variation in $U^*$ has been introduced. See the section "Generalization of Example C" below.

Operationally,

$$Y = -k(w^{-\beta}c + w^{1-\beta}t). \qquad (6)$$

When $\beta$ approaches 0, (6) becomes (5); and when $\beta$ approaches 1, (6) becomes (4). For values of $\beta$ between zero and one, the derivatives of the utility maximizing $L$ and $G$ with respect to income are greater than zero. The choice of $\beta$ is an empirical issue.

Generalization of Example C. Three elements are missing from Example C which prevent it from completely representing a mode choice model. First, mode choice models generally include various components of time, such as in-vehicle, wait, and walk times, with a different coefficient for each component. Similarly, various components of cost, such as out-of-pocket and "hidden" costs, might be included with separate coefficients. The analysis of Example C needs to be generalized to allow components of time and cost to enter "representative" utility. Second, mode choice models generally include socioeconomic variables other than wage and mode-specific constants. The analysis needs to be generalized to allow for these variables. Third, no variation in $U_1^*$ was specified in Example C. Yet, without variation in $U_1^*$, the mode choice model is deterministic rather than probabilistic, and $Y$ is not actually "representative" utility in a mode choice model.

Examples D and E in the appendix generalize Example C so as to account for these three elements of mode choice models. The notation in these examples is cumbersome, but the results are straightforward:

(1) Different components of time and cost can enter the "representative" utility function with different coefficients. Each time component is multiplied by $w^{1-\beta}$ when it enters "representative" utility and each cost component is divided by $w^{\beta}$, where $\beta$ is the parameter of the Cobb–Douglas utility function (defined in Example C).

(2) Socioeconomic variables and mode-specific constants can enter the "representative" utility function. These variables are to be interpreted as proxies for unobserved cost or time components. If a socioeconomic variable or a mode-specific constant is a proxy for an unobserved time component, then the variable is multiplied by $w^{1-\beta}$ when it enters "representative" utility. However, if the variable is a proxy for an unobserved cost component, then it is divided by $w^{\beta}$.

(3) An error term can be introduced into the analysis so that the model is probabilistic. This error term is interpreted as the difference between the unobserved cost and time components and the proxies which are used for these unobserved variables.

## 3. ESTIMATION

The existing estimation procedures only allow estimation of the parameters of "representative" utility functions which are linear in parameters. In Example E, $Y$ is linear in parameters for a given value of $\beta$, but not for $\beta$ treated as a parameter. Therefore, a search method is used in estimation. That is, $\beta$ is set at some value between zero and one, and the remaining parameters are estimated with existing estimation procedures. This process is repeated for various values of $\beta$ between zero and one. The value of $\beta$ which results in the largest value of the likelihood function is the estimate of $\beta$, and the estimates obtained with this $\beta$ for the other parameters are the final estimates of these parameters.

Unfortunately, estimation of a mode choice model with socioeconomic variables and mode-specific constants is not completely straightforward. As stated above, socioeconomic variables and mode-specific constants which are proxies for unobserved time components are multiplied by $w^{1-\beta}$, whereas those which serve as proxies for unobserved cost components are divided by $w^{\beta}$. Therefore, it is necessary that the researcher identify which socioeconomic variables and mode-specific constants are to be considered proxies for unmeasured time components and which are to be considered proxies for unmeasured cost components. The choice of which way to treat a socioeconomic variable or mode-specific constant affects the likelihood function and hence the estimates of $\beta$ and the other parameters.

The proper procedure is to enter each socioeconomic variable twice, both multiplied by $w^{1-\beta}$ and divided by $w^{\beta}$. In most cases, however, collinearity would prevent estimation of such a model. Consequently, the researcher must simply try various divisions of the socioeconomic variables and mode-specific constants and choose the model which is "best" by some loose criterion.

## 4. EVIDENCE

To obtain an indication of the value of $\beta$, work trip mode choice models were estimated, based on the generalization of Example C.† A sample of 161 workers was drawn from the East Bay area of the San Francisco Bay area. Each worker was considered to have a choice among four modes of travel to work: drive alone, carpool, bus with walk access, and bus with auto access. (No rail alternatives were included since the sample was taken before the rapid rail transit system, BART, was operating.) The mode which the worker actually chose was observed, and mode choice models were estimated to "explain" the choices. Several time and cost variables were entered as explanatory variables as well as socioeconomic variables and mode-specific constants. All of the time components were multiplied by $w^{1-\beta}$ and all of the cost components were divided by $w^{\beta}$.

Several divisions of the socioeconomic variables and mode-specific constants (into those which are proxies for unmeasured time components and those which are proxies for unmeasured cost components) were tried. When all of the socioeconomic variables and mode-specific constants were treated as proxies for unmeasured time components, the $\beta$ which obtained the highest value of the likelihood function was 1.0. In this model, the cost was divided by wage and all other variables did not interact with wage. When the socioeconomic and mode-

†These models and the sample upon which they are based are described more fully in Train and McFadden (1976).

specific constants were treated as proxies for unobserved cost components, the estimated $\beta$ was approx. 0.7. For other divisions of the socioeconomic variables and mode-specific constants, the estimated $\beta$ was between 0.7 and 1.0.

The model with the highest log likelihood was that in which all the socioeconomic variables and mode-specific constants were treated as proxies for unobserved cost components and $\beta$ was 0.7. However, the difference in log likelihood among the models based on various divisions of the socioeconomic variables and mode-specific constants was so small that choice among the models is arbitrary.

## REFERENCES

McFadden D. (1973) Conditional logit analysis of qualitative choice behavior. In *Frontier in Econometrics* (Edited by P. Zarembka). Academic Press, New York.

McFadden D. (1974) The measurement of urban travel demand. *J. Public Economics* 3, 303–328.

Train K. (1976) Work trip mode split models: an empirical exploration of estimate sensitivity to model and data specification. *Working Paper No.* 7602, Urban Travel Demand Forecasting Project, Institute of Transportation Studies, University of California, Berkeley.

Train K. and McFadden D. (1976) The measurement of urban travel demand IIA. *Working Paper No.* 7611, Urban Travel Demand Forecasting Project, Institute of Transportation Studies, University of California, Berkeley.

## APPENDIX

*Generalizations of Example* C

Example D. The analysis in Example C can be extended so that the terms in (6) have different coefficients and components of time and cost enter each with different coefficients. Let $U$ be the same as in Example C. The definitions of goods and leisure are respecified:

$$\text{"effective" leisure } L_i = \theta_{0i}T - W - \sum_{j=1}^{M} \theta_j t_i^j$$
$$\text{"effective" goods } G_i = \gamma_{0i}V + wW - \sum_{j=1}^{N} \gamma_j c_i^j \tag{7}$$

where $t^j$ is the $j$th component of travel time (say, on-vehicle time), $c^j$ is the $j$th component of travel cost, $M$ is the number of time components, $N$ is the number of cost components, $\theta_j$ is the psychometric weight attached to a minute in travel component $j$, in work time units, $\theta_{0i}$ is the psychometric weight attached to the total time budget, in work time units, when mode $i$ is used, $\gamma_j$ is the psychometric weight attached to a travel cost component, in wage income units, $\gamma_{0i}$ is the psychometric weight attached to non-wage income, in wage income units, when mode $i$ is used.

The $\theta_j$ and $\gamma_j$ reflect the relative onerousness or burden of different time or expenditure activities (associated, for example, with the exertion, fatigue, or bother involved). The parameters $\theta_{0i}$ and $\gamma_{0i}$ reflect the value of added units of time or income (again evaluated in working units). They may differ with mode if choice of mode itself affects the types of consumption and leisure activities available to the consumer. "Effective" income is:

$$E_i = (\gamma_{0i}V + w\theta_{0i}T) - \sum_{j=1}^{N} \gamma_j c_i^j - \sum_{j=1}^{M} \theta_j w t_i^j$$

Substituting this expression into the expenditure function for Example C and solving for $U$ obtains:

$$Y_i = k\left(\gamma_{0i}Vw^{-\beta} + \theta_{0i}w^{1-\beta}T - \sum_{j=1}^{N} \gamma_j c_i^j w^{-\beta} - \sum_{j=1}^{M} \theta_j t_i^j w^{1-\beta}\right)$$

This formula was used assuming the $\gamma_{0i}$ and $\theta_{0i}$ constant in $i$, so that operationally:

$$Y = -k\sum_{j=1}^{N} \gamma_j c^j w^{-\beta} - k\sum_{j=1}^{M} \theta_j t^j w^{1-\beta}. \tag{8}$$

Example E. Two elements are missing from Example D which prevent it from completely representing a mode choice model. First, mode choice models generally include socioeconomic variables other than wage and mode specific dummy variables in the "representative" utility function. The analysis needs to be generalized to allow for these variables. Second, no variation in $U_i^*$ has been specified. Yet, without variation in $U_i^*$, the mode choice model is deterministic rather than probabalistic, and $Y$ is not actually "representative" utility in a mode choice model. These two problems are confronted in this example. Redefine "effective" goods and leisure as:

$$\text{"effective" leisure } L_i = \theta_{0i}T - W - \sum_{j=1}^{M} \theta_j t_i^j - \sum_{j=1}^{P} \theta_{j+M} u_i^j$$
$$\text{"effective" goods } G_i = \gamma_{0i}V + wW - \sum_{j=1}^{N} \gamma_j c_i^j + \sum_{j=1}^{Q} \gamma_{j+N} v_i^j$$

where $u_i^j$ is the $j$th *unmeasured* time component of travel by mode $i$; $v_i^j$ is the *unmeasured* consumption of "good" $j$ in traveling by mode $i$.

An example of an $u^j$ is the length of time which one would usually arrive early to work so as not to be late for work if one's travel mode were delayed. If this variable were measured, it would assume a different value for each alternative, since the probability of being delayed a certain length of time varies across modes. An example of a $v^j$ is the privacy which one "consumes" in a particular travel mode.

Since the variables are unmeasured, they are approximated by other, measured variables, such as socioeconomic variables:

$$\sum_{j=1}^{P} \theta_{j+M} u_i^j = \sum_{j=1}^{R} \delta_j x_i^j + \epsilon_i$$
$$\sum_{j=1}^{Q} \gamma_{j+N} v_i^j = \sum_{h=1}^{S} \eta_j y_i^j + \mu_i \tag{9}$$

where $x_i^j$ is the $j$th measured variable used to approximate the unmeasured time components of travel by mode $i$; $y_i^j$ is the $j$th measured variable used to approximate unmeasured consumption from travel by mode $i$; $\delta_j$ and $\eta_j$ are parameters; and $\epsilon_i$ and $\mu_i$ are errors which allow eqns (9) to hold exactly rather than approximately.

Substituting (9) into the definitions of "effective" leisure and goods and solving for $U^*$ as in Example D obtains:

$$U^* = -k\sum_{j=1}^{N}\gamma_j c^j w^{-\beta} - k\sum_{j=1}^{M}\theta_j t^j w^{1-\beta} - k\sum_{j=1}^{S}\eta_j y_i^j w^{-\beta} - k\sum_{j=1}^{R}\delta_j x_i^j w^{1-\beta} - k(\epsilon+\mu). \quad (10)$$

For logit analysis, it is assumed that $-k(\epsilon+\mu)$ is distributed Weibull in the population and is independent across alternatives. Therefore, $Y$ equals the right hand side of (10) excluding the term $k(\epsilon+\mu)$, and the probability of choosing mode $i$ is:

$$P_i = \frac{e^{Y_i}}{\sum_{j=1}^{n} e^{Y_j}}.$$