Transpn. Res.-B, Vol. 25B, No. 5, pp. 341-350, 1991
Printed in Great Britain.

0191-2615/91 $3.00+.00

INCOME AND TASTE IN MODE CHOICE MODELS: ARE THEY SURROGATES?

SERGIO R. JARA-DÍAZ*
Departamento de Ingenieria Civil, Universidad de Chile, Casilla 228-3, Santiago, Chile

(*Received* 12 *April* 1989; *in revised form* 15 *August* 1990)

Abstract — The microeconomic foundations of mode choice models postulate modal utilities which are additive in income; this actually makes choice independent of this variable. On the other hand, it has been argued that income is correlated with variables that reflect taste and therefore, has a place in the utility specification as a proxi for taste. In this paper we propose a framework based on a generalization of our expenditure rate approach in order to explore the presumptive relation between income and taste empirically. We use data from Santiago, Chile, and the results suggest that the use of income may not be adequate to identify taste differences.

INTRODUCTION

Microeconomic analysis of consumer behavior is based on the presumption that individuals make choices which reflect their preferences among alternatives. Within this framework, the concept of *direct utility* (U) is used to represent the preferences for different levels and mixes of consumption; it does not depend on either prices or income. Observed behavior, however, reveals these preferences under a constraint, namely that total expenditure is limited by available income. Combining direct utility and the income constraint leads to the derivation of the *indirect utility function* (V) which takes account of preferences, prices, and income.

Discrete mode choice models require the estimation of parameters which belong to indirect utility functions. Thus, they intend to capture the influence of taste, through those parameters that help representing a system of preferences in U, and the influence of prices, income, and modal characteristics (that are one of the particular features of the discrete choice approach). The so-called socio-economic characteristics are usually introduced to account for differences within the population (e.g. age, sex, activity, etc.). After McFadden (1981), the assumptions behind his AIRUM (additive in income random utility model) structure have been implicitly or explicitly adopted by practitioners in mode choice modelling; this structure "yields choice probabilities that are independent of current income. However, tastes . . . 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 probabilistic choice system" (McFadden, 1981, p. 210). In other words, under these assumptions income is not purchasing power, but a surrogate for taste. Following this approach, the literature is absolutely dominated by income dependent specifications that have been discussed elsewhere. Some of them include a cost/income variable after Train and McFadden (1978); however, in that article the actual variable is cost/wage rate, with the wage rate explicitly defined as an exogenous variable such that the individual actually chooses his/her income by freely adjusting working hours. In this case, *income* is endogenous and does not belong into the indirect utility function. For specific discussions on the role of income in the specification of indirect utility in mode choice models, the interested reader can refer to Viton (1985), Hau (1987), and Jara-Díaz and Videla (1990a). The main idea in this literature is that the so-called modal utility reflects the level of utility that would be reached if that particular mode *was* chosen. Then, a conditional income variable appears which reflects purchasing power *after* transportation expenditure. As an indirect utility function must fulfill some analytical properties that link prices, utility, and income (like Roy's identity), a modal utility specification including income should meet these properties.

*The first version of this paper was written while the author was Visiting Associate Professor at MIT's Transportation Systems Division.

In this paper we want to analyze the relation between taste and income. The main question is whether it is reasonable to postulate that both variables are correlated. After all, income might be presented as a good proxy for lifestyle, type of neighborhood, or simply personal values, and these in turn are factors that contribute to generate a preference system as defined earlier. However, this is something that should be subject to empirical testing, not only for the sake of proper modelling and better forecasting, but also because of the important role of income in the welfare analysis that is consistent with mode choice; if the "taste-correlate" excuse is used, but the actual role of income is that of purchasing power, the analytics of welfare might change substantially (see Small and Rosen, 1981; Jara-Díaz and Videla, 1989; 1990b).

In the following section a theoretical framework is developed, specifically designed to allow distinguishment between pure preferences (direct utility) and the effect of income in the specification of modal utility. Next, the framework is applied to two corridors in Santiago, Chile, using income segmentation within each one; the results show no relation between income and taste. The final section contains a synthesis, conclusions, and directions for further research.

THE GENERALIZED EXPENDITURE RATE MODEL

Here we will present a model which generalizes the reformulation of Train and McFadden's (1978) goods-leisure trade off approach, developed by Jara-Díaz and Farah (1987) and empirically tested by Jara-Díaz and Ortúzar (1989). An individual whose income, I, and (paid) working hours, W, are fixed, maximizes utility, U, as a function of goods consumed, G, and "free" time, L (leisure) by appropriately choosing a transportation mode from an available choice set M. Each mode is described by its cost c_i and travel time t_i. If U is given a generalized Cobb-Douglas form, the problem can be stated as

$$\underset{G,L}{\text{Max}}\ U = A\, G^{\gamma} L^{\beta} \tag{1}$$

subject to

$$G + c_i = I \tag{2}$$

$$W + L + t_i = T \tag{3}$$

$$i \epsilon M \tag{4}$$

where T is the reference period and money is measured in G units. A is a scale factor (thus, it plays no role in the solution); γ and β represent preference for goods and leisure, respectively. As utility increases with both G and L, but not at an increasing rate, both γ and β belong to the interval (0,1].

The (conditional) indirect utility function V_i can be obtained by simply solving for G and L in constraints (2) and (3), and plugging the result into U. A first order approximation of this function yields a statistically tractable modal utility specification; this requires the evaluation of first derivatives of $U(G, L)$ at some point, which we choose to be $G = I$ and $L = T - W$. Note that this is equivalent to evaluate these derivatives at $c_i = t_i = 0$ after replacing the constraints. Formally,

$$\left.\frac{\delta U}{\delta G}\right|_{(I,T-W)} = A\gamma I^{\gamma+\beta-1} g^{-\beta}; \qquad \left.\frac{\delta U}{\delta L}\right|_{(I,T-W)} = A\beta I^{\gamma+\beta-1} g^{1-\beta}, \tag{5}$$

where $g = I/(T - W)$ is an expenditure rate, (Jara-Díaz and Farah, 1987). Using the budget and time constraints (2) and (3), we obtain

$$G - I = -c_i \quad \text{and} \quad L - (T - W) = -t_i. \tag{6}$$

Using eqn (5), a linear approximation of U around $(I, T - W)$ can be obtained. Introducing eqn (6) in that approximation, directly yields the conditional indirect utility function

$$V_i = A\, I^{\gamma}(T - W)^{\beta} + A\, I^{\gamma+\beta-1}\,(-\gamma\, g^{-\beta}\, c_i - \beta g^{1-\beta}\, t_i), \tag{7}$$

which has income, available time, and modal characteristics as arguments. This specific form of the expansion will be shown to be particularly useful for our purposes. First, note that V_i is conditional on the choice of mode: it gives the utility level that would be reached if mode i was chosen. This choice is determined by the problem that takes into account constraint (4), i.e.

$$\underset{i}{\text{Max}}\;\; V_i / i \epsilon M. \tag{8}$$

By inspection of eqn (7), the only term that influences choice (i.e. that determines the maximum V_i) is precisely the expression that depends on the modal characteristics. We will name $\overline{V}_i$ this term, given by

$$\overline{V}_i = -\gamma\, g^{-\beta} c_i - \beta g^{1-\beta} t_i. \tag{9}$$

Expression (9) has an important property: income in g can be measured for any time unit that contains a full cycle of leisure and work (a week, a month), as long as $T - W$ is consistently treated. If income was left as such (I), its units should be comparable with those of c_i. To obtain this result, however, a scale term multiplying $\overline{V}_i$ in eqn (7) needs to be cancelled; note that as this term involves income, alternative specifications could have been obtained as well.

The preceding development suggests a stochastic specification of modal utility as

$$V(c_i, t_i, I) = \alpha_i + \delta g^{-\beta} c_i + \theta\, g^{1-\beta} t_i + \epsilon_i, \tag{10}$$

where the α_i's, δ, θ, and β are parameters to be estimated; ϵ_i is an error which, if identically and independently distributed Gumbel, generates logit probabilistic mode choices in the usual manner. Specification (10) and its theoretical counterpart (9) are a generalized version of our expenditure rate model (Jara-Díaz and Ortúzar, 1989) that has c_i/g and t_i as variables; this corresponds to $\beta = 1$ in the general case of eqn (9).

The generalized expenditure rate model (GERM) in eqn (10) is particularly well suited for the purpose of investigating a relation between income and taste, because it contains both type of variables. Taste is represented by β which belongs to the direct utility U and reflects the preference for leisure (time) relative to goods (money); the g variable contains income as such. Therefore, the effects of taste and income are captured separately in GERM. Using standard software, the parameters α_i, δ, and θ can be estimated for a given β, treating $g^{-\beta}c_i$ and $g^{1-\beta}\, t_i$ as variables. Then β can be varied between 0 and 1, and the model with the highest log-likelihood will give the best estimator for β.

Whether a relation between income and taste exists or not in practice, can be studied estimating models for different income strata within otherwise similar populations (e.g. same socio-economic environment, location of residence, place of work, etc.). Then we can compare the values of β across different populations with similar income, or across income groups within the same population. If I actually represented taste, a monotonic relation with β should emerge; this is exactly the type of empirical analysis to be presented in the next section.

DATA AND RESULTS

The data set used refers to trips to work at the CBD in Santiago, Chile, with origin in either of two corridors that present markedly different income distributions. The information was gathered directly from 1354 individuals interviewed at their (stable)

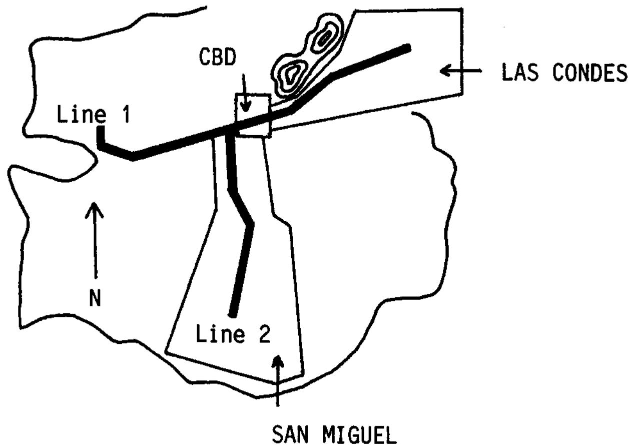

Fig. 1. Location of both corridors.

workplaces, 50% living on each corridor. The surveys were designed and conducted by Dr. Juan de Dios Ortúzar of the Chilean Catholic University during 1983 and 1985; a detailed description can be found in Ortúzar and Espinosa (1986).

Santiago is a city of roughly 4.5 million people, served by a fairly dense network of privately owned bus lines and shared taxis. Two subway lines intersect nearly at the CBD, one running east-west and the other running towards the south (with a short northern branch added during 1988). The bus fare is flat, unlike that of the share taxis which is unique within a line, but varies across lines. At the time of the surveys the subway fare was flat for movements within a line, with line 2 charging half of line 1; a transfer from line 2 to line 1 (equal to the difference) had to be paid, making a total about 20% less than the bus fare. The Las Condes corridor runs along the eastern branch of the subway's line 1, and includes the two wealthiest districts of the city, which also have the highest auto ownership rate in the country; average family income is 50% higher than the average of the richest 10% of Chilean families, and there are no families with incomes below the national average. On the other hand, the San Miguel corridor is served by the southern branch of the subway (line 2), a zone which can be classified as middle income as 30% of the individuals in the sample has a family income equal to or less than the national average and about 10% of the population falls within the highest income decile at a national level. Figure 1 shows the location of the corridors in space and Figure 2 gives an

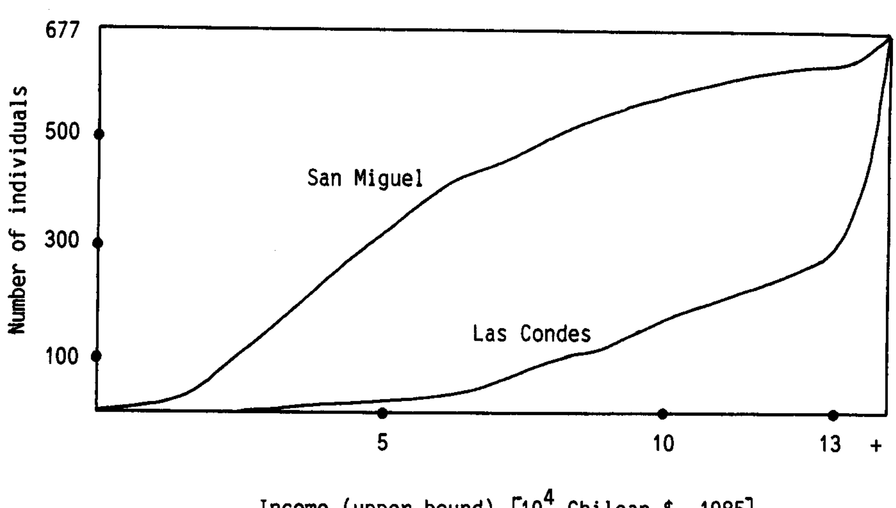

Fig. 2. Cumulative distribution of income, both corridors.

idea of income distribution. The public/private modal split (%) was 55/23 in Las Condes and 86/12 in San Miguel, the remainder being some form of public-private combination.

Taking into account the number of observations and the relative income distribution (both samples), four subsamples were created:

(i) Las Condes, high income (LCH), $FI > 130$; 398 individuals
(ii) Las Condes, middle income (LCM), $50 < FI < 130$, 266 individuals
(iii) San Miguel, middle income (SMM), $50 < FI < 130$; 297 individuals
(iv) San Miguel, low income (SML), $FI < 50$; 317 individuals

where *FI* is family income in thousand Chilean \$/month. Thirteen individuals with *FI* less than 50 in Las Condes, and 63 with *FI* greater than 130 in San Miguel were excluded.

Multinomial logit models were estimated for each subsample, specifying modal utility as in eqn (10) with the three usual components of travel time. Two socio-economic indicators were also included. The variables were defined as follows:

IV	:	in-vehicle travel time (minutes) times $g^{1-\beta}$
WK	:	walking time (minutes) times $g^{1-\beta}$
WT	:	waiting time (minutes) times $g^{1-\beta}$
COST	:	travel cost (Ch\$) times $g^{-\beta}$
CARLIC	:	number of cars per household driving licenses, with a limit of 1.
SEX	:	1 for men, 0 for women in car passenger modes and share taxi.

Mode specific constants are: CAR D (car driver), CAR P (car passenger), S TAXI (share taxi), SUBWAY, CAR D-S (car driver-subway), CAR P-S (car passenger-subway), S TAXI-S (share Taxi-subway), BUS-S (bus subway).

Bus was taken as the reference mode. The expenditure rate, g, was calculated using a reference period of a week or a month (i.e. a complete cycle work-leisure) and converted into Ch\$/minute. As explained in Jara-Díaz and Ortúzar (1989), income in g is the amount earned that can be spent; thus, both family income (*FI*) and family income per capita (*FIC*) can be used to calculate g, unlike the case of wage rate models, the logic of which justifies only individual income (what the individual "loses" due to one more minute travelling).

A number of models were run including all the possibilities in terms of the described socio-economic variables and income specifications. The values of β reported in Table 1, along with the estimated coefficients, correspond to those models that have the correct signs and the maximum likelihood within a class of comparable models (i.e. same or similar specifications).

Based upon previous models (Jara-Díaz and Ortúzar, 1989) we expected *FIC* to perform better than *FI* in general, as *FIC* reflects the amount of money available for one individual in a better way; but in the high income group, family size seems to have little impact on the perception of transportation cost. In terms of socio-economic variables, the inclusion of SEX was decided after an analysis in terms of actual choice relative to mode availability for both men and women in every group. The variable CARLIC was thought to be relevant for those segments in which car ownership was important. The results confirmed *a priori* expectations. Accordingly, SEX was relevant only in the high income segment, and neither SEX nor CARLIC were finally included in the low income group. *FI* performed slightly better in LCH. These are the results reported in Table 1; the values obtained for β deserve some additional discussion.

There are two ways of looking at the estimates of β: by comparison with other optimal values (max log-likelihood) corresponding to different specifications of variables, or analyzing the variation of the log-likelihood for the family of β's that originated the chosen value. From the first viewpoint, all optimal β's are fairly robust, as alternative

Table 1. Best models for selected segments†

Corridor Income Strata	Las Condes High	Las Condes Medium	San Miguel Medium	San Miguel Low
Beta	0.7	0.6	0.4	1.0
IV	−0.0677	−0.0873	−0.644	−0.0854
	(−4.1)	(−2.8)	(−1.2)	(−1.8)
WK	−0.1174	−0.1752	−0.0811	−0.0610
	(−7.2)	(−4.7)	(−2.3)	(−2.5)
WT	−0.2842	−0.1752	−0.4695	−0.1776
	(−2.1)	(−4.7)	(−3.4)	(−1.8)
COST	−0.0173	−0.0072	−0.0006	−0.0083
	(−2.3)	(−2.2)	(−0.2)	(−3.5)
CARLIC	2.57	2.04	1.89	—
	(4.7)	(2.7)	(2.4)	—
SEX	−0.7379	—	—	—
	(−2.4)			
CAR D	−2.131	−2.190	−2.423	−0.0855
	(−3.4)	(−2.8)	(−3.2)	(−0.1)
CAR P	−1.740	−2.415	−1.966	−3.164
	(−3.5)	(−5.2)	(−5.3)	(−4.7)
S TAXI	−1.241	−1.434	−2.093	−1.601
	(−2.7)	(−3.5)	(−4.7)	(−2.8)
SUBWAY	3.196	1.971	2.320	0.853
	(6.3)	(4.0)	(5.5)	(1.9)
CAR D-S	−1.593	−2.603	−3.282	—
	(−2.9)	(−3.6)	(−3.7)	—
CAR P-S	−0.488	−1.717	−0.959	−3.167
	(−1.3)	(−5.1)	(−2.0)	(−3.0)
S TAXI-S	−1.236	−1.821	−0.823	−0.901
	(−2.5)	(−3.9)	(−1.9)	(−1.9)
BUS-S	−0.423	−0.458	−2.103	−3.221
	(−0.1)	(−1.4)	(−3.5)	(−4.0)
$LL(0)$	−731.99	−460.02	−423.12	−377.82
$LL(\theta)$	−509.01	−370.80	−288.90	−191.86
ρ^2	0.3046	0.1939	0.3172	0.4922
% right	52.3	43.2	64.3	77.3
Sample size	398	266	297	317

†t-ratios are in parenthesis; the statistics describing the quality of each model are the usual ones (see Ortúzar, 1982).

specifications yielded the same (or very similar) optimal values. This is particularly clear in the case of SMM, where the 0.4 figure also gave the maximum likelihood for the family of models including SEX, and for those including SEX and CARLIC. Something similar occurs with SML, where the best value of β (1.0) is independent of the specification of income, as is the case with the LCM group, where $\beta = 0.6$. Lastly, all of our LCH models included SEX and CARLIC; the best alternative β (with *FIC*) was 0.8.

From the point of view of log-likelihood variation, the analysis can be better done with reference to Figs. 3 to 6, where we show $LL(\theta)$ as a function of β, for each member of the family of models reported in Table 1 (i.e. same variables for each segment). This is just a way to have an idea of the reliability of β for a given specification, as β was not estimated jointly with the other parameters, but provided as an exogenous value.† The gain in LL with respect to the minimum value, goes from one (SMM) to five (LCH); although the search was conducted using increments of 0.1 the best values of β emerge clearly from each figure.‡ The case of SMM in Fig. 5 seems to be an exception, but in fact the models for $0 < \beta < 0.3$ yield incorrect signs for the coefficients. Finally, if in all four cases we set a tolerance of 0.1 variation in LL with respect to the optimum, we

†Using a search process external to the estimation program for selection of β, results in an under-estimate of the standard error (overestimate of the t-ratio) for all other parameters.

‡We made a search varying β from 0.9 to 1.0 in 0.01 intervals for the SML segment; the reported 1.0 gives the maximum $LL(\theta)$. Note that this reduces specification (10) to our traditional expenditure rate model (Jara-Díaz and Ortúzar, 1989).

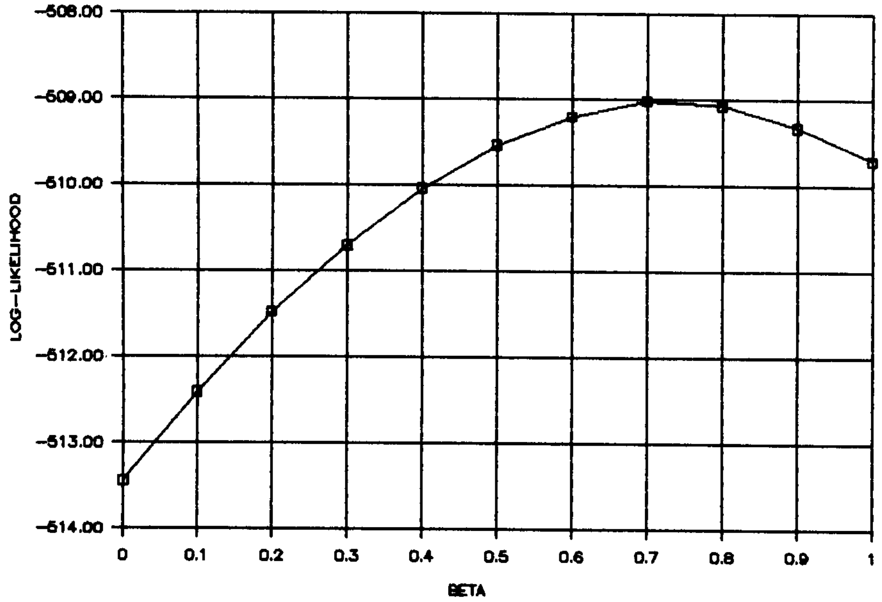

Fig. 3. Optimal beta value, LCH.

would accept the intervals in Table 2. Given the analysis presented in the previous paragraph (invariance of β with respect to the specification) and the shape of the $LL(\theta)$ curves, the values of β in Table 2 are the ones to be used for comparison.

The results highlight some interesting aspects regarding the perceptions of the corresponding population. As we have three income levels and two living environments (corridors), the analysis can be done better by pair of segments. In the poorer corridor (San Miguel), the two income groups present markedly different taste parameters; even if the extremes of the tolerance interval are chosen, the conclusion is the same (this is visually reinforced by the opposite shape of Figs. 5 and 6). Although the best estimates of β are also different between segments in the wealthier corridor (Las Condes), there is a common value of 0.7 within the tolerance interval in LL. The two middle income segments show

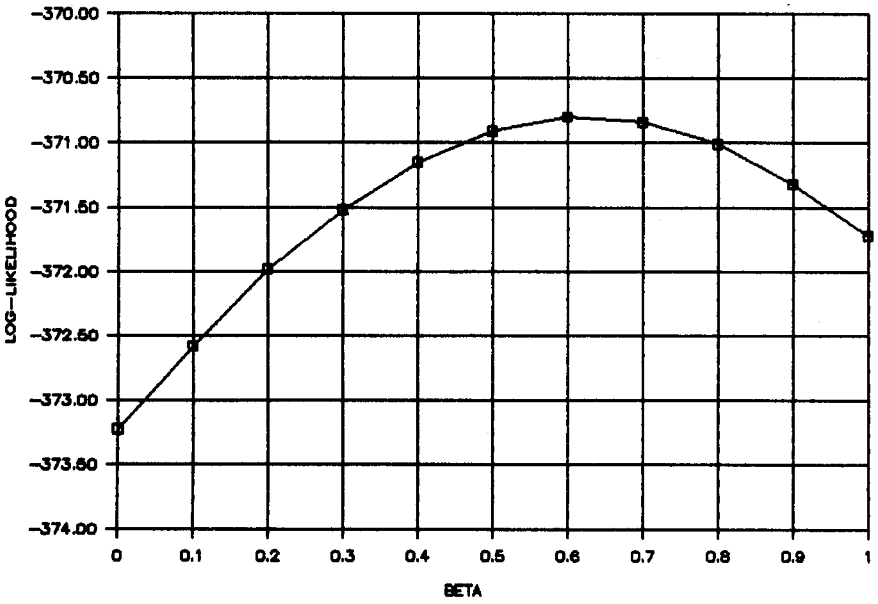

Fig. 4. Optimal beta value, LCM.

Fig. 5. Optimal beta value, SMM.

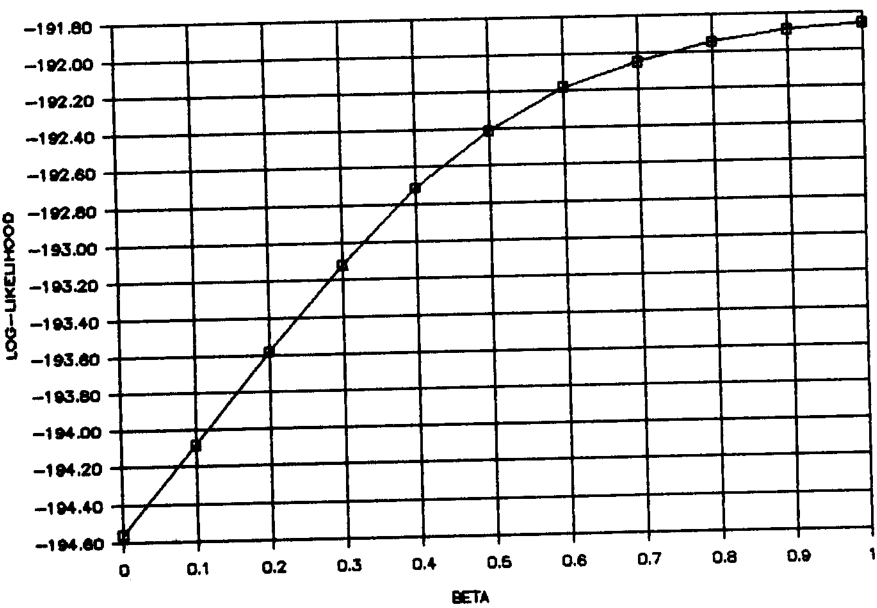

Fig. 6. Optimal beta value, SML.

Table 2. Values of β within a 0.1 tolerance in $LL(\theta)$

Corridor	Income	Optimum β	0.1 tolerance interval
Las Condes	High	0.7	0.7–0.8
Las Condes	Medium	0.6	0.5–0.7
San Miguel	Medium	0.4	0.4–0.6
San Miguel	Low	1.0	0.8–1.0

different optimal values of β (different tastes), but a clear overlapping of tolerance intervals is also present.

For synthesis, the best estimates of β are not monotonically related with income, and the analysis of tolerance intervals show:

(i) no clear taste differences between LCH and LCM;
(ii) a marked distinction in β between the two income groups in San Miguel;
(iii) common β values between the rich and poor groups; and
(iv) a clear overlapping of β intervals between the two middle income groups.

Thus, we could talk about a common preference structure (taste) in the rich corridor (Las Condes) but different taste parameters coexist in San Miguel. As an overall picture, the relative preference for leisure seems to increase towards the extremes. All these results tend to weaken the idea of a relation between income and taste.

SYNTHESIS AND CONCLUSIONS

We have developed a theoretical framework to generate a specification of modal utility that includes both a parameter related to direct utility (taste or preference parameter), and income. As income is a variable in the model and the parameter can be estimated, such a specification provides the necessary background to test the presumptive relation between taste and income argued by McFadden (1981) and implicitly adopted by most practitioners in seemingly income-dependent specifications.

We used the preceding framework to examine empirically the validity of using income as a socio-economic variable correlated with taste. A data set on trips to work originated in two corridors of Santiago, Chile, was manipulated to generate four segments (two income groups within each corridor). A taste parameter was estimated for each segment and no clear relation with income was found. In one corridor, no significant taste difference is found when income is taken into account through g. In the other case, a difference appears. Moreover, the extreme income groups (rich and poor) that belong to different corridors, exhibit similar preferences (β). All these findings suggest that use of income alone to identify taste differences of money versus time may not be adequate and sets the stage for further research to define a better representation of preferences. The theoretical framework has proved useful for our purpose, and the results tend to weaken current beliefs and practice.

It has not been our intention in this paper to fully explain or understand the motives behind people's mode choice behavior. We just wanted to verify what seemed to be a reasonable hypothesis. Our results, however, suggest that income and taste are different things. Preferences represented in the direct utility do not seem to receive a clear feedback from the constraint represented by income level. What people do is a matter of taste *and* income. At least, this seems to be the case for an heterogeneous middle class in a certain middle income, Third World country.

Acknowledgements—This research was partially funded by FONDECYT (Chile), the Andes Foundation, and the Fulbright Program. I am grateful to Rodrigo Parra for taking good care of the computational work, and to Jorge Videla for discussions. The detailed comments of the Associate Editor and both referees are greatly appreciated. As usual, I take full responsibility for the contents of the article.

REFERENCES

Hau T. (1987) Using a Hicksian approach to cost-benefit analysis in discrete choice: an empirical analysis of a transportation corridor simulation model. *Transpn. Res.*, **21B**, 339–357.

Jara-Díaz S. R. and Farah M. (1987) Transport demand and users' benefits with fixed income: the goods/leisure trade off revisited. *Transpn. Res.*, **21B**, 165–170.

Jara-Díaz S. R. and Ortúzar J. de D. (1989) Introducing the expenditure rate in the estimation of mode choice models. *J. Transport Economics and Policy*, **23**, 293–308.

Jara-Díaz S. R. and Videla J. (1989) Underestimation of users' benefits when income is misspecified in mode choice models. In Y. Okano et al. (Eds.) *Transport Policy, Management and Technology Towards 2001*, pp. 233–241, Western Periodicals, CA.

Jara-Díaz S. R. and Videla J. (1990a) On the role of income in the evaluation of users' benefits from mode choice models. In B. Gerardin (Ed.), *Travel Behaviour Research*, Gower, London.

Jara-Díaz S. R. and Videla J. (1990b) Welfare implications of the omission of income effect in mode choice models. *J. Transport Economics and Policy*, **24**, 83–93.

McFadden D. (1981) Econometric models of probabilistic choice. In C. Manski and D. McFadden (Eds.), *Structural Analysis of Discrete Data with Econometric Applications*, pp. 198–272, MIT Press, Cambridge, MA.

Ortúzar J. de D. (1982) Fundamentals of discrete multimodal choice modelling. *Transport Reviews*, **2**, 47–78.

Ortúzar J. de D. and Espinosa A. (1986) Influencia del ingreso y la tasa de motorización en la partición modal para el viaje al trabajo. *Tercer Congreso Latino-Iberoamericano de Investigación Operativa en Ingeniería de Sistemas*. Santiago, August.

Small K. and Rosen H. (1981) Applied welfare economics with discrete choice models. *Econometrica*, **49**, 105–130.

Train K. and McFadden D. (1978) The goods/leisure trade off and disaggregate work trip mode choice models. *Transpn. Res.*, **12**, 349–353.

Viton P. (1985) On the interpretation of income variables in discrete choice models. *Economic Letters*, **17**, 203–206.