# UNDERESTIMATION OF USERS' BENEFITS WHEN INCOME IS MISSPECIFIED IN MODE CHOICE MODELS

Sergio R. Jara-Díaz and Jorge I. Videla

Universidad de Chile
Casilla 228/3, Santiago, Chile

## 1. INTRODUCTION

Adequate demand modeling is important at least for two reasons: to have good estimates of the potential number of consumers or users after market variations, and to evaluate (in units that can be compared with costs) the eventual benefits or damages caused by such variations on consumers. In the case of mode choice models, their microeconomic formulation has been generated explicitly assuming that individual income does not influence mode choice, once a trip has been decided. Thus, such models do not presently consider income as a variable in the formal specification, including only fares and qualities of available modes, besides socioeconomic variables (different from income) that presumably capture individual specific aspects in mode choice.[1] As known, when there is no income effect the three measures of users' benefits (Marshallian consumer's surplus, equivalent and compensating variations) coincide; in the case of discrete choices, those benefits can be calculated from well established procedures (15), (9), (11) which are not valid when the income effect is present.

Lately, some interest has emerged in the literature for the understanding of the role of income in mode choice. Thus, Viton (13) uses some theoretical conditions on (indirect conditional) utility functions to analyze the internal consistency of possible specifications including income; Hau (1) develops models with income and uses a type of evaluation measure that can only be obtained by numerically inverting the expected maximum utility. Recently, Jara-Díaz and Videla (6) (8) have proposed new methods to calculate benefits from income sensitive mode choice models.

The objective of this paper is to apply and compare the different welfare measures calculated from mode choice models with and without income, estimated from the same population. In section 2 we summarize the measures; in section 3, data and models are presented. The measures are applied to those models assuming a variety of possible changes in modal costs; results are shown in section 4. The final section contains a synthesis and the main conclusions.

## 2. MEASURES OF USERS' BENEFITS: A SYNTHESIS

### 2.1. Direct Integration

The usual practice in the evaluation of users' benefits in transportation projects, has been to rely upon the rule-of-a-half to approximate the generalized version of the Marshallian consumer's surplus variation ($\Delta MCS$); such an approximation assumes small changes in fares and negligible second order effects in demand functions (14). In the last decade transportation welfare analysis has improved substantially, leading to analytical expressions which measure benefits exactly provided the income effect is null. The approaches by Williams (15) and Small and Rosen (11) respond to the general idea of directly solving the line integral representing $\Delta MCS$ (2)

$$\Delta MCS = -\int_{P^0}^{P^1} \sum_i X_i(P,I)dP_i, \tag{1}$$

where $X_i$ is the market demand for good i, $P^i$ is the price vector (state i), and I is income. When integrability conditions are fulfilled, a linear path of integration provides an operative way to solve equation (1); if such conditions are not met, the linear path can be used as an approximation (14). The discrete choice version of equation (1) was developed by Small and Rosen (11), explicitly assuming a negligible portion of income spent on transportation and a marginal utility of income $\lambda$ that is independent of model characteristics. If either approach is applied to the Logit specification, the increasingly popular log-sum measure of individual welfare $(1/\lambda)\,\mathrm{Ln}\,\Sigma \exp V_j$ is obtained, where $V_j$ is the (observed conditional indirect) utility function of mode j, and summation is overall available modes; $\lambda$ is calculated as the coefficient of cost in modal utility (for a full explanation of the relations among these measures, see (4)). It should be no surprise that the logarithmic part of the log sum measure is actually the value of the expected maximum utility (EMU) under the Logit specification; it can be shown to have very nice properties like fulfilling Roy's identity at an aggregate level. As known, $\Delta MCS$ is an exact measure of welfare when there is no income effect; however, equation (1) can always be chosen as an approximation when this is not the case.

## 2.2. Measures that Explicitly Consider the Income Effect

Assume that, somehow, income is specified in $V_i$. Then the EMU (which can be taken as representative) is a function of income, generally complex in analytical terms. If the EMU is viewed as an aggregate indirect utility function, inversion in I will generate an expenditure function which depends on all prices and utility, e(P,U), and either the compensating (CV) or equivalent (EV) variations could be calculated from it as a difference, e.g.

$$CV = e(P^0,U^0) - e(P^1,U^0). \tag{2}$$

This is exactly the procedure suggested by Hau (1), where all the steps are followed using numerical methods as analytical inversion is unfeasible. We will call CVH the value of CV obtained in this manner.[2]

We have developed another approach that explicitly accounts for the role of income (8). In essence, we approximate $e(P^1,U^0)$ using a second order Taylor expansion from $e(P^0,U^0)$, and make use of two important analytical properties of demand functions: the derivative property and the Slutsky equation. If $\pi_i$ denotes the probability of choosing mode i, the CV for each individual after a generalized change in modal costs, can be approximated as

$$CV \doteq -\sum_i \pi_i^0 \Delta P_i - \frac{1}{2}\sum_{ij} \frac{\partial \pi_j^0}{\partial P_i} \Delta P_i \Delta P_j - \frac{1}{2}\sum_{ij} \frac{\partial \pi_j^0}{\partial I} \pi_i^0 \Delta P_i \Delta P_j. \tag{3}$$

We have called the last term the Income Induced Welfare impact, IWI. Using the Logit model, equation (3) can be expressed in terms of partial derivatives of conditional (modal) utilities, and Roy's identity can be used either at a conditional level $(-\partial V_i/\partial P_i = \lambda_i)$ or under its stochastic form $(-\partial V_i/\partial P_i = \bar{\lambda})$, introduced by Hau (1). As results, we obtain either

$$CV \doteq -\overline{\Delta P}\left(1 + \overline{\lambda \Delta P} - \frac{1}{2}\bar{\lambda}\overline{\Delta P}\right) + \frac{1}{2}\overline{\lambda \Delta P^2} \equiv CVC, \quad \text{or} \tag{4}$$

$$CV \doteq -\overline{\Delta P}\left(1 + \frac{1}{2}\overline{\lambda \Delta P}\right) + \frac{1}{2}\bar{\lambda}\overline{\Delta P^2} \equiv CVS, \tag{5}$$

where the bar denotes expected value of the corresponding variable or function of variables, in the initial state. Note that if $\lambda_i$ was independent of modal qualities and cost (i.e. if $\lambda_i = \lambda, \forall i$), both measures would coincide.

It is very important to realize that in this last approach, careful calculation of $\lambda_i$ is required. This is not explicitly the case in the previous methods, since equation (1) is the result of solving a line integral in modal costs, and Hau's inversions in income yields an income independent equation. If one recalls the microeconomic foundations of discrete choices, the key variable is not income itself but disposable income after transportation $I - P_i$; since this is stated in terms of one trip, which income should be used to calculate $\lambda_i$ is unclear. Therefore, formulas (4) and (5) required ad hoc calculation of $\lambda_i$.

## 3. DATA AND MODELS

The data corresponds to trips to work at the CBD, originated in the San Miguel corridor in Santiago (Chile). It consists of 617 individual observations, each one containing the mode choice set, level of service characteristics, costs (fares), socioeconomic characteristics, and choice. The survey was done in 1985 (10). In Table 1, we have summarized information regarding availability and choice of mode, as well as family and individual income (averages) of users choosing each mode. The national mean family income in the same period was around 32,000 Chilean$/month; it is apparent that the auto modes are associated to high income, while bus is related to the lowest income.

Two types of model were estimated. One follows the framework developed by Train and McFadden (12); there, working hours can be adjusted and modal utility results in a specification given by

$$V_i = \alpha_i + \beta P_i/\omega + \gamma t_i \tag{6}$$

Here, $t_i$ represents travel time (or its components) and $\omega$ is the individual wage rate, usually approximated through individual income divided by working hours (W). The second model corresponds to the reformulation due to Jara-Diaz and Farah (3), where income and working hours are fixed; this results in

$$V_i = \alpha_i^1 + \beta' P_i/g + \gamma t_i, \tag{7}$$

where g is an expenditure rate, defined as income divided by non-working hours (disposable time). The logic behind this latter specification indicates that per capita family income should be used, since g represents disposable income to be spent in a reference period (for a discussion and an application of this theory, see (5)).

Both types of models were estimated using the (fairly traditional) three components of travel time (in vehicle, walking, waiting) and two socioeconomic variables: SEX (1 for men in modes 2, 3, and 8) and CARLIC (number of cars over number of driving licenses in the household) which reflects real access to car in modes 1 and 6. The estimated coefficients are shown in Table 2. Note that from these results, both models actually yield very similar statistical performance.

**Table 1. Modal split and income (Chilean $/month, 1985)**

| Mode | Description | % Available | % Choice | Fam. I | Indiv. I |
|---|---|---|---|---|---|
| 1 | car driver | 38.55 | 7.83 | 113359 | 84180 |
| 2 | car pass. | 39.00 | 4.28 | 105000 | 51241 |
| 3 | share taxi | 40.32 | 2.22 | 72200 | 50853 |
| 4 | subway | 39.29 | 27.33 | 70601 | 46994 |
| 5 | bus | 99.56 | 49.94 | 54387 | 33900 |
| 6 | 1 + 4 | 22.90 | 0.89 | 130000 | 65000 |
| 7 | 2 + 4 | 25.55 | 1.33 | 90000 | 46111 |
| 8 | 3 +4 | 58.08 | 6.20 | 67694 | 42230 |
| 9 | 5 + 4 | 56.87 | 0.89 | 76667 | 44000 |

**Table 2. Estimated Coefficients**

| Variable | Wage Rate Model | Expenditure Rate Model |
|---|---|---|
| IVT | -0.052(-1.76) | -0.0486(-1.66) |
| WALK | -0.0631 (-3.69) | -0.0692(-3.98) |
| WAIT | -0.2633(-3.78) | -0.2660(-3.78) |
| $P_i/\omega$ | -0.0341(-3.73) | |
| $P_i/g$ | | -0.0054(-4.12) |
| SEX | -1.4728(-4.41) | -1.4473(-4.34) |
| CARLIC | 1.6554(2.63) | 1.6594(2.63) |
| | | |
| Mode 2 | 0.2737(0.45) | 0.1534(0.25) |
| Mode 3 | 0.6498(1.06) | 0.5700(0.93) |
| Mode 4 | 3.5082(5.40) | 3.4009(5.22) |
| Mode 5 | 1.6851(2.97) | 1.5759(2.76) |
| Mode 6 | -1.4201(-2.85) | -1.4992(-2.99) |
| Mode 7 | 0.6801(0.92) | 0.5993(0.81) |
| Mode 8 | 0.9092(1.47) | 0.8651(1.40) |
| Mode 9 | -0.9615(-1.34) | -1.0234(-1.42) |
| $\rho^2$ | 0.060 | 0.063 |
| % right | 69.4 | 69.4 |

As explained before, ad hoc calculation of $\lambda_i$ is required for the approach based upon expansion of the expenditure function. This was done using the method fully explained in (7) with the results shown in Table 3.

Tables 2 and 3 contain what we need to experiment with the welfare measures previously described.

**Table 3. Marginal Utility of Income**

| Income Level ($10^3$ Chilean pesos/month) | $\lambda(P_j)$ | $\lambda(\bar{P})$ |
|---|---|---|
| I < 30 | 0.3740 - 0.00482 $P_j$ | 0.19788 |
| 30 < I < 50 | 0.1416 - 0.00120 $P_j$ | 0.09167 |
| 50 < I < 75 | 0.0413 - 0.00027 $P_j$ | 0.0300 |
| I > 75 | 0.00547 | 0.00547 |

## 4. USERS' BENEFITS FROM DIFFERENT APPROACHES

Direct integration ($\Delta MCS$), Hau's numerical method (CVH), and the two approximations (CVC and CVS), were programmed in order to allow for extensive experimentation. Given the models and the policies to be analyzed (cost changes), each welfare measure can be calculated for each individual, such that aggregate benefits can be obtained through complete sample enumeration. The experiments are divided in two groups: single mode cost changes, and simultaneous cost changes; these latter were designed based upon both the results of the former and our previous qualitative analytical discussion on the IWI (8).

In Table 4, two sets of equal single mode cost changes are presented. The first column of welfare results corresponds to direct integration of the wage rate model (no income effect), which means application of equation (1) directly, which results in the log-sum measure with $\lambda = \beta/\omega$. The second column refers to the same method applied to the expenditure rate model, which again is like using the log-sum measure, but now "$\lambda$" should be interpreted as $\beta'/g$. As evident, the three other columns refer only to the specifically income sensitive model. Three types of comparative analysis can be made on the results: relative values with the same method applied to different models, comparison of the different welfare measures applied to the same model, and comparison among CV results.

The first regularity observed is the pattern of differences between the Marshallian measures applied to the two models. Estimated benefits with the non-income-sensitive model are lower in the case of mode 5, which is not only the one associated with low income users but also the one that presents the highest share. For all other cases $\Delta MCS_1 > \Delta MCS_2$. Since price variations are equal and relatively small, the differences occur due to different demand predictions, i.e. the increase in bus users after a price drop is underestimated by the wage rate model, and the opposite occurs in the other cases.[3] This is intuitively correct, since an income sensitive model should predict a higher share for "low income modes" in this case. The second regularity is the nearly identical outcome between $\Delta MCS_2$ and CVH; this is consistent with Hau's own work, which reports negligible differences (1). This is suggestive of an eventual (hidden) property of the method, i.e. equality might hold always. Thirdly, both CVC and CVS are systematically greater than CVH; it should be recalled that the two approximations are based upon expansions from the base case and exogenously determined marginal utilities of income. However, yet a fourth regularity arises and should be mentioned here because of its intuitively appealing interpretation: CVS is greater than $\Delta MSC_1$ for all modes whose users have an average income below $90000, and the relation reverses for the relatively "wealthy" modes (1, 2 and 6); CVS is nearly equal to $\Delta MSC_1$ for the limiting income case (mode 7). This is saying that the traditional measure applied to the traditional model would underestimate the benefits caused by projects that favor those segments of the population that present income effect.

**Table 4. Users' benefits for a single mode cost change**

| Cost Change | $\Delta MCS_1$ | $\Delta MCS_2$ | CVC | CVS | CVH |
|---|---|---|---|---|---|
| $\Delta P_1 = -5$ | 122 | 72 | 52 | 93 | 72 |
| $\Delta P_2 = -5$ | 378 | 268 | 320 | 310 | 268 |
| $\Delta P_3 = -5$ | 265 | 234 | 243 | 275 | 233 |
| $\Delta P_4 = -5$ | 285 | 254 | 317 | 295 | 254 |
| $\Delta P_5 = -5$ | 1692 | 1981 | 2129 | 2101 | 1981 |
| $\Delta P_6 = -5$ | 43 | 26 | 31 | 32 | 26 |
| $\Delta P_7 = -5$ | 106 | 93 | 126 | 110 | 93 |
| $\Delta P_8 = -5$ | 268 | 245 | 270 | 285 | 245 |
| $\Delta P_9 = -5$ | 250 | 230 | 252 | 269 | 230 |
| $\Delta P_1 = -10$ | 250 | 147 | 68 | 229 | 146 |
| $\Delta P_2 = -10$ | 770 | 547 | 750 | 713 | 546 |
| $\Delta P_3 = -10$ | 543 | 480 | 314 | 646 | 480 |
| $\Delta P_4 = -10$ | 584 | 523 | 775 | 687 | 523 |
| $\Delta P_5 = -10$ | 3335 | 4013 | 4603 | 4494 | 4014 |
| $\Delta P_6 = -10$ | 88 | 53 | 72 | 77 | 53 |
| $\Delta P_7 = -10$ | 219 | 191 | 320 | 259 | 191 |
| $\Delta P_8 = -10$ | 556 | 309 | 607 | 668 | 509 |
| $\Delta P_9 = -10$ | 518 | 477 | 562 | 630 | 477 |

[1] Wage rate model
[2] Expenditure rate model

In Table 5 we have produced a set of cases dealing with simultaneous changes in model costs, where complexity increases as one moves from case 1 to 9. Some regularities can be detected. In all cases where $\Delta P_4, < 0$, it occurs that $\Delta MCS_1$ is less than $\Delta MCS_2$, which is very much in line with our first observation in Table 3, but now the effect of the accompanying cost changes can also be looked at; one can see that the relative difference between measures (i.e. the difference over $\Delta MSC_2$) increases towards those cases which include both advantages to high share - low income modes and disadvantages to low share - high income ones. A second regularity is the equality between $\Delta MSC_2$ and CVH, which holds in all these cases of simultaneous changes in costs, as was the case in Table 3. Similarly, CVC and CVS are systematically greater than CVH. But the most striking fact is the relatively large difference in values that the various measures can yield; they can even have different signs (cases 8 and 9), although only $\Delta MSC_1$ can be blamed for this.

**Table 5. Users' benefits for a simultaneous change in modal costs**

| | Cost Changes | $\Delta MCS_1$ | $\Delta MCS_2$ | CVC | CVS | CVH |
|---|---|---|---|---|---|---|
| 1. | $\Delta P_4 = \Delta P_5 = -5$ | 1967 | 2224 | 2330 | 2294 | 2224 |
| 2. | $\Delta P_4 = \Delta P_5 = -10$ | 3977 | 4488 | 4912 | 4491 | 4491 |
| 3. | $\Delta P_8 = \Delta P_9 = -5$ | 516 | 473 | 518 | 546 | 473 |
| 4. | $\Delta P_8 = \Delta P_9 = -10$ | 1062 | 975 | 1151 | 1264 | 975 |
| 5. | $\Delta P_8 = \Delta P_9 = -10, \Delta P_4 = -5, \Delta P_1 = \Delta P_2 = 10$ | 404 | 585 | 979 | 1189 | 584 |
| 6. | $\Delta P_4 = \Delta P_5 = +5, \Delta P_8 = \Delta P_9 = -10$ | -812 | -1158 | -719 | -624 | -1159 |
| 7. | $\Delta P_4 = \Delta P_5 = -5, \Delta P_8 = \Delta P_9 = +10$ | 1059 | 1395 | 1852 | 1942 | 1394 |
| 8. | $\Delta P_1 = \Delta P_2 = \Delta P_3 = -5, \Delta P_4 = \Delta P_5 = +1.8$ | 73 | -210 | -120 | -64 | -211 |
| 9. | $\Delta P_1 = \Delta P_2 = \Delta P_3 = +5, \Delta P_4 = \Delta P_5 = -1.8$ | -28 | 248 | 339 | 395 | 247 |

[1] Wage rate model
[2] Expenditure rate model

An interesting exercise is to order the "projects" according to the level of benefits for each welfare measure. The result is that, with the exception of $\Delta MCS_1$, <u>all</u> the measures yield the same ordering; this is indeed an important observation, as the form of computing benefits is certainly quite different. The Marshallian measure applied to the non-income sensitive model matches the ordering of benefits only in the extreme cases (2, 1, and 6); the evident reverse ordering in cases 8 and 9 makes even clearer the fact that $\Delta MCS_1$ seems to work against projects that favor low income users.

## 5. CONCLUSIONS

Measures of welfare having different theoretical foundations have been presented, applied and compared. After analyzing the results, it seems clear that the (generalized) Marshallian consumers' surplus measure applied to mode choice models that are not sensitive to income, underestimates the benefits caused by projects which particularly favor low income users. When income is better accounted for in a mode choice model, even the Marshallian measure captures benefits more properly; this is due to the difference in modal share forecasts, as income sensitive models predict higher shares for cheaper modes when their fares are reduced.

The method of numerically inverting in income the expected maximum utility yields the same figures as the Marshallian method. This should be theoretically investigated. All Hicksian compensating variation measures yield the same ordering of pricing projects, in spite of differences in the absolute level of benefits. This is a particularly nice result, as the approximations which include the income effect (CVS and CVC) are, in our experiments, practically model independent; they depend upon the shares in the base case (approximately equally reproduced by both models), and the value of the marginal utility of income for each individual (which are exogenously provided). As these measures use either mode specific or average marginal utilities of income, their differences are built-in from the beginning.

As a general recommendation, appropriate specifications of mode choice models including income should be investigated, not only for good demand forecasts, but also the appropriate evaluation of benefits, which might be otherwise biased against projects that favor low income users. Finding better measures, both in analytical and operational terms, is also needed.

## NOTES

1. As we have stated elsewhere, the "income" variable that sometimes appear in the specification of utility in applied mode choice analysis, is there as a proxy for either taste or wage rate, based upon the discussion of McFadden (1981) and Train and McFadden (1978), respectively.
2. Note that measures based upon income variations that are either equivalent to or compensatory of price changes, can be easily extended to quality changes as well. We have also developed another approach based upon the inversion in income of the stochastic conditional indirect modal utility, working out an expected minimum expenditure function (EME) which can be used to calculate CV both at an individual and population level (Jara-Diaz and Videla, 1987).
3. This can be intuitively grasped with the rule-of-a-half, where $\Delta MCS$ is approximated by $\frac{1}{2}\Delta Pi(\pi_i^0 + \pi_i^1)$. Since $\pi_i^0$ is similarly reproduced by both models, the discrepancy is due to $\pi_i^1$.

## ACKNOWLEDGEMENTS

This research has been supported by FONDECYT (Chile), the Fulbright Program, and the Andes Foundation. The paper was written while the first author was Visiting Associate Professor at the Transportation Systems Division of M.I.T.

REFERENCES

1. Hau, T.D. (1987), "Using a Hicksian Approach to Cost-Benefit Analysis in Discrete Choice: An Empirical Analysis of a Transportation Corridor Simulation Model," Transportation Research 21B, pp. 339-357.

2. Hotelling, H. (1938), "The General Welfare in Relation to Problems of Taxation and of Railway and Utility Rates," Econometrica 6, pp. 242-269.

3. Jara-Diaz, S.R. and Farah, M. (1987), "Transport Demand and Users' Benefits with Fixed Income: The Goods/Leisure Tradeoff Revisited," Transportation Research 21B, pp, 165-170.

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

5. Jara-Diaz, S.R. and Ortuzar, J. de D. (1988), "Introducing the Expenditure Rate in Mode Choice Models," Forthcoming in the Journal of Transport Economics and Policy.

6. Jara-Diaz, S.R. and Videla, J. (1987), "On the Role of Income in the Evaluation of Users' Benefits from Mode Choice Models," Proceedings of the 5th International Conference on Travel Behaviour, Gower, London (in press).

7. Jara-Diaz, S.R. and Videla, J. (1988a), "Detection of Income Effect in Mode Choice: Theory and Application," Forthcoming in Transportation Research B.

8. Jara-Diaz, S.R. and Videla, J. (1988b), "Welfare Implications of Income Effect Omission in Mode Choice Models," Submitted to the Journal of Transport Economics and Policy.

9. McFadden, D. (1981), "Econometric Models of Probabilistic Choice," In Ch. Manski and D. McFadden (eds.), "Structural Analysis of Discrete Data with Econometric Applications," MIT Press, Cambridge, MA.

10. Ortuzar, J. de D. and Espinosa, A. (1986), "Influencia del Ingreso y Tasa de Motorizacion en la Particion Modal para el Viaje al Trabajo," Third Latin-Iberoamerican Meeting of Operations Research and Systems Engineering, Santiago, Chile.

11. Small, K. and Rosen, H. (1981), "Applied Welfare Economics with Discrete Choice Models," Econometrica 49, pp. 105-130.

12. Train, K. and McFadden, D. (1978), "The Goods/Leisure Tradeoff and Disaggregate Work Trip Mode Choice Models," Transportation Research 12, pp. 349-353.

13. Viton, P. (1985), "On the Interpretation of Income Variables in Discrete Choice Models," Economic Letters 17, pp. 203-206.

14. Williams, H. (1976), "Travel Demand Models, Duality Relations and User Benefit Analysis," Journal of Regional Science 16, pp. 147-166.

15. Williams, H. (1977), "On the Formation of Travel Demand Models and Economic Evaluation Measures of User Benefit," Environment and Planning A9, pp. 285-344.