# Results of Discrete Choice Model for a Brazilian City

**Mateus Araújo e Silva, USP**
**Kay W. Axhausen, ETH**

**Conference paper STRC 2010**

**STRC** | **10th Swiss Transport Research Conference**
Monte Verità / Ascona, September 1 - 3, 2010

# Results of Discrete Choice Model for a Brazilian City

Mateus Araújo e Silva
Departamento de Transportes (STT)
Universidade de São Paulo
Escola de Engenharia de São Carlos
São Carlos / 13566-590 Brazil

Phone: +55 16 3373 9613
Fax: + 55 16 3373 9602
mateus.araujo@terra.com.br:

Kay W. Axhausen
Institute for Transport Planning
and Systems (IVT)
ETH Zurich
CH-8093 Zurich

Phone: +41 44 633 3943
Fax: + 41 44 633 1057
axhausen@ivt.baug.ethz.ch

April 2010

## Abstract

This paper examines the application of a random utility model to analyze human choice behavior among discrete alternatives from different modes of transport in a given set to a Brazilian medium sized city. Mode choice data collected in a 2002 RP survey in Uberlândia, Brazil, was used to obtain parameter estimates of utility functions, as well as to analyze trade-offs between observed attributes (including travel time and travel cost both based on geocoding origins and destinations). The calibration of the discrete choice model (multinomial logit model) has been carried out with the software BIOGEME version 1.8. Model results outline the subjective value of time, fare policy (discount and free rides) and household characteristics (auto ownership and household income).

This research was developed during the planned stay of the first author at the Institute for Transport Planning and Systems from the Swiss Federal Institute of Technology Zurich as a part of the first author DSce Thesis at the Universidade de São Paulo.

## Keywords

Discrete Choice Model, Mode Choice, Brazilian Traveler's Behavior

# 1. Introduction

The choices that individual make in the transport system produce aggregate outcomes that shape urban traffic conditions, access to economic opportunities and to particular service provisions in society.

This aspect leads to the conclusion that for analysis of many transportation-related policies, it is useful to know the change which a particular policy would induce in human choice behavior among discrete alternatives from different modes of transport (Train, 1980).

Discrete choice models based on the random utility maximization are intended to provide information about the effects of various policies in human behavior related to transportation decisions. This information can be used, along with other information and ideas, in deciding which policies should be implemented to tackle some of the transportation problem in urban areas.

To take these factors into consideration, specific in urban studies, the last decades witnessed the effort of researchers to develop discrete choice models to achieve the urban issues and several authors have reported the advances in developing theory and measure building and their implication for future studies on this topic (McFadden, 2000).

Following this perspective, this paper proposes the application of a random utility model to analyze human choice behavior among discrete alternatives from different modes of transport in a given set to a Brazilian medium sized city.

Mode choice data collected in a 2002 Origin-Destination survey in Uberlândia, Brazil, was used to obtain parameter estimates of utility functions, as well as to analyze trade-offs between observed attributes (including travel time and travel cost both based on geocoding origins and destinations).

The calibration of the discrete choice model (multinomial logit model) has been carried out with the software BIOGEME version 1.8. Model results outline the subjective value of time, fare policy (discount and free rides) and household characteristics (auto ownership and household income).

The aims of this paper are: (1) is to contribute to the understanding of human choice behavior in the transport field and (2) to serve as foundation for future studies of choices among different modes of transport in Brazil.

Following this line, the mode choice data is presented on the next section. The model specification is presented in Section 3, the model results in Section 4, the travelers subjective value of time in Section 5. The paper summarized by some implications and recommendations for further research.

# 2. Data

The analyses presented in this paper are based on data collected in 2002 from trips in Uberlândia. In the following subsections are presented more information about the study context, the revealed preferences survey and the descriptive data collected, the geocoding method of trip origins and destinations, a travel time estimation proceeding from perceived values reported by respondents, and the travel cost alternatives for motorized trips.

Two aspects were considered in the selection of the Uberlândia as case study: the characteristics of its mobility systems (presented in Table 2 and illustrated in Figure 1 and Figure 2) and the availability of data (specially because the trips within the study area could be geocoded). Unfortunately these elements are not common in all cities of the same size, or even in bigger sized city.

## 2.1 The study context: Uberlândia, Brazil

The application of a random utility model to analyze human choice behavior among discrete alternatives from different modes of transport was tested in Uberlândia, which is a medium sized city located in the state of Minas Gerais, Brazil.

According to Census Bureau, Uberlândia had 485,492 inhabitants in 2002. Its economy is based on commerce, agriculture, productive industry and services. More socioeconomic information about Uberlândia is provided in Table 1.

Uberlândia possesses a radial spatial growth (Figure 1) that associated with the real estate speculation, concentrates an expressive spatial activity density in the central area and a slight spatial activity density in the outlying areas.

The existence of arterial streets with radial characteristics strengthens the attraction for trips to the central area. Furthermore, the city core shows a highly density of activities like commerce, public administration and services.

The central area is surrounded by neighborhoods with relative density diversified activities and a spatial distribution of high income population. Moving away from the core to the peripheral area the activity density decreases (except by the industrial area, located in the periphery) and spatial distribution of low income population increases.

Table 1 Socioeconomic attributes of Uberlândia

| Attribute | Value |
|---|---|
| Area ($km^2$)[1] | 83.80 |
| Population[1] | 485,492 |
| Households[1] | 141,319 |
| Density: number of people per hectare[1] | 57.94 |
| Average household size[1] | 3.44 |
| Gross Domestic Product (USD)[2] | 2,479,763,829.70 |
| Agriculture sector (USD)[2] | 90,951,849.93 |
| Industry sector (USD)[2] | 909,553,546.54 |
| Commerce and service sector (USD)[2] | 1,234,965,550.20 |
| Others sectors (USD)[2] | 244,292,883.03 |
| Gross Domestic Product per Capita (USD)[2] | 2,479,763,829.70 |
| Number of cars[3] | 98,637 |
| Number of cars per household[3] | 0.70 |
| Number of cars per individual | 0.20 |

Source:

[1] National Census 2000.

[2] CESPE (2007). Rate applied: Secretariat of the Federal Revenue of Brazil, 1 USD = 2.5109 BRL / 1 BRL ≈ 0.40 USD at May 15th, 2002.

[3] National Department of Transity, 2002.

The transportation system facility of Uberlândia is made by streets (with hierarchical levels presented among them), well provided side walkers, bus stops and five bus stations (the available transit system). Fare integration in public transport system occurs in the bus stations, this mean that an individual can take several buses to complete the trip with one ticket, since the bus interchanges occurs in the bus stations. More details about the transport system are presented in Table 2 and Figure 1 and Figure 2.

Table 2 Facility description of Uberlândia transportation system

| Attribute | Value |
|---|---|
| Links | 25,101 |
| Nodes | 16,951 |
| Transit lines (one-way) | 160 |
| Physical stops (bus stations included) | 1,731 |
| Bus stations | 5 |

Figure 1 Network aspect of Uberlândia

Source: RP survey of Uberlândia (2002).

Figure 2 Transit System aspect of Uberlândia (bus station in green)

Source: RP survey of Uberlândia (2002).

## 2.2 Revealed preference survey

The mode choice data was collected in the spring of 2002 in Uberlândia by the Universidade Federal de Uberlândia through an application of a questionnaire of revealed preference survey. The aim of the sponsoring project was to understand the traveler behaviors in the city.

A total of 3,126 households interviews were completed yielding information about the household, the chosen mode and the main socio-economic characteristics of the individual. A final sample of 10,757 trip observations was left after removing errors in the data collection (Table 3).

Table 3 RP survey sample

| Sample | People | Trips |
|---|---|---|
| Original | 5,078 | 11,887 |
| Cleaned | 4,666 | 10,757 |

Source: RP survey of Uberlândia (2002).

## 2.3 Description of data

This section provides descriptive results in detail about the data collected in Uberlândia. Table 4 provides the frequency of mode and trip purpose shares; Table 5 presents the auto ownership in households and family role; Table 6 shows the frequencies of trip purpose per mode; Figure 4 shows the population age distribution by gender; Figure 5 presents the population income distribution by gender and Figure 5 provides the population household income distribution by gender.

The first observation in the analysis of mode choice shares given by Table 4 in approximately 70% of the trips the modal usage is not the auto. This can be related to the decision to use the car in the in household, since 53% of the households own a vehicle, 31% of the respondents are family head and 27% of the work trips the car (auto driver) is used.

Table 4 Mode share and trip purpose of the RP survey data

| Attribute | Value | |
|---|---|---|
| Mode | | |
| Walk | 3,179 | 29.55% |
| Cycle | 451 | 4.19% |
| Bus | 3,936 | 36.59% |
| Auto driver | 1,975 | 18.36% |
| Auto passenger | 1,216 | 11.30% |
| Trip Purpose | | |
| Home | 5,068 | 47.11% |
| School | 1,910 | 17.76% |
| Work | 2,732 | 25.40% |
| Other Activities | 1,047 | 9.73% |

Source: RP survey of Uberlândia (2002).

The high share of walk trips to school (47%, Table 6) can be in part explained by the local and state policy issue for education to force the individuals (from basic to high school level, individuals under the age of 18 years old, Figure 3) to study in public school next to household. This policy issue is not implemented for students from private schools and universities.

Table 5 Auto ownership in households and family role

| Attribute | Value | |
|---|---|---|
| Number of Cars in Household | | |
| 0 | 1,048 | 46.66% |
| 1 | 953 | 42.43% |
| 2 | 203 | 9.04% |
| ≥3 | 42 | 1.87% |
| Family Role | | |
| Head | 1,574 | 31.00% |
| Husband/Wife | 810 | 15.95% |
| Son/Daughter | 2,134 | 42.02% |
| Housemate | 419 | 8.25% |
| Non-relative | 36 | 0.71% |
| Maid | 8 | 0.16% |
| Other non-relative | 97 | 1.91% |

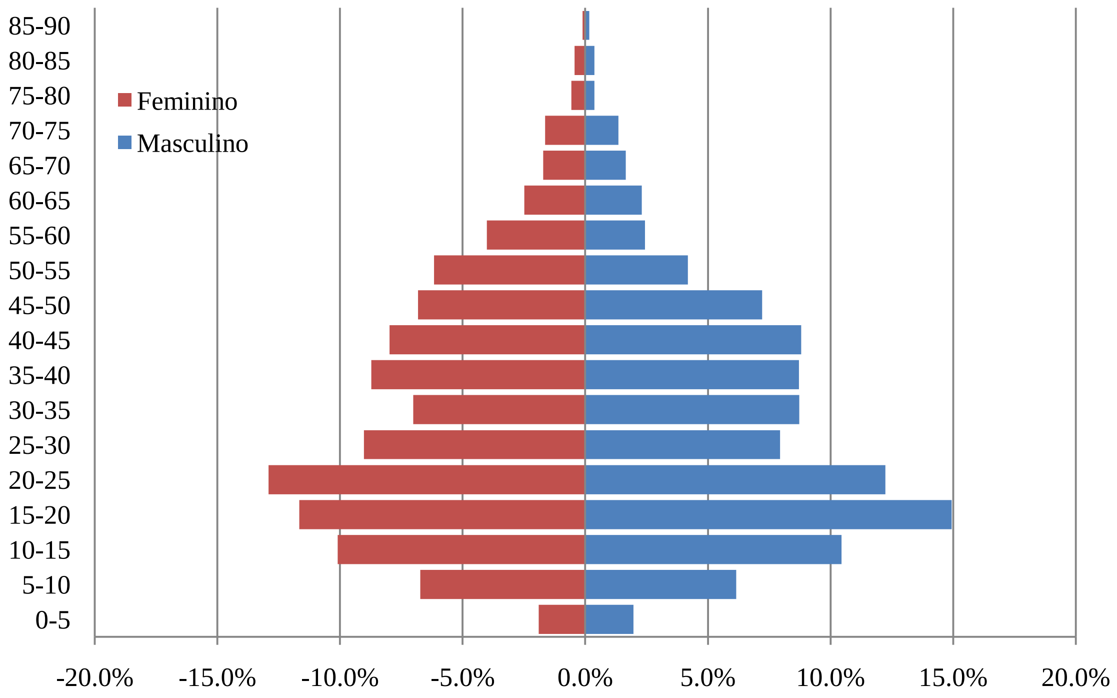
Source: RP survey of Uberlândia (2002).

Table 6 Trip purpose per mode

| Mode | Other Activities | | School | | Work | |
|---|---|---|---|---|---|---|
| | Frequency | | Frequency | | Frequency | |
| Walk | 228 | 21.78% | 909 | 47.59% | 497 | 18.19% |
| Cycle | 18 | 1.72% | 47 | 2.46% | 175 | 6.41% |
| Bus | 453 | 43.27% | 586 | 30.68% | 1,040 | 38.07% |
| Auto driver | 205 | 19.58% | 120 | 6.28% | 755 | 27.64% |
| Auto passenger | 143 | 13.66% | 248 | 12.98% | 265 | 9.70% |

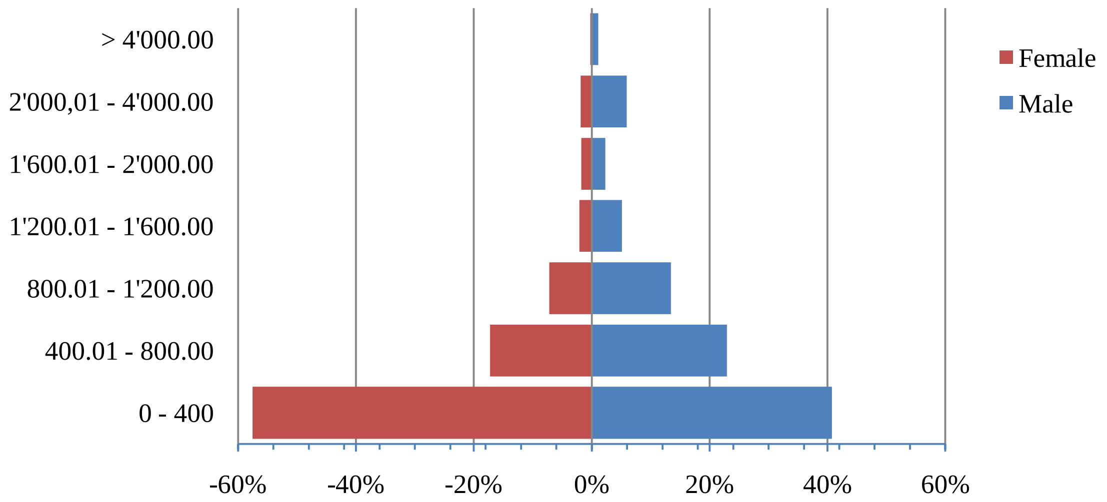
Concerning about the information presented Figures 4 to 5, one can be observed is the differences among the income and household income distributions, possibly because more than one individual work in the household.

Figure 3 Population age distribution (5 year age bands) by gender



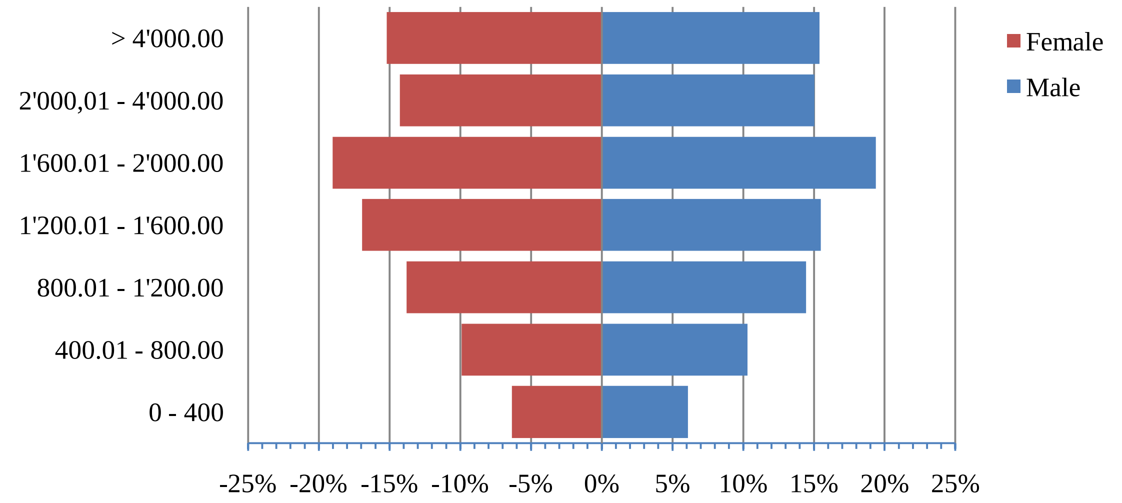

Source: Origin-Destination Survey of Uberlândia (2002).

Figure 4 Population income (in BRL) distribution by gender




Source: Origin-Destination Survey of Uberlândia (2002).

Figure 5 Population income (in BRL) distribution by gender




Source: Origin-Destination Survey of Uberlândia (2002).

## 2.4 Geocoded trips

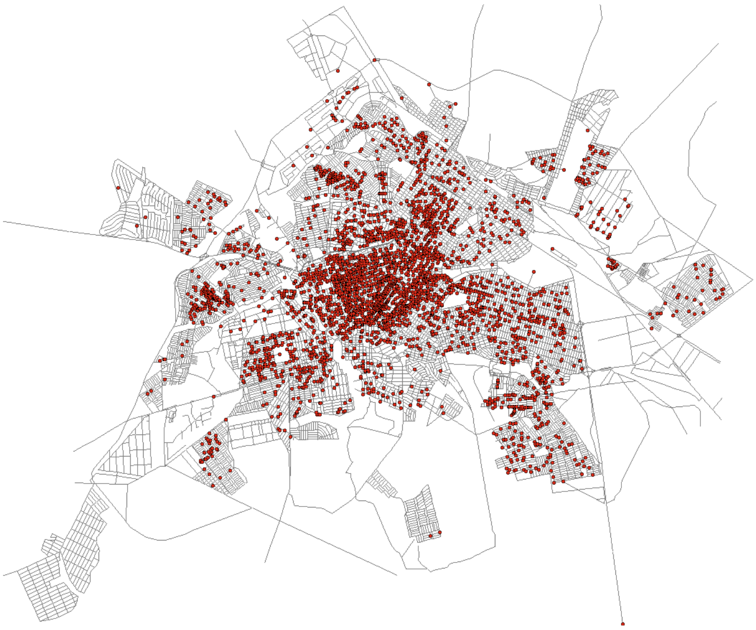
The 93.7% reported trips within the study area could be geocoded, as presented in Table 7 and Figure 6. The analysis of geocoded locations estimates revealed accuracy with absolute percent error of 19.13% over all geo-code addresses at correct traffic zone (65 traffic zones were considered on the validation test).

Based on the geocodes network-based information about the trips were added for both individual and public modes: car, road distances; public transport, public transport network distances and road distances; walking, Euclidean distances and road distances; and cycling road distances.

Table 7 – Sample aspects of geo-code addresses of trip origins and destinations

| Sample | Number of street addresses |
|---|---|
| Original data | 23,774 |
| Exclusive cases (without repetition on database) | 5,123 |
| Non-identified cases | 145 |
| Geo-code street addresses | 4,798 |

Figure 6 Geocode results of origins and destination trips



Source: Origin-Destination Survey of Uberlândia (2002).

Based on the geocode results average distances were estimated and Table 8 presents the values per socioeconomic attribute. In Table 9 is presented the average distances per mode and trip purpose.

The analysis of Table 8 indicates that trip lengths average are slightly sensitive to socioeconomic attribute: men are higher than woman and head are higher than son/daughter (one that can explain is the different among purpose, work and schools, and the fact that the school is closer to household, as discussed above).

In Table 9, as expected, walk mode has the lower average trip length and work purpose have the higher trip length average.

Table 8 Average trip length (network distance) per socioeconomic attribute

| Attribute | Trips | | Average distance (km) |
|---|---|---|---|
| Gender | | | |
| Male | 5,429 | 50.47% | 3.84 |
| Female | 5,328 | 49.53% | 3.47 |
| Family Role | | | |
| Head | 3,498 | 32.52% | 4.18 |
| Husband/Wife | 1,707 | 15.87% | 3.91 |
| Son/Daughter | 4,467 | 41.53% | 3.29 |
| Housemate | 822 | 7.64% | 3.18 |
| Non-relative | 72 | 0.67% | 3.50 |
| Maid | 13 | 0.12% | 1.77 |
| Other non-relative | 163 | 1.52% | 2.49 |
| Income (BRL)[1] | | | |
| 0 to 400 | 5,279 | 49.08% | 3.30 |
| 400.01 to 800 | 2,163 | 20.11% | 4.34 |
| 800.01 to 1.200 | 1,111 | 10.33% | 4.02 |
| 1,200.01 to 1,600 | 392 | 3.64% | 3.99 |
| 1,600.01 to 2,000 | 220 | 2.05% | 3.84 |
| 2,000.01 to 4,000 | 422 | 3.92% | 3.93 |
| > 4.000 | 75 | 0.70% | 4.03 |
| Household Income (BRL)[1] | | | |
| 0 to 400 | 669 | 6.22% | 3.73 |
| 400.01 to 800 | 1,088 | 10.11% | 3.83 |
| 800.01 to 1.200 | 1,519 | 14.12% | 3.90 |
| 1,200.01 to 1,600 | 1,743 | 16.20% | 3.58 |
| 1,600.01 to 2,000 | 2,065 | 19.20% | 3.60 |
| 2,000.01 to 4,000 | 1,576 | 14.65% | 3.46 |
| > 4.000 | 1,646 | 15.30% | 3.80 |

Source: RP survey of Uberlândia (2002), [1] Secretariat of the Federal Revenue of Brazil, 1 USD = 2.5109 BRL / 1 BRL ≈ 0.40 USD at May 15th, 2002.

Table 9 Average trip length (network distance) per mode and purpose

| | Trips | | Average distance (km) |
|---|---|---|---|
| Mode | | | |
| Walking | 3,179 | 29.55% | 1.64 |
| Cycling | 451 | 4.19% | 3.38 |
| Transit - Bus | 3,936 | 36.59% | 5.03 |
| Auto Driver | 1,975 | 18.36% | 4.21 |
| Auto Passenger | 1,216 | 11.30% | 3.66 |
| Purpose | | | |
| Household | 1,047 | 9.73% | 3.63 |
| School | 5,068 | 47.11% | 3.62 |
| Other activities | 1,910 | 17.76% | 2.75 |
| Work | 2,732 | 25.40% | 4.37 |

Source: RP survey of Uberlândia (2002).

## 2.5 Real-travel times estimation

Research on travel behavior is often based on travel times reported by travelers. It is expected an accurate measurement since most travelers have their own watch; in addition, travelers have to keep a close look at their travel times when they want to arrive in time for scheduled activities. Nevertheless, it is clear that also here inaccuracies occur (Rietveld, 2002).

To analyze human choice behavior among discrete alternatives from different modes of transport there was a need to estimate travel times for the mode choices alternatives: walk, cycle, bus, auto driver and auto passenger (a discussion about the design of the alternative choice set is presented in section 3.1). For simplification auto driver and auto passenger were considered as auto for travel time estimation.

In the RP survey the respondent were asked to report their actual travel time for the mode used on the trip. The travel time of alternative modes were not reported. To overcome this problem approaches were used to estimate the entire alternative set of travel times.

The perceived values were grouped in five databases, namely walk, cycle, bus, auto and prediction. For each individual in the data base it was associated the reported travel time and based on geocodes the trip length were also added: walk (Euclidean distance), cycle (road distance), bus (transit network distance), auto (road distance) and prediction (euclidean, road and transit distances).

For modeling purposes, speeds were estimated for in each mode observation. Table 10 provides the averages values extract in the sample. As expected, there were speeds values out of the range (upper and lower bound were define) and these observations were excluded for model building estimation.

Table 10 Mode speed statistics

| Mode | Trips | Average | Std deviation |
|---|---|---|---|
| Walk (range up to 7.93km/h) | 3,216 | 3.21 | 1.71 |
| Cycle (range up to 30.29 km/h) | 462 | 9.13 | 5.09 |
| Bus (entire sample)[1] | 3,842 | 19.57 | 14.90 |
| Auto driver (up to 87.63 km/h) | 2,042 | 16.53 | 12.00 |

[1]Bus speeds are base on bus trip lengths (transit network)

Ideally, individual speeds should vary in the sample; but after a sensitive analysis about the influence of socio (age and gender), locations and timing attributes on individual speeds the value walk and cycle mode speeds were considered the same value for all individuals.

So the average speeds of **3.21 km/h** and of **9.13** km/h were used to predict the travel time of walk and cycle mode, respectively, in the sample (including those who reported walk and cycle mode on their trips).

In the case of bus and auto the sensitive analysis about speed indicates the influence peak hour (if the bus/car trip occurs during peak hours, or not), number of transfers (bus), and location displacement (car). The travel time by bus is given to the linear regression model described in Table 11 and the prediction of the auto speeds is given by cluster model, described in Table 12.

Table 11 Model estimation results for bus travel time estimation

| Variable | Estimates | SE Estimates | t-ratio |
|---|---|---|---|
| Interceptor | - | - | - |
| Bus Length (km) | 0.0457 | 0.0015 | 30.38 |
| Peak hour (1 – Yes; 0 – No) | 0.0816 | 0.0110 | 14.90 |
| Transfers (number of transfers) | 0.0780 | 0.0126 | 6.18 |

$R^2$ = 0.3626

Table 12 Cluster estimation results for auto driver speed estimation

| Cluster | Obs. | Origin | Destination | Peak | Average | Std deviation |
|---|---|---|---|---|---|---|
| 1 | 128 | Core | Core | Yes | 13.05 | 9.11 |
| 2 | 174 | | | No | 13.52 | 9.24 |
| 3 | 206 | | NW, SW and SE | Yes | 15.09 | 8.19 |
| 4 | 221 | | | No | 15.94 | 8.83 |
| 5 | 37 | | NE | - | 16.69 | 8.75 |
| 6 | 186 | SE | SE | - | 12.92 | 10.18 |
| 7 | 174 | | Core | - | 15.94 | 7.71 |
| 8 | 177 | NE and NW | Core | - | 15.77 | 9.02 |
| 9 | 118 | SW | | - | 17.11 | 10.82 |
| 10 | 138 | NE, NW, SW | SE | - | 21.11 | 11.36 |
| 11 | 156 | NE, NW and SW | SW | - | 16.85 | 10.45 |
| 12 | 146 | NE and NW | NE and NW | - | 15.92 | 10.41 |
| 13 | 41 | SW | | - | 26.23 | 17.46 |
| 14 | 140 | SE | NE, NW and SW | - | 26.23 | 11.42 |

## 2.6 Travel cost alternatives

Table 13 shown the travel cost alternatives were considered for test for the choice of public transport and Table 14 for the choice of auto driver:

Table 13 Travel cost alternatives for the bus mode

| Respondent Characteristic | Value (BRL) | |
|---|---|---|
| | Marginal Cost (discount ticket) | Fixed Cost (full price ticket) |
| Individual age ≥ 65 years | 0.00 | 1.25 |
| Child age ≤ 6 years | | |
| Student | 0.75 | |
| Worker | 0.00 6% Wg | |

- Individual age older than 65 years: there is a transit national regulation which states that individuals older than 65 years are free of cost to use the public transport systems;
- Child age younger than 6 years: there is no transit regulation on this theme; however, in practice there is an "informal" agreement among transport public drivers that children younger than 6 years are free of cost to use the public transport systems;

- Student discount: students have 40% of discount in the ticket. This is a local transit regulation;

- Worker: according to the Brazilian labor regulation, the employers are responsible to cover the monthly expenses of the employee trips from employee’ home to work, and from work to employee’ home and charge from employee 6% of the employee wage, or the trip budget, what does come first;

Table 14 Travel cost alternatives for the auto driver mode

| Cost Alternative | Value (BRL/Km) |
|---|---|
| (1) Fuel burned cost | 0.24 |
| (2) Marginal cost: including taxes, maintenance and fuel burned costs | 0.44 |
| (3) Fixed cost: price of the good (car) and capital costs | 0.24 |
| (4) Total cost: (2) + (3) | 0.68 |

- Fuel burned cost: is the price of the fuel burned to complete the trip;

- Marginal cost: are the costs to use the car;

- Fixed cost: are the costs to own the car;

- Total cost: is the global cost, to own and use the car.

# 3. Modelling approach

In general terms, mode choice with disaggregate data can be explored on temporal-dependent, trip-based and activity-based decisions. Temporal-dependent decisions means that mode choice can be undertaken at different temporal context: in the long-run there is question to which mode the travellers pre-commitment themselves through the purchase of car, motorcycles or public transport season tickets (e.g. mode-choice and auto-ownership jointly), in the short run, one can look at the price policy for public transport and address the analysis to the effect of a new price policy on modal shares.

In trip-based context, each trip is seen as independent event, for which the traveller made new and separate decision regards to mode, in contrast to activity-based decisions that seen the trips (tours) as an outcome of a daily activity scheduling of an individual.

Traditionally the analysis has been explored in trip-based context with different temporal-decisions (long- and short-run). However, the last decades witnessed the effort of researchers to develop behavioral travel demand models (trip generation, timing, destination and mode choice) based on the activity-based theory to achieve the complexities of human choice behavior in the urban context.

These distinguishing approaches can be viewed by considering the concept of problem-solving in the context of systems. Understanding human choice behavior is complex and require a number of simplifying assumptions, otherwise the level of problem-solving shift from simple problems, described by a limited number of variables, to problems of organized complexity, which involve dealing simultaneously with sizeable number of factors which are interrelated into an organic whole, or problems of disorganized complexity, which the number of variables is very large, and one in which of the many variables has an individually erratic behavior and may be totally unknown. Further details about simple, organized and disorganized problems in the context of system are presented by Weaver (1961).

Trip-based models keep the analysis of human choice behavior less inclusive and place some limitations for planning strategies. However, activity-based models are relatively demanding on data and embedded by a more complex model building structure.

In this paper, the line of problem-solving followed to analyze human choice behavior among discrete alternatives from different modes of transport in Uberlândia is the trip-based and the model specification is designed to explore short- and long-run decisions.

This approach addresses the relevant issues applied in the literature of mode choice studies and highlights the effects of the measured attributes (variables in the modeling process) on the context of individuals and their mode decisions for traveling to a particular destination. On the other hand, dealing with simplifying assumptions, the case of trip-based approach, place some limitations on the interpretation results.

To progress, a general model of individual behavior requires that three key factors be taken into account (McFadden, 1973): (1) the objects of choice and sets of alternatives available to decision-makers, known as choice set generation; (2) the observed attributes of decision-makers, and; (3) the model of individual choice and behavior and distribution of behavior patterns in the population.

## 3.1 The objects of choice and sets of alternatives available to decision-makers

In the RP survey of Uberlândia, respondents were asked to inform the main mode or combination of modes (e.g. cycle mode until bus station, mode shift to bus until the final destination, so the combination is cycle and bus) used in each trip.

This information was readily available, because a set of ten modes alternatives (1- bus, 2- scholar bus, 3 – auto driver, 4 – auto passenger, 5 – shuttle, 6 – motorcycle, 7 – cycle, 8 – walk, 9 – truck, 10 – others) was offered in the RP survey form and the respondents were requested to choose one (main mode) or more than one (combination of modes) option on the list. Table 15 presents the values of the modes shares.

Table 15 Travel mode alternatives and relative frequency

| Mode alternative (number) | Frequency (%)[1] |
|---|---|
| Bus (1) | 32.5 |
| Scholar bus (2) | 0.5 |
| Auto driver (3) | 14.9 |
| Auto passenger (4) | 12.9 |
| Shuttle (5) | 2.3 |
| Motorcycle (6) | 4.6 |
| Cycle (7) | 4.1 |
| Walk (8) | 27.1 |
| Truck (9) | 0.4 |
| Others (10) | 0.7 |

[1]Frequencies based on original RP sample.

Based on some similarities among travel modes the individual were considered to have a choice among five alternatives: walk, cycle, bus[1] (bus, scholar bus, shuttle and others, because all of them are not private transport modes), auto driver[2] (auto driver and motorcycle, because both are private motorized modes) and auto passenger. The resized modal shares are presented on Table 16.

Table 16 Resized travel mode alternatives and relative frequency

| Mode alternative (number) | Frequency (%) |
|---|---|
| Walk | 27.1 |
| Cycle | 4.1 |
| Bus | 36.0 |
| Auto driver | 19.9 |
| Auto passenger | 12.9 |

## 3.2 The observed attributes of decision-makers

Following the literature on this subject and especially the literature about empirical applications of random utility models to analyze human choice behavior among different modes of transport in Brazilian cities, e.g. Swait and Ben-Akiva (1987), an extensive set of variables were considered for test:

- Social attributes: age (dummy codes for age bands), education level (dummy codes for non-undergraduate, undergraduate), family role (dummy codes for head, wife/husband, son, etc.) and gender (dummy codes for male/female);
- Economic attributes: household income (BRL), income (BRL), transport budget (wage %);
- Mobility attributes: auto ownership (number of autos);
- Transportation attributes: travel time (hours), travel cost alternatives (BRL) described in section 2.6., transit supply measure, and;
- Policy issues: free rides for older than 65 years (dummy code) and discount tickets for workers (dummy code).

[1] For convenience the term was kept.

[2] For convenience the term was kept.

## 3.3 The model of individual choice and behavior

The basic transportation choice behaviour of urban individuals is considered here to be a result of decision-makers' trade-offs between costs and benefits perceived to be associated with decision alternatives.

These trade-offs are modelled within the analytical structure of utility functions and individuals are considered to have a rational economic consumer behavior. Let $U_{ni}$ be the utility of the $i$th alternative for an individual $n$. Further assume each utility value can be portioned into two components: a systematic component or "representative utility", $V_{ni}$ and a random component, $\varepsilon_{ni}$, the latter reflecting unobserved individual idiosyncrasies of tastes. Then, the utility function can be expressed as:

$$U_{ni} = V_{ni} + \varepsilon_{ni} \tag{1}$$

The probability that a person with certain observed socioeconomics characteristics and facing a choice among several alternatives each of which exhibiting certain measure of attributes will chooses alternative $i$ is

$$P_{ni} = Prob(V_{ni} + \varepsilon_{ni} > V_{nj} + \varepsilon_{nj}, \forall\, j \neq i)$$

$$P_{ni} = Prob(\varepsilon_{nj} < \varepsilon_{ni} + V_{ni} - V_{nj}\,, \forall\, j \neq i) \tag{2}$$

Some algebraic manipulation of this general choice probability and assuming that each $\varepsilon_{ni}$ is independently, identically distributed extreme value, the closed expression of the probability is given by Eq. (3):

$$P_{ni} = \frac{e^{V_{ni}}}{\sum_j e^{V_{nj}}} \tag{3}$$

which is the multinomial logit choice probability. The representative utility is usually specified to be linear in parameters:

$$V_{nj} = \beta' \cdot x_{nj} \tag{4}$$

where $x_{nj}$ is a vector of observed variables relating to alternative $j$. With this specification, the logit probabilities become:

$$P_{ni} = \frac{e^{\beta' \cdot x_{ni}}}{\sum_j e^{\beta' \cdot x_{nj}}} \quad (5)$$

Eq. (5) express the probability that a person with certain observed socioeconomic characteristics and facing a choice among several alternatives will choose a particular alternative.

# 4. Model specification and results

This section describes the estimated model of the conditional probabilities of the trip mode choice for the data collected in Uberlândia. Individuals were considered to have a choice among the five alternative modes of travel: walk, cycle, bus, auto driver and auto passenger

The model estimates the probability that a particular individual will choose each of the five modes for his trip, given the observed attributes. The utility functions take the forms of Eq. (6)-(10) and is specified to be linear- and not-linear-in parameters (representative utility).

$$U_{walk} = \beta_{time\ walk} \cdot \delta_{time\ walk} \tag{6}$$

$$U_{cicle} = \beta_{time\ cicle} \cdot \delta_{time\ cicle} + \varepsilon_{cicle} \tag{7}$$

$$U_{bus} = \beta_{time\ bus} \cdot \delta_{time\ bus} + \beta_{bus\ discount} \cdot \delta_{bus\ discount} + \beta_{cost} \cdot \delta_{bus\ free\ ride} + +\beta_{bus\ distance} \cdot \delta_{bus\ distance} + \beta_{household\ Income} \cdot log(\delta_{household\ income} + 1) + \beta_{transit\ supply} \cdot log(\delta_{transit\ supply} + 1) + e^{\beta_{bus\ fare} \cdot \delta_{bus\ fare}} + \varepsilon_{bus} \tag{8}$$

$$U_{auto\ driver} = \beta_{time\ auto\ driver} \cdot \delta_{time\ auto} + \beta_{auto\ driver} \cdot \delta_{auto} + \beta_{head} \cdot \delta_{head} + \beta household\ Income \cdot log(\delta household\ income+1)+e\beta cost \cdot \delta\ auto\ driver\ cost+\varepsilon auto\ driver \tag{9}$$

$$U_{auto\ passeger} = \beta_{time\ auto\ passenger} \cdot \delta_{time\ auto} + \beta_{auto\ passenger} \cdot \delta_{auto} + \varepsilon_{auto\ passenger} \tag{10}$$

which

- $\delta_{time\ walk}$ is the travel time attribute (door-to-door in hours) used for the walk mode alternative.
- $\delta_{time\ cicle}$ is the travel time attribute (door-to-door in hours) used for the cycle mode alternative.
- $\delta_{time\ bus}$ is the travel time (attribute door-to-door in hours, including waiting, access and egress times) used for the bus mode alternative.
- $\delta_{time\ auto}$ is the travel time attribute (door-to-door in hours) used for auto alternatives (auto driver and auto passenger).
- $\delta_{bus\ discount}$ is set to 1 if the respondent is worker and has a discount ticket ("work pass"). This term is included only for the bus alternative.
- $\delta_{bus\ free\ ride}$ is set to 1 if the respondent is 65 or more years old and does not have to pay for the ride. This term is included only for the bus alternative.

- $\delta_{bus\ distance}$ is the total trip distance length attribute (in km) in the transit system. This term is included only for the bus alternative.

- $\delta_{household\ income}$ is the monthly household income attribute (in BRL). This term is included for the bus and auto driver alternatives.

- $\delta_{transit\ supply}$ is a measure of the transit supply system. This term is included only for the bus alternative.

- $\delta_{bus\ fare}$ is the price attribute (in BRL) for a single trip ticket. This term is included only for the bus alternative.

- $\delta_{auto}$ is the number of autos attribute in the respondent household. This term is included only for the auto alternatives (auto driver and auto passenger).

- $\delta_{head}$ is set to 1 if the respondent is the head member of the family. This term is included only for the auto alternatives (auto driver and auto passenger).

- $\delta_{auto\ driver\ cost}$ is the auto cost (in BRL) attribute associated with auto driver mode. The cost is based on the fuel burned cost to complete the trip.

- $\varepsilon_{cicle}$, $\varepsilon_{bus}$, $\varepsilon_{auto\ driver}$, $\varepsilon_{auto\ passenger}$ are the unobserved attributes associated with the mode alternatives.

Table 17 presents the model estimated. Estimation of coefficients (β) was performed by the maximum likelihood containing in BIOGEME (Bierlaire, 2003), that provides the estimated parameter along (column 2) with its statistics (column 3).

For each person in the sample, the available choice set was considered to be the five alternatives listed above with the following expectation. The walk, cycle and auto passenger modes were considered available to all individuals. The bus mode was considered unavailable to individual if the trip distance is not so far that the individual can walk (less than 500m), and/or if the bus mode is not available to the individual (distance to the access or egress stops are more than 3km from the origin or to the destination). The auto driver was considered unavailable to an individual if the individual's household had chosen to own no autos and/or if the individual is under 18 years old.

Mode travel times are based on the estimation presented in section 2.5 and all travel cost variable alternatives were considered in the modeling trial, and the one that provides the best fitting was choose (correct in signal and statistic significance). The analysis of the results and conclusions are presented in the Discussion section.

Table 17 Parameter estimates and model fit for the mode choice model

| Utility function parameters | Estimates | (t-ratio) |
|---|---|---|
| 1. *Constants* | | |
| Walk | - | - |
| Cycle | -3.300 | (-35.71) |
| Bus | -3.060 | (-15.77) |
| Auto Driver | -3.920 | (-25.10) |
| Auto Passenger | -3.450 | (-43.38) |
| 2. *Travel Time, door-to-door, one-way, in hours.* | | |
| Walk | -3.270 | (-35.22) |
| Cycle | -1.940 | (-8.66) |
| Bus | -0.660 | (-2.09) |
| Auto Driver | -0.883 | (-2.74) |
| Auto Passenger | -2.740 | (-8.87) |
| 3. *Distance, door-to-dor, transit network, in km.* | | |
| Bus | 0.044 | (2.43) |
| 4. *One-way travel-cost, marginal cost fuel burned (BRL) for auto driver mode and ticket price (BRL) for the bus mode.* | | |
| Bus<br>Auto Driver | -0.540 | (-2.44) |
| 5. *Household monthly income (BRL.)* | | |
| Bus<br>Auto Driver | 0.039 | (2.89) |
| 6. *Auto Ownership, number of autos in household* | | |
| Car Driver | 0.936 | (17.33) |
| Auto Passenger | 1.080 | (25.71) |
| 7. *Transit Supply Measure:* | | |
| Bus | 0.267 | (6.47) |
| 8. *Bus free ride for 65 or more years old passenger = 1 , yes / = 0, no* | | |
| Bus | 0.648 | (5.75) |
| 9. *Bus discount ticket for workers = 1 , yes; = 0, no* | | |
| Bus | 0.212 | (4.16) |
| 10. *Household head member = 1 , yes; = 0, no* | | |
| Auto Driver | 1.060 | (17.26) |

Observations 10,594; Null log-likelihood -15,553.03; Final log-likelihood 10,645.87; Adjusted $\rho^2$ 0.314.

# 5. Subjective value of travel time savings

The subjective value of travel time savings (SVTTS) is an important willingness-to-pay indicator, used for example cost-benefit analysis in the context of planning new transport systems, or for pricing (Hess et al., 2005). In discrete choice measure is given by Eq. (11), Jara-Díaz (2007):

$$SVTTS = \frac{\partial V / \partial TT}{\partial V / \partial TC} \quad (11)$$

which, $V$ is the deterministic part of the utilities in the model that contains a travel time attribute TT and travel cost attribute TC.

For the commonly used linear in-variables utility functions, this formula reduces to $\beta_{TT}/\beta_{TC}$ , where $\beta_{TT}$ and $\beta_{TC}$ are the time and cost coefficients (produced by calibrating the model on the choice data used in the estimation). For the non-linear formulation case of the model specification described in section 4, the VTTS computation is not relatively straightforward and the application of Eq. (11) yields:

$$\frac{\beta_{time\ i}}{\beta_{cost} \cdot e^{\beta_{cost} \cdot \delta_{cost\ i}}} \quad (12)$$

Eq. (12) implies that the values of travel time savings measure are mode dependent. Therefore the SVTTS are based on arithmetic averages of modal cost values. The results are shown in Table 18 and the analysis is presented in the Discussion section.

Table 18 Subjective value of travel time savings per mode

| Mode | Time parameter | Cost paramenter | Average Cost (BRL) | SVTTS (BRL) | SVTTS (USD)[1] |
|---|---|---|---|---|---|
| Walk | -3.270 | | | 10.75 | 4.28 |
| Cycle | -1.940 | | | 6.38 | 2.54 |
| Bus | -0.660 | -0.540 | 1.06 | 2.17 | 0.86 |
| Auto Driver | -0.883 | | | 2.90 | 1.16 |
| Auto Passenger | -2.740 | | | 9.00 | 3.59 |

[1] Secretariat of the Federal Revenue of Brazil, 1 USD = 2.5109 BRL / 1 BRL ≈ 0.40 USD at May 15th, 2002.

# 6. Discussion

This paper has presented the findings of a study looking at the analysis of human choice behaviour among discrete alternatives from different modes of transport in a given set to Brazilian medium sized city.

A method of finding associated geographic coordinates to street addresses (latitude and longitude) have been applied for the mode data collected and it able to estimate the impedance measure (trip length based) at street level.

In fact the availability of street based origins and destinations adds significantly to the applicability to the choice behaviour analysis, since the unavailability of street level detail level is one of the deficiencies (usually traffic zones aggregation are used) noted in the data usually collected in RP surveys in Brazil.

Drawback of the geocoding, lack of empirical evidence of goodness-of-fit, can be overcome by spatial validation tests (e.g. actual against mapped geocoded locations). The analysis of geocoded locations estimates revealed accuracy with absolute percent error of 19.13% over all geo-code addresses at correct traffic zone. However, this validation test is a weak approximation of the true differences (it can exhibits error less than 19.13% or more than it). An in-depth study of the true errors in geocoding locations is indicated for future researches.

Although real-travel time estimations are based on reported values in the sample, it can be argued that they cannot be truly representative of the population future's perceptions, so, models based on this information may not be useful for forecasting. However, it is also true that, since real-travel time estimations are closer to individual cognitive decisions it could be expected better models fits using this type of data.

Furthermore, what matter in utility is the difference among the travel time of alternative modes. If the estimations are not fitted with actual travel time, but it keeps the differences among travel time of modes, therefore these values are right in sense for the utility functions.

In the logit model we estimate 14 parameters and 4 alternative specific constants (ASC), all coefficients in the model are of the expected right sign. All the ASC are negative, and the effect on null parameter on the representative utility is a probability of 0.88 to choose the walk mode. It can be assumed that the representative utility captured the advantages attributes of other alternatives modes in comparison with the walk mode.

It is expected that faster is the mode, lesser is the impact of travel time disutility. However, the time parameter of auto driver mode is 33.8% higher than bus mode. This results can be related to low travel speeds of auto mode in the core of Uberlândia, as reported in Table 10.

The trip length is positive for bus alternative, indicating that for long trips the traveler's are more likely to use the bus. This can be in part explained by the average trip length of bus travelers in Table 6. It is the highest average among modes.

Furthermore, the sensitivity of the public transport measure indicates that travelers in well provided public transport locations have the probability of using the bus increased, as expected, and the policy issues of free rides for individuals older than 65 years and discount tickets for workers have a positive impact on bus utility.

The public transport measure and the policy issues attributes have special meanings for researchers and policy makers to tackle some of the problems of urban transportation planning. It indicates that providing access to the public transport service (high frequency and travel time minimization among locations) associated with fare policies, like discount ticket, or subsidies for low income people, will have a positive impact on individual probability to use the bus.

It has long been understood that the interaction between two locations declines with the increasing disutility between (e.g. cost) them. Several forms of decay function have been established in the literature: negative power, negative exponential function, Gaussian function and modified (log)logistic function. In the model, the best fitted decay cost function was the negative exponential function and it suggested a strong sensitivity of the travelers to short distances (as the cost is a function of distance).

The long-run decision of more than one auto ownership in household is positive and significant for the utilities of auto drivers and auto passengers. Thus, as auto availability in household increases (i.e. more cars per individual) the probabilities decrease for all other modes except auto driver and passenger.

Approximately 53% of Uberlândia households owned one or more automobiles in 2002 (Table 4), so the effect discussed above is likely to be applicable to a great portion of the population, which does own a vehicle. Furthermore, being the family head and have auto in household increases the probability to use auto.

In terms of cost-benefit practice, the subjective value of travel time was found to be higher for walk (BRL 10.75), followed by auto passenger (BRL 9.00), cycle (BRL 6.38), auto driver (BRL 2.90) and auto passenger (BRL 2.17).

The results suggest that improvements in transport systems appear to be more valued for walk travelers than other mode travelers and may it turns the question to what improvements in transport system can be implemented to travel faster.

However, it should be noted that there the desire to speed trip up does seem to be an inconsistency in the travel time savings arguments within urban context, where much effort is now going into slowing traffic down for environmental and safety reasons. The key policy objective now becomes that of reasonable travel time, rather than travel time minimization.

It has important implications for transport planning, if it is to embrace the concepts of sustainable mobility (to reduce the need to travel, to levels of car use trough the promotion of walk and cycle, to reduce trip lengths and to encourage greater efficiency in the transport system).

In closing, it should be said that the application presented in this papers made use of one specific functional relationship between taste coefficients socio-demographic, transport system and policy issue indicators. There are many open research problems associated with the choice in transport system in the context of random utility maximization models and it is expected to see more research on this topic in the future.

# 7. Acknowledgements

This research was given financial support by the National Council for Scientific and Technological Development (Brazil), through grants 141058/2006-4, the São Paulo Research Foundation (Brazil), through grants 06/57487-5 and the CAPES Foundation (Brazil), through grants BEX2188/07-4. The first author is greatly indebted to Prof. Carlos Faria from Universidade Federal de Uberlândia on his kind help for proving the data used in this research and to Prof. Eiji Kawamoto from Universidade de São Paulo on his support along with the study on travel demand models.

# 8. References

Bierlaire, M., 2003 BIOGEME: a free package for the estimation of discrete choice models. In: STRC (Swiss Transport Research Conference), *3rd Swiss Transport Research Conference.* Monte Verita, Ascona, Switzerland. STRC: Monte Verita.

Hess, S., Bierlaire, M., Polak, J., 2005. Estimation of value of travel-time savings using mixed logit models. *Transportation Research A*, 39, pp. 57-66.

Jara-Díaz, S., 2007. *Transport economic theory*. 1st ed. Amsterdam: Elsevier.

McFadden, D., 2000. Disaggregate behavioral travel demand's RUM side: a 30-year retrospective. In: IATB (International Association for Travel Beavior), *6th International Conference of the Association for Travel Behaviour*. Gold Coast, Queensland, Australia 2-7 July 2000. IATB: Gold Coast.

McFadden, D., 1973. Conditional logit analysis of qualitative choice behavior. In: P. Zarembka, ed. 1973. *Frontiers in econometrics*. New York: Academic Press, pp. 105-142.

Rieveld, P., 2002. Rounding of arrival and departure times in travel surveys. *Journal of Transportation and Statistics*, 5, pp. 71-82.

Swait, O., Ben-Akiva, M., 1987. Empirical test of a constrained choice discrete model: mode choice in São Paulo, Brazil. *Transportation Research B*, 21 (2), pp. 103-115.

Weaver, W., 1961. A quarter century in the natural sciences. *Public Health Report*, 76 (1), pp. 57-66.