Everyday travel mode choice and its determinants: trip attributes versus lifestyle

Braun Kohlová, Markéta

Faculty of Social Sciences and Charles University Environment Center, Charles University in Prague, Czech Republic

Abstract

Lifestyle, indicating preferences toward a particular type of neighborhood and its transport accessibility is an important determinant of travel mode choice. In the paper a discrete choice model estimates are presented which include beyond the attributes of trip and socio-economic determinants also latent variables capturing the differences in individual preferences of neighborhood type and access to downtown and amenities (traditional city-like versus suburban). Empirical results confirm that the inclusion of latent variables capturing access and city lifestyle preferences into the travel mode choice model increase its explanatory power. The demonstrated connection between city lifestyle and transit and walk choice and connection between suburban lifestyle and car choice imply that policies aimed at behavior modification may be limited by the large proportion of households who have strong preferences towards suburban and car oriented lifestyle.

Introduction

Transport is one of the most important, but also hardest-to-solve, problems of modern-day cities. In developed countries, the transport sector contributes to the GDP by about 10%, and its services are a precondition for economic activities as well as leisure activities. At the same time, transport, especially in cities, causes a number of environmental problems and negative impacts on human health due to emissions, noise or vibrations. Accidents and congestion are other visible negative manifestations of transport. Both represent a social as well as economic problem.

All the mentioned types of damage grow as the mobility of the population increases – the total number of trips, as well as the distances travelled rise. The Czech Republic is no exception – the road passenger transport grows faster than 5% a year. The passenger-kilometers by cars were increasing constantly by more than 70% between 1990 and 2003. The growth rate has been slowing down in the last few years but car use is expected to continue to grow with the citizens' economic situation improving. We experience the changes in the modal split of passenger transport. The share of so-called "sustainable modes of transport", which include public transport, cycling and walking, is decreasing, while the share of cars is growing.

In the paper I present the estimates of a discrete choice model including beyond the attributes of trip and socio-economic predictors also latent variables capturing the differences in individual preferences of residence and access to downtown and amenities (traditional city-like versus suburban) which are considered to determine alternative specific utilities and travel mode choice. The latent variables and its indicators are based on the results of a qualitative survey on residential choices and its subsequent effects on accessibility, travel time and travel costs. I present the differences in estimates between the model with and without latent variables.

The following text is structured as follows: The next section formulates behavioral hypotheses on the effects of latent attitudinal characteristics on travel mode choice. The second section describes the data. The third section develops statistical model of travel mode choice. The fourth section describes the latent variables constructs and the hypothesis test. The fifth section presents the model estimates and discusses results while the sixth section summarizes key findings and outlines possible directions of future research.

1. Background and hypothesis formulation

The research question explored in this paper is how lifestyle, indicating preferences towards a particular way of living affects travel behavior and travel mode choice in particular. Discrete choice models have been trying to employ psychological factors since 1980s (See, for example, Ben-Akiva et al. 1999, Walker and Li 2006). A guiding idea in this effort is that the incorporation of psychological or psychometric factors leads to a more behaviorally realistic representation of the choice process, and consequently, better explanatory power.

This study was preceded by a qualitative survey on *Meaning of accessibility in decisions about residential relocation* (Braun Kohlová, 2008) which demonstrates that the choice of travel mode for regular trips within urban area is besides "constraints variables" affected by the individual or family lifestyles comprising of housing and neighborhood preferences. In this qualitative survey two basic subpopulations were identified. Individuals in the first group stated great importance of a good access of their new place of residence meaning mostly public transport connection of good quality to the city center and other frequently visited destinations, short travel time, few mode changes etc. For the individuals in the second group the access does not matter because they "use car anyway". The qualitative survey further shows that the individuals in the first group in varying degree succeed to meet the claim on good access when moving which affects their travel time, travel costs and subsequent mode choice. While those who succeed in moving to the place of residence with good accessibility tend to choose public transport and non-motorized means of transport, those, who do not, subsequently choose car as "reluctant drivers or car passengers".

Although the instruments of a qualitative survey provide a valuable piece of information on how the travel mode choice process might be determined by the residential characteristics it is necessary to validate the effect on a representative data sample and in a reliable statistical model.

I employ a choice model with latent variables to represent this behavior, where the latent variables are capturing lifestyles concerning place of residence and neighborhood preferences and the choice model is the travel mode choice on regular trip. The choice model with latent variables including heterogeneity of taste allows inferring how preferences regarding place of residence and access impact travel mode choice. I employ a sequential (two-stage) estimation method combining factor analysis and random utility choice model (multinomial logit). Responses to attitudinal survey questions are used as indicators of the latent psychological factors. Inclusion of latent (psychological) factors in the utility function improves the explanatory power of the model.

Scholars have paid attention to the association between travel behavior and attitudes in various research contexts: Handy et al. (2005) showed that differences in attitudes largely explain observed differences in travel behavior between suburban and traditional neighborhoods. Their multivariate analysis of cross-sectional data shows that the differences in level of driving are largely explained by attitudes and that the effect of the built environment mostly disappears when attitudes and socio-demographic factors have been accounted for. Ory and Mokhtarian (2005) used attitudes as explanatory variables for survey responses related to how much subjects like to travel for various purposes. Schwanen and Mokhtarian (2005) modeled commute mode choice accounting for "cognitive dissonance" (the mismatch between ones current neighborhood type and ones preference for neighborhood type) as measured by attitudinal indicators. They found that "neighborhood type dissonance is statistically significantly associated with commute mode choice: dissonant urban residents are more likely to commute by private vehicle than consonant urbanites but not quite as likely as true suburbanites. However, differences between neighborhoods tend to be larger than between consonant and dissonant residents within a neighborhood" (Schwanen and Mokhtarian, 2005, 83).

2. Data

The data were obtained from a travel behavior survey conducted in 7 selected Czech cities and their suburban areas in 2008. I make use of the revealed preference survey of individual travel mode choices that was administered to 1723 adult individuals including retiree. The populations of the smallest city in the sample is 10 thousands, the biggest (capital Prague) one million. We used quota sampling (quotas for gender, age, educational level, economic activity and residential locality/address). The main sampling criterion was the fact that the respondent realized at least one trip during the previous working day and that the trip was realized only within the respective urban area. The travel mode choice model is based on revealed preferences elicited as characteristics of the first trip and the following trips chain made on a randomly selected working day. Thus trips of different purposes (not only work trips) are included in the sample. Only trips within the respective urban area were included (radius of public transport service or 25 kilometers). The total choice set included twelve options (*car as driver, car as passenger, motorcycle, tram, trolleybus, bus, subway, train, regional bus, bicycle, walk, other*). The available options varied across localities. Motorcycle was represented in only 3 cases and therefore excluded from the analysis. For this analysis six public transport modes were included into a common alternative (*public transport)*. Besides the attributes of travel mode choice on selected day and socio-economic characteristics of individuals indicators of attitudes toward different lifestyles regarding preferred neighborhood and its accessibility were elicited. Two modified indicators on *neighborhood preferences* suggested by Walker and Li (2006) were appended with three *accessibility* indicators formulated as results of qualitative pre-survey conducted before the survey (Table 2). All together five lifestyle indicators were included into the questionnaire.

Data collection - one-on-one interviews - was done in respondents household using paper questionnaires. The data were collected by a professional agency. After data cleaning, the estimation results presented below are based on responses from 1438 individuals.

3. Model formulation

I include latent factors into the random utility choice model because my hypotheses are that lifestyle (neighborhood and access) preferences exist, that they are not directly identifiable from the data and that people with different values of preference factors exhibit different travel behavior.

A convenient functional form for analyzing the influence of potential explanatory variables on a categorical dependent variable is the multinomial logit (MNL) model. The MNL model assumes that travelers have unobservable, latent preferences or utilities for different transport modes and that they choose the mode providing the highest utility (Ben-Akiva and Lerman, 1985). The utility associated with a transportation mode consists of two components - a deterministic part reflecting the influence of observed factors relating to socio-demographic, mobility limitations, personality, lifestyle etc. and a random part capturing all unobserved impacts.

The applied random utility choice model assumes that the probability that alternative i in C_n is chosen by decision maker n is given by

$$
\begin{aligned}
P_n(i) &= \Pr(U_{in} \geq U_{jn}, \forall j \in C_n) \\
P_n(i) &= \Pr\left[V_{in} + \varepsilon_{in} \geq \max_{j \in C_n, j \neq i} (V_{jn}, + \varepsilon_{in}) \right]
\end{aligned}
\tag{1}
$$

The choice probability for alternative i is given by

$$
P_n(i) = \frac{e^{V_{in}}}{\sum_{j \in C_n} e^{V_{jn}}} \tag{2}
$$

If we assume that $U_{in} = V_{in} + \varepsilon_{in}$, for all $i \in C_n$, and that all the disturbances ε_{in} are independently and identically distributed and Gumbel-distributed with a location parameter η, and a scale parameter $\mu > 0$, then

$$
P_n(i) = \frac{e^{\mu V_{in}}}{\sum_{j \in C_n} e^{\mu V_{jn}}} \tag{3}
$$

If V_{in} and V_{jn} are linear in their parameters,

$$
P_n(i) = \frac{e^{\mu \beta' X_{in}}}{\sum_{j \in C_n} e^{\mu \beta' X_{jn}}} \tag{4}
$$

Where X_{in}, X_{jn} are vectors of observable characteristics of individuals and attributes of alternatives.

The universal choice set C includes 5 alternatives:

1. car as driver (CAR),
2. car as passenger (PASS),
3. bicycle (BIKE),
4. walk (WALK) and
5. public transport (TRANS).

Not all alternatives are assumed to be available for each individual. The set of availability variables (CAR_AV, PASS_AV, BIKE_AV, WALK_AV, TRANS_AV) specifies which alternatives are feasible for each individual. The following rules were used, defined as follows:

1. Anybody without a driver's license cannot drive alone.
2. Anybody in a household without a car cannot drive alone.
3. Anybody in a household without a bicycle cannot use bicycle.

The remaining alternatives were assumed to be feasible to everybody.

The variables entering into the utility functions are following (denoted as in parentheses):

Alternatives attributes (alternative specific varying over alternatives):
Car travel time (min for single trip incl. in-vehicle and out-of-vehicle time): CAR_TIME Car travel cost (CZK[1]): CAR_COST Car passenger travel cost (CZK): PASS_COST Bicycle travel time (min): BIKE_TIME Walk time (min): WALK_TIME Public transport travel cost (out-of-pocket cost in CZK; 0 if individual owns PT pass): TRANS_COST Public transport travel time (min for single trip incl. in-vehicle and out-of-vehicle time): TRANS_TIME Note: Bicycle and walk specific cost is assumed to equal zero and therefore not having an effect on linear-in-parameters utility
Socio-economic characteristics:
Gender (1=male): GENDER Personal net month income (11 income categories) – *including input missing*: P_INC The individual has at least one subordinate in work (1, 0): BOSS Public transport pass ownership (1, 0): ABO
Latent variables (generic):
Importance of travel accessibility: ACCESS Enjoyment of downtown /city lifestyle: CITYLIFE

The specification of the systematic utilities is given in Table 1.

Table 1: Travel mode choice model specification

	β_1	β_2	β_3	β_4	β_5	β_6	β_7
Car	ASC_CAR	0	0	0	CAR_COST	0	0
Pass	0	ASC_PASS	0	0	0	PASS_COST	0
Bike	0	0	ASC_BIKE	0	0	0	0
Walk	0	0	0	ASC_WALK	0	0	0
Trans	0	0	0	0	0	0	TRANS_COST

[1] Exchange rate ranged from 24 to 25 CZK for EURO in the period of data collection: May – June 2008.

	β_8	β_9	β_{10}	β_{11}	β_{12}	β_{13}	β_{14}
Car	CAR_TIME	0	0	0	GENDER	P_INC	BOSS
Pass	CAR_TIME	0	0	0	0	0	0
Bike	0	BIKE_TIME	0	0	GENDER	0	0
Walk	0	0	WALK_TIME	0	0	0	0
Trans	0	0	0	TRANS_TIME	0	0	0
	β_{15}	β_{16}	β_{17}				
Car	0	*0*	*0*				
Pass	0	*0*	*0*				
Bike	0	*0*	*0*				
Walk	0	*0*	*FAC_CITYLIFE*				
Trans	ABO	*FAC_ACCESS*	*FAC_CITYLIFE*				

4. Latent variables

Based on the results of the preceding qualitative survey and literature (Walker and Li, 2006) attitudinal questions capturing desired characteristics of neighborhoods and its public transport accessibility were included into the questionnaire. The responses were measured using five-point Likert-type scale of agreement. The wording of the employed questions is displayed in Table 2.

Table 2: Indicators of latent variables

How much do you agree with the following? *
I am willing to travel 15 minutes longer a trip to live in green, quiet locality and realize my hobbies.
I like to live within walking distance to shops, restaurants and other amenities.
I enjoy the hustle and bustle of the city.
I consider very important the public transport access of my neighborhood from the city center.
I consider very important the public transport access to frequently visited destinations of mine.

*The neighborhood and accesses indicators / variables were measured using five-point Likert-type scale of agreement

To reduce the information gained through individual responses I first perform factor analysis. Application of factor analysis presumes that relatively few latent variables may underlie the resulting data. I used principal component method of factor extraction and Varimax rotation which minimizes the number of variables that have high loadings on each factor. The two extracted factors explained 64.2 % of variance in the observed variables (for details see table I in Appendix). The rotated component matrix (see table II in Appendix) helped to formulate an interpretation of the components. The first component captures the importance of transport accessibility of home, city center and frequently visited destinations. The second component captures positive attitudes toward living in neighborhoods with shops, restaurants and other amenities with downtown character. Both extracted factor scores were saved as variables. The extracted factor scores entered in the utility of travel choice model alternatives as generic variables, varying among individuals, but not alternatives.

The hypothesis test

To test whether the inclusion of latent variables capturing access and city life preferences into the travel mode choice model increase its explanatory power, I use the likelihood ratio test (Ben-Akiva and Lerman, 1985, 164-167). I first estimate the unknown parameters in the restricted model including only trip specific and socio-economic variables. Then I compare the log likelihood functions of the restricted model and unrestricted model including two latent variables. Hence, the null hypothesis is

$$H_0: \beta_{ACCESS} = \beta_{CITYLIFE} = 0$$

And the test statistic for null hypothesis is given by

$$-2 (L_R - L_U)$$

Which is asymptotically distributed as χ^2 with df = $K_U - K_R$ degrees of freedom, where K_U and K_R are the numbers of estimated parameters in the unrestricted and restricted models, respectively.

I reject the null hypothesis that the restrictions are true because, using $\alpha = 0.05$,

$$- 2 (- 990.178 + 963.787) = 9.392 > 5.99$$

I therefore conclude that the latent variables capturing access and city life preferences should be included in the travel mode choice model, because it increases the explanatory power of the model.

5. Estimation results

The unknown parameters are estimated using maximum likelihood estimation. The software BIOGEME (Bierlaire 2003, 2008) is used for estimation of multinomial logit model. The software SPSS is used for the estimation of latent variables scores.

For estimation purposes, I have normalized the alternative specific constant of public transport to zero. The estimation results for the model specification above are shown in Table 3. The results show the statistical significance of two alternative specific constant, ASC_{PASS} and ASC_{WALK}. The former has a negative sign indicating a negative preference of travelling as passenger in car with respect to means of public transport. The latter has a positive sign and a higher value indicating greater positive preference of walking. The alternative specific constants for car (as driver) alternative and bicycle alternative have both negative signs, indicating negative preference for these alternatives, with all the rest remaining constant, but are statistically significant only at 10% level. As expected, car cost and transit cost, both significant at 5% level, have negative sign meaning negative effect on utility of respective alternatives. Transit cost coefficient is about 1.6 time higher showing greater effect of transit cost compared to car cost. The passenger specific cost coefficient has a positive sign, but is not statistically significant. All travel time coefficients are statistically significant with negative signs. The differences in values indicate the highest negative effect of a minute spent walking and lowest of a minute spent on or off public transport vehicle. As expected ownership of public transport pass has statistically significant positive and high effect on the public transport utility. Three socio-economic characteristics of individuals have statistically significant and positive effect on the related utility, i.e. personal income and superior position in work on car utility, male gender on car and bicycle utility. Interestingly, the effect of superior position in work and gender is considerably higher than that of personal income, which indicates that the utility function is beyond physical or objective constraint affected by psychological or social behavioral factor. We can hypothesize that driving a car is an indispensable attribute of superior position in organizational hierarchy or male identity.

Two latent variables capturing the impact of good access and city life preference, which eventually entered the model, have statistically significant and positive impact on transit and walk utility. It indicates that the more individual (verbally) prefers good public transport access and city lifestyle the higher is the transit and walk utility with respect to car both as driver and passenger and bicycle.

Including the latent variables into the travel mode choice model does not significantly change the estimates of the other coefficients. The only exception is transit cost, the coefficient of which decreased from -0.0207 (with robust std. error 0.00727) to -0.0137 (with robust std. error 0.00766) and stopped being statistically significant. It indicates that the coefficients of latent variables capturing access and city lifestyle preferences take away a part of the transit cost variability. We can conjecture that the more individuals prefer good public transport access and city lifestyle the less important are for them the transit cost. However the interpretation might be double; these people either prefer transit and use is disregarding the cost or they neglect the cost because they can walk thanks to shorter travel distances.

To summarize the results, the estimated parameters are both significant and have the expected signs. The log likelihood increased from the initial value -1704.182, to the final value -990.178 (without latent variables) and -963.787 (with latent variables). The summary statistics for both models are shown in Table 4.

Table 3: Travel mode choice multinomial logit model with latent variables

	Value	Robust Std err	Robust t-test	P-value	
ASC_{BIKE}	-0.436	0.250	-1.75	0.08	*
ASC_{CAR}	-0.495	0.274	-1.81	0.07	*
ASC_{PASS}	-2.18	0.175	-12.43	0.00	
ASC_{WALK}	3.92	0.520	7.55	0.00	
β_{CAR_COST}	-0.00844	0.00401	-2.10	0.04	
β_{PASS_COST}	0.00600	0.0136	0.44	0.66	*
β_{TRANS_COST}	-0.0137	0.00766	-1.79	0.07	*
β_{BIKE_COST}	-0.0479	0.00684	-7.00	0.00	
β_{CAR_TIME}	-0.0191	0.00685	-2.78	0.01	
β_{TRANS_TIME}	-0.0165	0.00486	-3.38	0.00	
β_{WALK_TIME}	-0.145	0.0259	-5.58	0.00	
β_{ABO}	2.28	0.241	9.47	0.00	
β_{GENDER}	0.654	0.153	4.28	0.00	
β_{INCOME}	0.118	0.0359	3.30	0.00	
β_{BOSS}	0.689	0.198	3.49	0.00	
β_{ACCESS}	0.505	0.0743	6.79	0.00	
$\beta_{CITYLIFE}$	0.221	0.0743	2.97	0.00	

Table 4: Multinomial logit model: summary statistics

Summary statistics:	Without latent variables	With latent variables of lifestyle
Number of observations:	1438	1438
Null log-likelihood:	-1704.182	-1704.182
Final log-likelihood:	-990.178	-963.787
Likelihood ratio test:	1428.007	1480.789
Adjusted rho-square:	0.410	0.424

6. Summary and future directions

The above empirical study supports the hypothesis that lifestyle preferences exist and that they are determinants of travel mode choice. In the paper discrete choice model estimates are presented which include beyond the attributes of trip and socio-economic predictors also latent variables capturing the differences in individual preferences of residence and access to facilities (urban versus suburban). The indicators of latent variables are based on the results of the qualitative survey on residential choices and its subsequent effects on accessibility, travel time and travel costs. A guiding idea in this effort was that the incorporation of psychological or psychometric factors leads to a more behaviorally realistic representation of the choice process, and consequently, better explanatory power.

A sequential (two-stage) estimation method combining factor analysis and random utility choice model (multinomial logit) was employed. The applied likelihood ratio test of the models with and without latent variables confirmed that the inclusion of latent variables capturing access and city life preferences into the travel mode choice model increase its explanatory power. The estimates of the other determinants were not significantly affected by the inclusion of the latent variables.

The research demonstrates the potential of latent constructs in uncovering heterogeneity of lifestyle and travel mode choice preferences. By making explicit connection between travel mode choice and neighborhood and accessibility preferences we can better understand the manner in which an urban and suburban area develops and the impact that changes in urban form will have on modal split and its impacts on environment and public health. The demonstrated connection between city lifestyle and transit choice and walk and connection between suburban lifestyle and car choice imply that policies aimed at behavior modification may be limited by the large proportion of households who have strong preferences towards suburban and car oriented lifestyle.

I aim to extent this work in several directions. The two-stage estimates including factor loading estimates and discrete choice model estimates allowed me to validate the hypothesis of the effect of latent lifestyle variables on the travel mode choice. However, to avoid measurement errors bias, it is suggested (Ben-Akiva et. al 1999, Walker and Li, 2006) to estimate latent variables scores or classes and choice model simultaneously. Another extension is to match the latent neighborhood and access preferences, which are stated, with the actual neighborhood characteristics and as such revealed preferences. I hypothesize that the framework incorporating stated and revealed preferences may extent and better specify the lifestyle segments of interest.

In summary, the choice modeling approach including latent variables of lifestyle provides a method allowing for heterogeneity of lifestyle while explaining travel mode choices. Further, the empirical results point to intricate policy implications implied by the complexity of preferences towards residential characteristics.

Acknowledgements

This research has been supported by National Czech Foundation GAČR No. 403/08/1694, *Application of the model of environmentally significant behavior in the Czech Republic* and by the R&D project 2D06029 VaV/320/1/03 *Analysis of distribution and social effects of sector policies* funded by the Czech Ministry of Education Youth and Sports. This support is gratefully acknowledged.

References:

BEN-AKIVA, M.; LERMAN, S. *Discrete Choice Analysis: Theory and Application to Travel Demand*. Cambridge (Mass.): The MIT Press, 1985.

BEN-AKIVA, M.; WALKER, J.; BERNARDINO, A. T.; GOPINATH, D. A.; MORIKAVE, T.; POLYDOROPOULOU, A. 1999. *Integration of Choice and Latent Variable Models*. MIT Working Paper.

BIERLAIRE, M. (2003). BIOGEME: A free package for the estimation of discrete choice models , Proceedings of the 3rd Swiss Transportation Research Conference, Ascona, Switzerland.

BIERLAIRE, M. (2008). An introduction to BIOGEME Version 1.7, biogeme.epfl.ch.

BRAUN KOHLOVÁ, M. (2008). Meaning of accessibility in decisions about residential relocation – results of a qualitative sociological survey in Czech cities. *In: Becker, U.; Gerike, R. (Ed.) (2008) How to define and measure access and need satisfaction in transport.* Issue 7 in the series of Dresden Institute for Transportation and Environment (DIVU), Dresden.

HANDY, S., CAO, X., MOKHTARIAN, P.L. (2005). Correlation or causality between the built environment and travel behavior? Evidence from Northern Kalifornia. *Transportation Research Part D 10*, 427–444.

ORY, D.T., MOKHTARIAN, P.L. (2005). When is getting there half the fun? Modeling the liking for travel. *Transportation Research Part A 39,* 97–123.

SCHWANEN, T., MOKHTARIAN, P.L. (2005). What affects commute mode choice: neighborhood physical structure or preferences toward neighborhoods? *Journal of Transport Geography* 13, 83–99.

WALKER, J.; LI, J. Latent Lifestyle Preference and Household Location Decisions. *Journal of Geographical Systems*, 2007, vol. 9, no. 1, s. 77–101.

Appendix:

Table I: Factor analysis: total variance explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	1.940	38.805	38.805	1.940	38.805	38.805	1.861	37.226	37.226
2	1.270	25.391	64.195	1.270	25.391	64.195	1.348	26.969	64.195
3	0.965	19.299	83.494						
4	0.632	12.641	96.136						
5	0.193	3.864	100.000						

Extraction Method: Principal Component Analysis.

Table II: Rotated component Matrix (factor loadings)

	Component	
	1	2
	ACCESS	CITYLIFE
I am willing to travel 15 minutes longer a trip to live in green, quiet locality and realize my hobbies.	0.202	**-0.576**
I like to live within walking distance to shops, restaurants and other amenities.	0.250	**0.589**
I enjoy the hustle and bustle of the city.	0.110	**0.815**
I consider very important the public transport access of my neighborhood from the city center.	**0.931**	0.048
I consider very important the public transport access to frequently visited destinations of mine.	**0.937**	0.054

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.