Hybrid Choice Modeling of New Technologies for Car Use in Canada

Denis Bolduc, Nathalie Boucher, Ricardo Alvarez-Daziano
Département d'Économique, Université Laval

Abstract

In the last decade, a new trend in discrete choice modeling has emerged in which psychological factors are explicitly incorporated in order to enhance the behavioral representation of the choice process. In this context, Hybrid Models expand on standard choice models by including attitudes and perceptions as latent variables.

The complete model is composed of a group of structural equations describing the latent variables in terms of observable exogenous variables, and a group of measurement relationships linking latent variables to certain observable indicators. Although the estimation of Hybrid Models requires the evaluation of complex multi-dimensional integrals, simulated maximum likelihood is implemented in order to solve the integrated multi-equation model.

In this paper we study empirically the application of Hybrid Choice Modeling to data from a survey conducted by the EMRG (Simon Fraser University, 2002-2003) of virtual personal vehicle choices made by Canadian consumers when faced with technological innovations. The survey also includes a complete list of indicators, allowing us to apply a Hybrid Choice Model formulation.

We conclude that Hybrid Choice is genuinely capable of adapting to practical situations by including latent variables among the set of explanatory variables. Incorporating perceptions and attitudes in this way leads to more realistic models and gives a better description of the profile of consumers and their adoption of new private transportation technologies.

1. Introduction

In the current context of global warming, it is more pertinent than ever to explore sustainable solutions to problems created by personal transportation. Technological innovation and the use of cleaner alternative fuels are among the solutions that have traditionally been proposed to contain, or at least alleviate, these problems. In this way it becomes important to model the effect of introducing new technologies for car use on transportation choice behavior in a more realistic and expanded way. The modeling challenge is to develop transportation demand models that are not only able to adequately predict individual preferences but also to recognize the impact of psychological factors, such as perceptions and attitudes.

In this paper we analyze the practical implementation of hybrid choice models in order to include perceptions and attitudes in a standard discrete choice setting. In section 2, we first discuss the derivation of discrete choice models (section 2.1) and their integration into a hybrid model framework using latent variables (section 2.2). Then we concentrate on how to estimate a hybrid model (section 2.3). In section 3, we present the empirical data about car choice made by Canadian consumers, while in section 4 we show the application of a Multinomial Logit model. In section 5 we expand this standard model showing the results of each partial model that configures the Hybrid model setting. We finish by presenting the main conclusions of this work (section 6).

2. Hybrid Choice Modeling

2.1. Standard Discrete Choice Modeling

In discrete choice modeling, the most common approach is based on *random utility theory* (McFadden, 1974). According to this theory, each individual has a *utility function* associated with each of the alternatives. This individual function can be divided into a systematic part, which considers the effect of the explanatory variables, and a random part that takes into account all the effects not included in the systematic part of the utility function. In other words, choices are modeled using a *structural equation* (1) – the utility function – representing the individual preferences, where the inputs are the alternative attributes and individual characteristics. The observed choice corresponds to the alternative which maximizes the individual utility function, a process represented by a *measurement equation* (2). Because the utility function has a random nature, the output of the model actually corresponds to the choice probability of individual *n* choosing alternative *i*.

The set of equations describing the standard discrete choice setting is given by:

$$U_{in} = X_{in}\beta + \upsilon_{in} \tag{1}$$

$$y_{in} = \begin{cases} 1 & if\ U_{in} \geq U_{jn},\ j \neq i \\ 0 & otherwise, \end{cases} \tag{2}$$

where U_{in} corresponds to the utility of alternative *i* as perceived by individual *n*; X_{in} is a row vector of attributes of alternative *i* and socioeconomic characteristics of individual *n*; β is a column vector of unknown parameters; υ_{in} is an error term and y_{in} corresponds to an indicator of whether alternative *i* is chosen by individual *n* or not.

Different choice models can be derived depending on the assumptions considered for the distribution of the random error term (Ben-Akiva and Lerman, 1985). So far the workhorses in this area have been the Multinomial Logit model (McFadden, 1974) and the Nested Logit model (Williams, 1977). Both offer closed choice probabilities but with not always properly justified restrictive simplifying assumptions. In an aim to gain generality, more flexible models have been incorporated in practice. Indeed, one powerful modeling alternative is the Logit Mixture model (Ben-Akiva and Bolduc, 1996; Brownstone and Train, 1999), which can approximate any random utility maximization model (McFadden and Train, 2000). The main idea of this kind of model is to consider more than one random component, allowing the presence of a more flexible covariance structure. The estimation implies the evaluation of integrals without a closed form, although it is possible to use computer aided simulation.

In this way it can be seen that the development of discrete choice modeling has evolved quickly and that powerful models can be used; however, under this whole approach discrete choice models represent the decision process as an obscure black box, where attitudes, perceptions and knowledge are neglected. This is why in the last decade discrete choice modeling has evolved towards an explicit incorporation of psychological factors affecting decision making (Walker *et al.*, 2005). According to 2002 Nobel Laureate Daniel Kahneman, there still remains a significant difference between economist modelers that develop practical models of decision-making and behavioral scientists which focus on in-depth understanding of agent behavior. Both have fundamental interests in behavior but each work with different assumptions and tools. In a 1986 paper, Daniel McFadden point out the need to bridge these worlds by incorporating attitudes in choice models. In his 2000 Nobel lecture, McFadden emphasized the need to incorporate attitudinal constructs in conventional economics models of decision making.

2.2. Latent Variables and Discrete Choice: The Hybrid Choice Model

The new trend in discrete choice modeling, initially promoted by McFadden and Ben-Akiva (see Brownstone et al, 2002), is to enhance the behavioral representation of the choice process. As a direct result, the model specification is improved and the model gains in predictive power. Econometrically, the improved representation model involves dealing with a choice model formulation that contains unobserved psychometric variables (attitudes, opinions) among the explanatory variables incorporated as latent variables (see Figure 1).

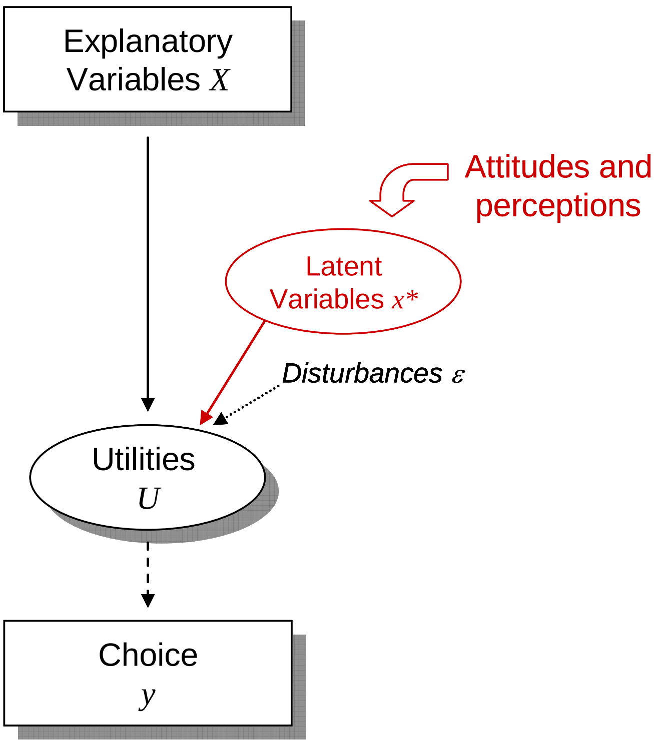


Figure 1: Latent variables and discrete choice

Since the latent variables are not observed, they are normally linked to questions of a survey: the *indicators*. These indicators can be continuous, binary or categorical variables expressed by responses to attitudinal and perceptual survey questions. The latent variable model is composed of a *structural equation*, which describes the latent variables in terms of observable exogenous variables, and a group of measurement relationships (*measurement model*) linking latent variables to indicators (Jöreskog and Sörbom, 1984). It is possible to integrate the latent variable model into the standard choice model setting, obtaining a group of structural and measurement equations – the Hybrid Choice model – which may be written as follows:

Structural equations

$$x_n^* = Bw_n + \zeta_n, \quad \zeta_n \sim N(0, \Psi) \tag{3}$$

$$U_n = X_n\beta + Cx_n^* + \upsilon_n, \tag{4}$$

Measurement equations

$$I_n = \alpha + \Lambda x_n^* + \varepsilon_n, \quad \varepsilon_n \sim \mathrm{N}(0, \Theta) \tag{5}$$

$$y_{in} = \begin{cases} 1 & \text{if } U_{in} \geq U_{jn},\ j \neq i \\ 0 & \text{otherwise,} \end{cases} \tag{6}$$

where x_n^* is a $(L \times 1)$ vector of latent variables; w_n is a $(M \times 1)$ vector of explanatory variables causing the latent variables; B is a $(L \times M)$ matrix of unknown parameters; and Ψ is a $(L \times L)$ variance covariance matrix. The choice model in equation 4 is written in vector form where we assume that there are J alternatives. Therefore, U_n is a $(J \times 1)$ vector of utilities; υ_n is a $(J \times 1)$ vector of i.i.d. Type 1 extreme value error terms. X_n is a $(J \times K)$ matrix with X_{in} designating the ith row. β is a $(K \times 1)$ vector of parameters. C is a $(J \times L)$ matrix of unknown parameters associated with the latent variables present in the utility function, with C_i designating the ith row of matrix C.

In the set of measurement equations, I_n corresponds to a $(R \times 1)$ vector of indicators of latent variables associated with individual n; α is a $(R \times 1)$ vector of constants and Λ is a $(R \times L)$ matrix of unknown parameters that relate the latent variables to the indicators. The term ε_n is a $(R \times 1)$ vector of independent error terms. We thus assume that Θ is a diagonal matrix with variance terms on the diagonal. Finally, we stack the y_{in}'s into a $(J \times 1)$ vector called y_n.

If the latent variables were not present, the choice probability would correspond exactly to the standard choice probability $P(y_n | X_n, \beta)$. In a setting with observed latent variables x_n^*, the choice probability would be represented by $P(y_n | x_n^*, X_n, \theta)$ where θ contains all the unknown parameters in the choice model of equation 4. Since latent variables are not actually observed, the choice probability is obtained by integrating the latter expression over the whole space of x_n^*:

$$P(y_n | X_n, w_n, \theta, B) = \int_{x_n^*} P(y_n | x_n^*, X_n, \theta) g(x_n^* | B, w_n) dx_n^*, \tag{7}$$

which is an integral of dimension equal to the number of latent variables in x_n^*.

Indicators are introduced in order to characterize the unobserved latent variables, and econometrically they permit identification of the latent variables. Indicators also provide efficiency in estimating the choice model with latent variables, because they add information content. The joint probability of observing y_n and I_n may thus be written as:

$$P(y_n, I_n | X_n, w_n, \delta) = \int_{x^*} P(y_n | x_n^*, X_n, \theta) f(I_n | \Lambda, x_n^*) g(x_n^* | B, w_n) dx_n^*, \tag{8}$$

where by assumption, y_n and I_n are assumed to be correlated only via the presence of the latent variables x_n^* in equations 4 and 5. The choice model $P(y_n | x_n^*, X_n, \theta)$ may be assumed to be of the logit mixture type. The indicators in I_n can be continuous, binary or categorical variables. The number of latent variables has an impact on the computation of the joint probability in equation 8. The vector δ designates the full set of parameters to estimate.

2.1. Estimation

In practice, with large number of latent variables (more than 3), we replace the multidimensional integral with a smooth simulator with good properties. Indeed, if υ_n is i.i.d. extreme value type 1 distributed, then conditional on *x** the choice probability has the Multinomial Logit form, which leads to the following expression:

$$P(y_n, I_n \left| X_n, w_n, \delta \right.) = \int_{x^*} \frac{\exp(X_{in}\beta + C_i x_n^*)}{\sum_j \exp(X_{jn}\beta + C_j x_n^*)} f(I_n \left| \Lambda, x_n^* \right.) g(x_n^* \left| B, w_n \right.) dx_n^* . \qquad (9)$$

Taking advantage of the expectation form, we can replace the probability with the following empirical mean:

$$\tilde{P}(y_n, I_n \left| X_n, w_n, \delta \right.) = \frac{1}{S} \sum_{s=1}^{S} \frac{\exp(X_{in}\beta + C_i x_n^{*s})}{\sum_j \exp(X_{jn}\beta + C_j x_n^{*s})} f(I_n \left| \Lambda, x_n^{*s} \right.) \qquad (10)$$

where x_n^{*s} corresponds to a random draw s from its distribution. This sum is computed over S draws. This simulator is known to be unbiased, consistent and smooth with respect to the unknown parameters. Replacing $P(y_n, I_n \left| X_n, w_n, \delta \right.)$ with $\tilde{P}(y_n, I_n \left| X_n, w_n, \delta \right.)$ in the log likelihood leads to a maximum simulated likelihood (MSL) solution. We therefore consider the following objective function:

$$\max_{\delta} \sum_{n=1}^{N} \ln \tilde{P}(y_n, I_n \left| X_n, w_n, \delta \right.) .$$

In the past few years, a lot of progress has been made regarding MSL estimation. Train (2002) gives an in-depth analysis of the properties of MSL estimation. Recent results, mostly attributable to Bhat (2001), suggest that simulation should exploit *Halton draws*. Halton type sequences are known to produce simulators with a given level of accuracy using many fewer draws than when using conventional uniform random draws (Ben Akiva *et al*, 2001; Munizaga and Alvarez-Daziano, 2005). Simulated maximum likelihood is now well known and has been applied in numerous circumstances. The Logit probability kernel makes the simulated log likelihood fairly well behaved. Asymptotically, meaning as $S \to \infty$ and as $N \to \infty$, the solution becomes identical to a solution arising from maximizing the usual log likelihood function $\sum_{n=1}^{N} \ln P(y_n, I_n \left| X_n, w_n, \delta \right.)$.

Regarding the measurement model, we assume that each equation that links the indicators and the latent variables corresponds to a continuous, a binary, or a multinomial response. A given measurement equation r in the continuous case is given by:

$$I_{rn} = \alpha_r + \Lambda_r x_n^* + \varepsilon_{rn}, \quad \varepsilon_{rn} \sim \mathrm{N}(0, \theta_r^2) . \qquad (11)$$

In the binary case, calling I_{rn}^*, an unobserved continuous indicator, we get instead:

$$I_{rn} = \begin{cases} 1 & \text{if } I_{rn}^* \geq 0 \\ 0 & \text{otherwise,} \end{cases} \tag{12}$$

while in the multinomial case we obtain:

$$I_{rn} = \begin{cases} 1 & if\ \gamma_0 < I_{rn}^* \leq \gamma_1 \\ 2 & if\ \gamma_1 < I_{rn}^* \leq \gamma_2 \\ \vdots & \\ L & if\ \gamma_{L-1} < I_{rn}^* \leq \gamma_L, \end{cases} \tag{13}$$

where I_{rn} and ε_{rn} are the rth element of I_n and ε_n respectively. θ_r^2 is the rth element on the diagonal of Θ, and Λ_r denotes row *r* of Λ. In the multinomial cases, the γ's are estimated. By convention, γ_0 and γ_L are fixed to $-\infty$ and ∞ respectively. For identification, the constant terms α_r must be set to 0 in the multinomial case. In the other cases –continuous and binary– it is estimable (Bolduc and Giroux, 2005).

Given our assumptions, the density $f(I_n)$ simply corresponds to:

$$f(I_n) = \prod_{r=1}^{R} f(I_{rn}). \tag{14}$$

If measurement equation *r* is continuous, then

$$f(I_{rn}) = \frac{1}{\theta_r}\phi\left(\frac{I_{rn} - \alpha_r - \Lambda_r x_n^*}{\theta_r}\right), \tag{15}$$

where ϕ denotes the probability density function (pdf) of a standard normal. If the measurement equation *r* corresponds to a binary response, then

$$f(I_{rn}) = \Phi\left(\frac{\alpha_r + \Lambda_r x_n^*}{\theta_r}\right)^{I_{rn}} \left(1 - \Phi\left(\frac{\alpha_r + \Lambda_r z_n^*}{\theta_r}\right)\right)^{(1-I_{rn})}, \tag{16}$$

where Φ denotes the cumulative distribution function (cdf) of a standard normal. Finally, if measurement equation *r* corresponds to a multinomial response, then

$$f(I_{rn} = l) = \Phi\left(\frac{\gamma_l - \Lambda_r x_n^*}{\theta_r}\right) - \Phi\left(\frac{\gamma_{l-1} - \Lambda_r z_n^*}{\theta_r}\right). \tag{17}$$

Except for the continuous case, the variances θ_r cannot be estimated. We thus fix them to 1.

3. The Data

The data for this study are from a survey conducted by EMRG (Energy and Materials Research Group, Simon Fraser University) in 2002-2003 on 1500 Canadian consumers living in urban areas (Horne, 2003). Survey participants were first contacted in a telephone interview for personalizing a detailed questionnaire that was then mailed to them. The telephone conversation allowed general information to be gathered on the socioeconomic characteristics of the participants and their households, as well as on their automobile fleet.

The mailed questionnaire consisted in:

- Part 1: Transportation Options, Requirements and Habits
- Part 2: Virtual personal vehicle choices made by Canadian consumers when faced with technological innovations
- Part 3: Transportation Mode preferences
- Part 4: Views on Transportation issues
- Part 5: Additional Information (gender, education, income)

The Stated Preferences (SP) hypothetical car choices in Part 2 considered four vehicle types:

- A conventional vehicle (operating on gasoline or diesel)
- A natural-gas vehicle
- A hybrid vehicle (gasoline-electric)
- A hydrogen fuel cell vehicle

The characteristics of the vehicles were depicted as:

- Capital Cost: purchase price
- Operating Costs: fuel costs
- Fuel Available: percentage of stations selling the proper fuel
- Express Lane Access: whether or not the vehicle would be granted express lane access
- Emissions Data: emissions compared to a standard gasoline vehicle
- Power: power compared to their current vehicle

Under the SP experimental design of the survey, each participant was asked to make up to four consecutive virtual choices while the features of the vehicle were modified after each round. The sample has 866 completed surveys (77% response rate) and since each respondent provided up to 4 vehicle choices, after some clean up, there remains 1877 usable observations. The whole SP design is described in Horne (2003). The various values assumed by the characteristics of the vehicles that were used as a basis for developing the experimental design of the SP survey are set out in Table 1. The basis of comparison of the experimental design refers to the motor vehicle that is the most frequently used by the respondent.

Table 1. SP design

Vehicle Type	**Gasoline Vehicle**	**Alternative Fuel Vehicle**	**Hybrid-Electric Vehicle**	**Hydrogen Fuel Cell Vehicle**
Number of Choices	**213**	**70**	**917**	**677**
Capital Cost CC	100% $_{CC}$ 105 % $_{CC}$ 110% $_{CC}$ 115% $_{CC}$	105 % $_{CC}$ 110 % $_{CC}$	105 % $_{CC}$ 120% $_{CC}$	110% $_{CC}$ 120% $_{CC}$
Operating Costs FC	100% $_{FC}$ 110% $_{FC}$ 120% $_{FC}$ 130% $_{FC}$	110% $_{FC}$ 120% $_{FC}$	Equals 75% Gasoline Value	110% $_{FC}$ 120% $_{FC}$
Fuel Available	100%	25% 75%	100%	25% 75%
Express Lane Access	No	No Yes	Equals AFV Value	No Yes
Emissions Data (compared to current vehicle)	Equal	10% Less	25% Less	100% Less
Power Compared to Current Vehicle	Equal	Equal	Equal	Equal
		10% Less	10% Less	10% Less

4. Multinomial Logit Results

We first apply a standard Multinomial Logit model, using as variables the list of attributes for the vehicle choice. The choice model is depicted in Figure 2, while in Table 2 we present the estimation results, which are equivalent to those reported by Horne (2005).

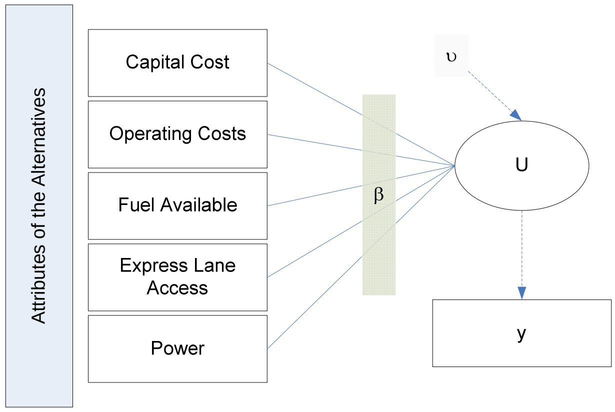


Figure 2. Choice model

Table 2. MNL Results

	Estimates	**Robust t.student**
ASC_AFV	-4.500	-6.81
ASC_HEV	-1.380	-2.18
ASC_HFC	-2.100	-3.26
capcost	-0.856	-4.07
fuelcost	-0.826	-4.18
fuelavail	1.360	7.32
expaccess	0.156	2.29
power	2.700	4.12

Number of individuals	1877
Final log-likelihood	-1984.55
Adjusted rho-square	0.234

Interpretation comes here.

5. Hybrid Choice Model Results

5.1. Definition of the Latent Variables

The first step to build a Hybrid Choice model is to define the latent variables involved. In the survey people answered to the following questions:

- Question 7: Considering your transportation requirements, do you think your family could meet its needs with fewer vehicles, either by traveling less, or using different methods of transportations? Yes/No

- Question 30: What is your level of support for/opposition to the following government actions that would influence your transportation system? Degree of support: 7 levels from Strongly Opposed to Strongly Supportive.

    - Improving traffic flow by building new roads and expanding existing roads
    - Discouraging automobile use with road tolls, gas taxes and vehicle surcharges
    - Making neighborhoods more attractive to walkers and cyclists using bike lanes and speed controls
    - Reducing vehicle emissions with regular testing and manufacturer emissions standards
    - Making carpooling and transit faster by giving them dedicated traffic lanes and priority at intersections
    - Making transit more attractive by reducing fares, increasing frequency, and expanding route coverage
    - Reducing transportation distances by promoting mixed commercial, residential and high-density development
    - Reducing transportation needs by encouraging compressed workweeks and working from home

- Question 33: Thinking about your daily experiences, how serious do you consider the following problems related to transportation to be? Degree of seriousness of problem: 7 levels from Not a Problem to Major Problem.

    - Traffic congestion you experience while driving
    - Traffic noise you hear at home, work, or school
    - Vehicle emissions which impact local air quality
    - Accidents caused by aggressive or absent-minded drivers
    - Vehicle emissions which contribute to global warming
    - Unsafe communities due to speeding traffic

- Question 6: What importance did the following factors have in your family's decision to purchase this vehicle? Degree of importance: 7 levels from Not at All Important to Very Important.

    - Purchase Price
    - Vehicle Type
    - Fuel Economy
    - Horsepower
    - Safety
    - Seating Capacity
    - Reliability
    - Appearance and Styling

Considering those questions as indicators we identify two latent variables:

- **Environmental Concern (EC)**: related to transportation and its environmental impact

- **Appreciation of new car features (ACF)**: related to car purchase decisions and how important are the characteristics of this new alternative

The Hybrid Model setting we consider is represented in Figure 3, where the complete set of structural and measurement equations is sketched, depicting the relationships between explanatory variables and each partial model. Indeed, we can distinguish the Choice Model, which is centered on the utility function modeling; the latent variables structural model, which links the latent variables with the characteristics of the traveller and the latent variables measurement model, which link each latent variable with the indicators.

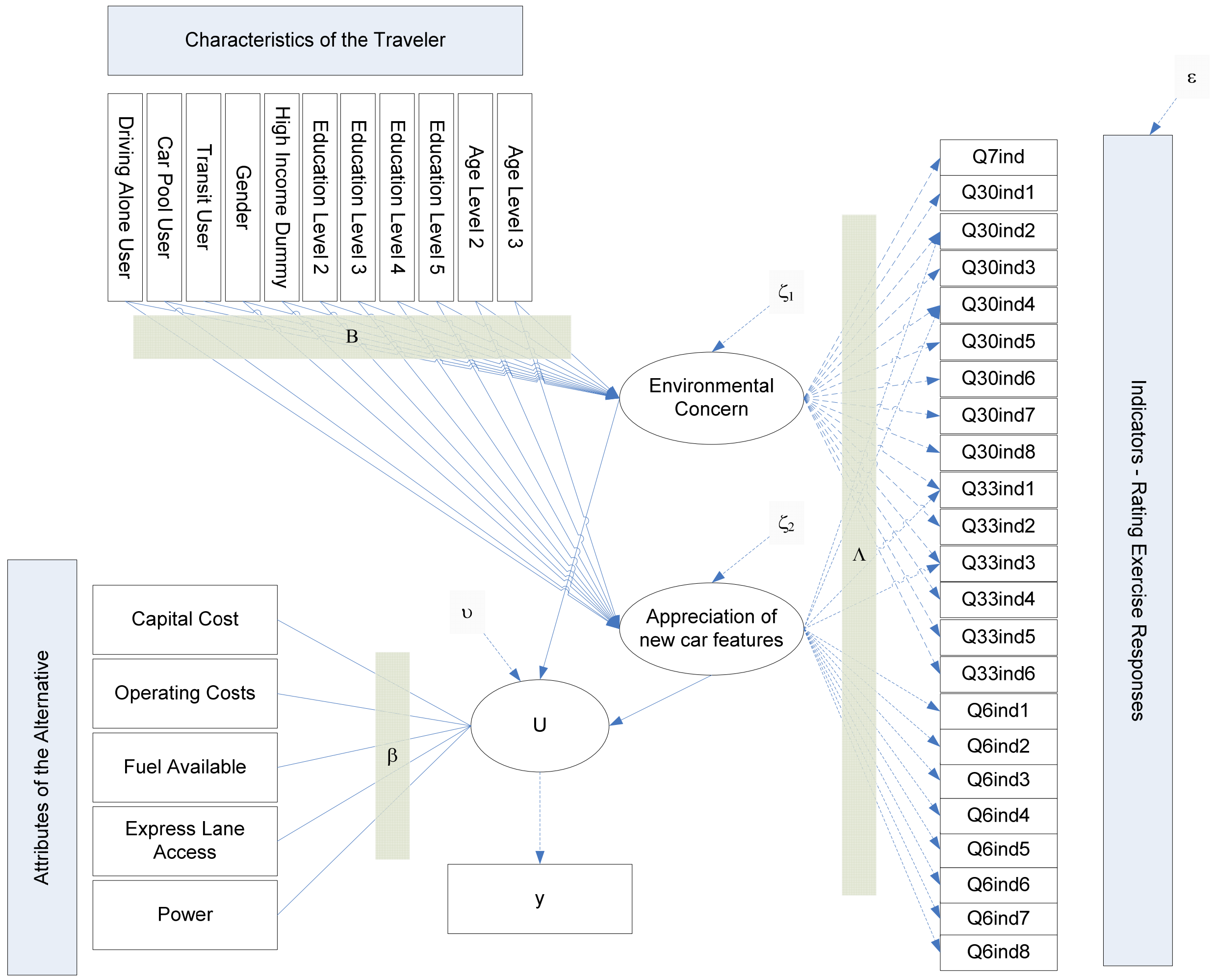


Figure 3. Hybrid Choice model

5.2. Car Choice Model

The set of equations for the mode choice model alone are given by:

$$U_{SGVn} = V_{SGVn} + c_{1,2}x^{*}_{2n} + \nu_{SGVn} \tag{18}$$

$$U_{AFVn} = V_{AFVn} + c_{2,1}x^{*}_{1n} + c_{2,2}x^{*}_{2n} + \nu_{AFVn} \tag{19}$$

$$U_{HEVn} = V_{HEVn} + c_{3,1}x^{*}_{1n} + c_{3,2}x^{*}_{2n} + \nu_{HEVn} \tag{20}$$

$$U_{HFCn} = V_{HFCn} + c_{4,1}x^{*}_{1n} + c_{4,2}x^{*}_{2n} + \nu_{HFCn} \tag{21}$$

Note that the model shares a common deterministic part with the standard Multinomial Logit model. However we add the effect of the latent variables on each utility function. The latent variable related to environmental concern was not considered on the standard fuel vehicle. In Table 3, we report the results of the choice model.

Table 3. Car Choice Model Results

	Estimates	**Robust t.student**
ASC_AFV	-4.494	-4.530
ASC_HEV	0.022	0.040
ASC_HFC	-1.700	-3.053
Capcost	-0.880	-4.056
Fuelcost	-0.868	-4.095
Fuelavail	1.404	7.196
Expaccess	0.168	2.327
Power	2.784	4.094
Latent Variables		
ACF on SGV	1.557	10.389
EC on AFV	0.289	3.012
ACF on AFV	1.115	4.747
EC on HEV	0.294	4.844
ACF on HEV	0.576	5.358
EC on HFC	0.441	7.026
ACF on HFC	0.712	6.201

Number of individuals	1877
Final global function	- 61477.52
Final global function c=0	- 61531.85
Adjusted rho-square	0.235
Number of Halton draws	500

Common parameters with the standard multinomial logit model have the same sign and magnitude, except for alternative specific constants. In addition, the significance of the latent variables' parameters shows that they have a significant effect on the individual utilities.

5.3. Latent Variables Structural Model

The set of equations for the latent variables structural model is given by:

$$\begin{aligned} x_{1n}^{*} = {} & B_{1,1}Intercept + B_{1,2}user\_car + B_{1,3}user\_carpool + B_{1,4}user\_transit + B_{1,5}gender + B_{1,6}HighInc_5 \\ & + B_{1,7}EDUC_2 + B_{1,8}EDUC_3 + B_{1,9}EDUC_4 + B_{1,10}EDUC_5 + B_{1,11}AGE_2 + B_{1,12}AGE_3 + \varsigma_{1n} \end{aligned} \tag{22}$$

$$\begin{aligned} x_{2n}^{*} = {} & B_{2,1}Intercept + B_{2,2}user\_car + B_{2,3}user\_carpool + B_{2,5}gender + B_{2,6}HighInc_5 \\ & + B_{2,7}EDUC_2 + B_{2,8}EDUC_3 + B_{2,9}EDUC_4 + B_{i,10}EDUC_5 + B_{2,11}AGE_2 + B_{2,12}AGE_3 + \varsigma_{2n} \end{aligned} \tag{23}$$

Table 4. Latent Variables Structural Model Results

	EC		**ACF**	
	Estimates	**Robust t.student**	**Estimates**	**Robust t.student**
Intercept	2.832	7.234	1.415	5.346
Driving Alone User : *user_car*	-0.044	-0.320	0.297	6.950
Car Pool User : *user_carpool*	0.307	1.171	0.242	2.596
Transit User : *user_transit*	0.674	3.021	-	-
Gender (Female Dummy)	0.697	5.764	0.185	4.614
High Income Dummy (>80,000$)	0.650	0.927	0.119	0.500
Education level 2 (Grade 9)	-0.120	-0.184	0.154	0.739
Education level 3 (Grade 12)	0.044	0.067	0.138	0.680
Education level 4 (College)	0.317	0.473	0.095	0.473
Education level 5 (University)	0.521	2.166	0.520	7.009
Age level 2 (26 to 40 years)	0.898	3.665	0.741	8.910
Age level 3 (41 to 55 years)	0.913	3.556	0.834	9.392

The structural equation links consumer characteristics with the latent variables. For example, we can conclude that environmental concern is more important for transit users than for car pool users. We in fact observe a negative parameter for a driving alone user. We can also see that mean environmental concern increases with the education level.

Note that we also estimate the elements of the covariance matrix. As the results show, the elements in the diagonal are significant and show the presence of a heteroskedastic nature. In this version, we assumed that the two latent variables are uncorrelated.

Table 5. Latent Variables Structural Model Covariates

	Estimates	**Robust t.student**
Var(EC)	1.787	11.820
Var(ACF)	0.643	23.115

5.4. Latent Variables Measurement Model

Finally, the latent variable measurement model links the latent variables with the indicators, and a typical equation for this model has the form:

$$q_{30ind2} = \alpha_{q30ind2} + \lambda_{1,q30ind2} x^{*}_{1n} + \lambda_{2,q30ind2} x^{*}_{2n} + \varepsilon_{30ind2}\,. \tag{24}$$

In this example, we can see that the effects on the second indicator of question 30 are measured using a constant and the two latent variables. We have considered 23 indicators, so it is necessary to specify 23 equations. Their relation with the latent variables is depicted in Figure 3 and the results are displayed in Table 6.

Table 6. Latent Variables Structural Model Covariates

	Estimates	**Robust t.student**
EC on Travel Needs with Fewer Vehicles	0.044	1.830
EC on Expanding & Upgrading Roads	-0.189	-6.902
EC on Road Tolls & Gas Taxes	0.170	5.803
ACF on Road Tolls & Gas Taxes	-0.286	-3.837
EC on Bike Lanes & Speed Controls	0.165	7.365
EC on Reducing Car Emissions	0.158	6.398
ACF on Reducing Car Emissions	0.230	4.470
EC on HOV & Transit Priorities	0.171	7.059
EC on Improving Transit Service	0.139	6.628
EC on Promoting Compact Communities	0.055	2.566
EC on Encouraging Short Work Weeks	0.118	6.137
EC on Traffic Congestion	0.238	8.115
ACF on Traffic Congestion	0.069	1.120
EC on Traffic Noise	0.288	8.104
EC on Poor Local Air Quality	0.511	11.095
ACF on Poor Local Air Quality	-0.273	-6.653
EC on Accidents Caused by Bad Drivers	0.251	9.791
EC on Emissions & Global Warming	0.365	11.277
EC on Speeding Drivers in Neighborhoods	0.316	10.398
ACF on Vehicle Type Importance	0.951	13.128
ACF on Fuel Economy Importance	0.198	3.321
ACF on Horsepower Importance	0.991	12.410
ACF on Safety Importance	0.826	15.034
ACF on Seating Capacity Importance	0.852	14.091
ACF on Reliability Importance	0.501	13.779

As explained before, this model measures the effect of the latent variables on each indicator. Some interesting conclusions can be seen from the results presented in Table 6. For example, the effect of environmental concern on the indicator related to the support of expanding and upgrading roads is negative. This sign reflects the idea that car priority in the context of rising road capacity is negatively perceived because of the negative impact on the environment.

On the one hand, we see that the effect of environmental concern on the indicator related to the support of applying road tolls and gas taxes is positive, indicating a perceived positive environmental impact of measures allowing a rational use of the car. Note also that the effect on the same indicator of the appreciation of new car features, the other latent variable considered, is negative. This sign can be explained because of the perceived negative impact of this kind of car use restrictions, especially if the user is considering buying a new car. A similar analysis can be done for the other indicators. For example, the positive sign of the effect of both latent variables on the support for reducing vehicle emissions with regular testing and manufacturer emissions standards: it is perceived with a positive environmental impact but also it is positively perceived by consumers as a good attribute of a potential new car.

It is worth noting that such results permit us to establish a consumer profile in a way not possible with standard discrete choice models. Using simple questions we can enrich the model and obtain better knowledge about the user's characteristics and his behavioral attitudes and perceptions.

Prevision and simulation exercise will come here. We will also compare prediction of Hybrid versus a conventional choice model with many variables in the specification.

6. Conclusions

In the last decade, discrete choice modeling has evolved towards an explicit incorporation of psychological factors affecting decision making. Traditionally, discrete choice models represented the decision process as an obscure black box for which attributes of alternatives and characteristics of individuals are inputs and where the observed choice made by the individual corresponds to the output of the system. The new trend in discrete choice modeling is to enhance the behavioral representation of the choice process. As a direct result, the model specification is improved and the model gains in predictive power. Econometrically, the improved representation – called the Hybrid Choice model – involves dealing with a choice model formulation that contains latent psychometric variables among the set of explanatory variables. Since perceptions and attitudes are now incorporated, this leads to more realistic models.

In this paper we have described the Hybrid Choice model, composed of a group of structural equations describing the latent variables in terms of observable exogenous variables, and a group of measurement relationships linking latent variables to certain observable indicators. We have shown that although the estimation of Hybrid Models requires the evaluation of complex multi-dimensional integrals, simulated maximum likelihood can be successfully implemented and offers an unbiased, consistent and smooth estimator of the true probabilities.

Using real data about virtual personal vehicle choices made by Canadian consumers when faced with technological innovations, we show that Hybrid Choice is genuinely capable of adapting to practical situations. Indeed, the results provide a better description of the profile of consumers and their adoption of new private transportation technologies. We identify two dimensions with a significant impact: Environmental concern and appreciation of new car features. A final positive feature of Hybrid Choice is the fact that enriching the model is quite easy: people are used to rating exercises from marketing surveys, and responses to such questions are simple to include in econometric studies.

References

Ben-Akiva, M.E. and Bolduc, D. (1996). “Multinomial probit with a logit kernel and a general parametric specification of the covariance structure” Working Paper, Departement d’Économique, Université Laval.

Ben-Akiva M. and S.R. Lerman (1985). “Discrete choice analysis: Theory and application to travel demand”, the MIT Press, Cambridge, Massachusetts.

Ben-Akiva, M., D. McFadden, T. Gärling, D. Gopinath, J. Walker, D. Bolduc, A. Börsch-Supan, P. Delquié, O. Larichev, T. Morikawa and V. Rao (1999). “Extended Framework for Modeling Choice Behavior”, *Marketing Letters*, 10(3), pp. 187-203.

Ben-Akiva, M., D. McFadden, K, Train, J. Walker, C. Bhat, M. Bierlaire, D. Bolduc, A. Börsch-Supan, D.

Bhat, C., (2001), Quasi-random maximum simulated likelihood estimation of the mixed multinomial logit model. *Transportation Research B*, **35**, 677-693.

Brownstone, D.S. Bunch, A. Daly, A. de Palma, D. Gopinath, A. Karlstrom and M.A. Munizaga (2002). “Hybrid Choice Models: Progress and Challenges”. *Marketing Letters,* 13(3), pp. 163-175.

Brownstone, D. and Train, K. E., 1999, Forecasting new product penetration with flexible substitution patterns. *Journal of Econometrics*, **89**, 109-129.

Bolduc, D. and A. Giroux (2005). “The Integrated Choice and Latent Variable (ICLV) Model : Handout to accompany the estimation software”, Département d’économique, Université Laval.

Bunch, D. A.. (1991). “Estimability in the multinomial Probit model”. *Transportation Research* B, 25, pp. 1-12.

Domencich, T. and McFadden, D., 1975, *Urban Travel Demand: A Behavioural Analysis.* North-Holland, Amsterdam.

Horne, M. (2003) “Incorporating Preferences for Personal Urban Transportation Technologies into a Hybrid Energy-Economy Model”. PhD Thesis. Simon Fraser University

Jöreskog, K. and D. Sörbom (1984). “LISREL VI: Analysis of linear structural relations by maximum likelihood, instrumental variables and least squares methods”, User’s guide, Department of Statistics, University of Uppsala, Uppsala, Sweden.

McFadden, D. (1974) “Conditional logit analysis of qualitative choice behavior” En P. Zarembka (Ed.), *Frontier in Econometrics*. Academic Press, New York.

McFadden, D. and Train, K., 2000, Mixed MNL models of discrete response. *Journal of Applied Econometrics,* **15**, 447-470.

Munizaga, M.A. and R. Alvarez-Daziano, R. (2005). “Testing Mixed Logit and Probit models by simulation”. *Transportation Research Record* 1921, pp. 53-62.

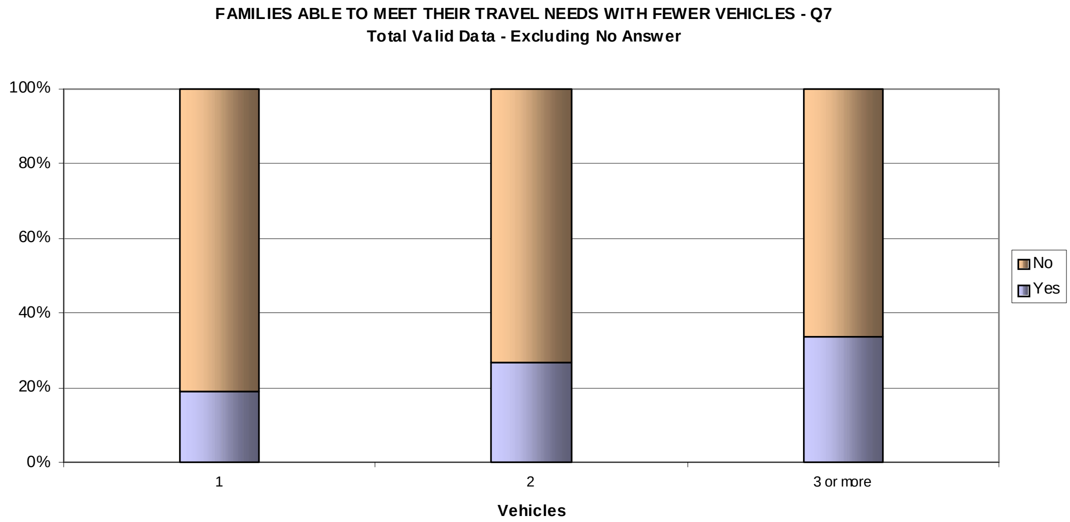
Train, K. (2003) “Discrete Choice Methods with Simulation”. Cambridge University Press.

Walker, J. (2002) “The Mixed Logit (or Logit Kernel) model: dispelling misconceptions of identification”. *Transportation Research Record* 1805, pp. 86-98.

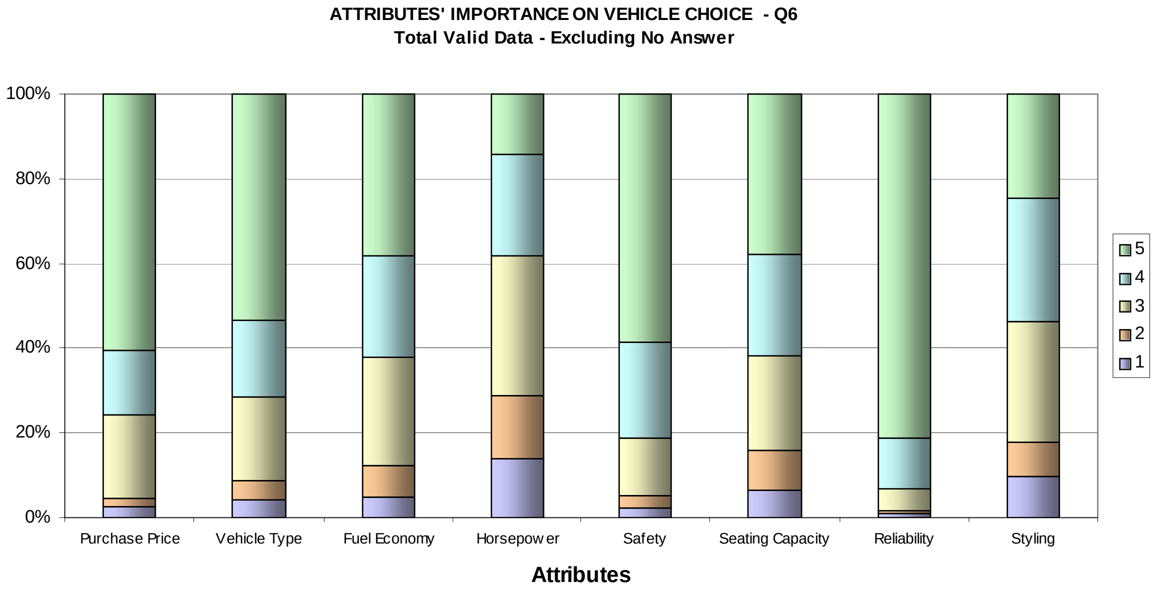
Walker, J.W., M. Ben-Akiva, M. and D. Bolduc (2005). “Specification & Estimation of the Logit Kernel (or Mixed Logit) Model”, *Journal of Applied econometrics*.

Williams, H.C.W.L. (1977) “On the formation of travel demand models and economic evaluation measures of user benefit” *Environment and Planning*, Vol**9A**, pp. 285-344.

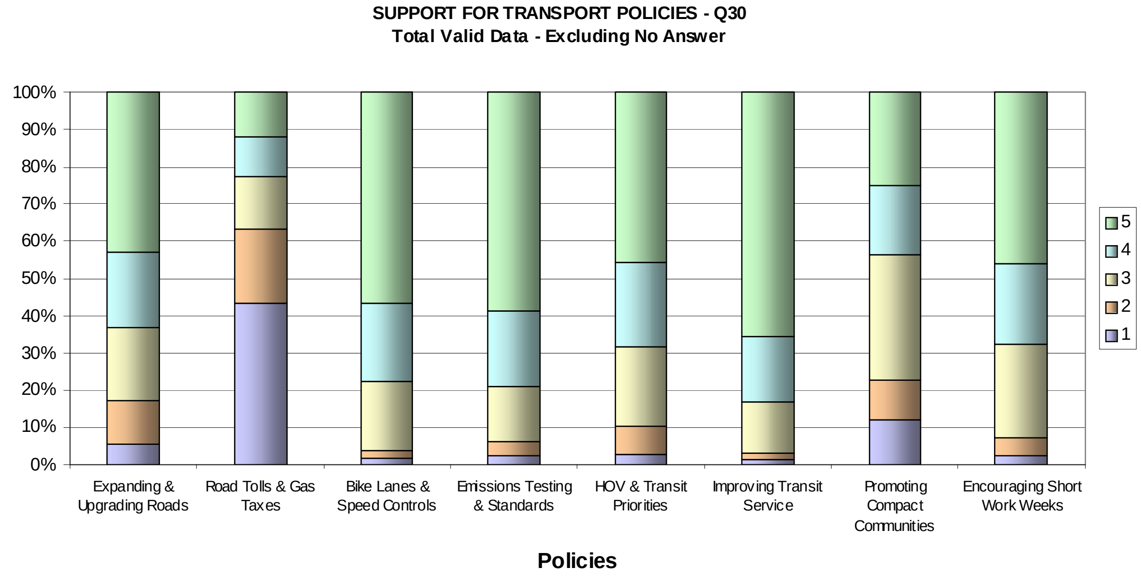
Appendix

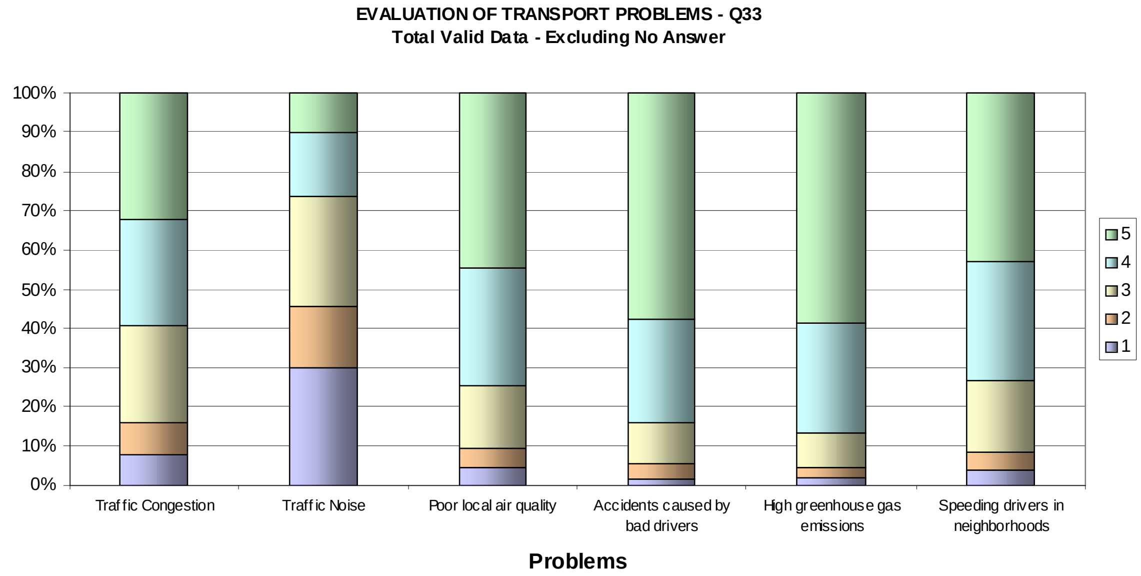


FAMILIES ABLE TO MEET THEIR TRAVEL NEEDS WITH FEWER VEHICLES - Q7
Total Valid Data - Excluding No Answer
100%
80%
60%
40%
20%
0%
1
2
3 or more
Vehicles
No
Yes



ATTRIBUTES' IMPORTANCE ON VEHICLE CHOICE - Q6
Total Valid Data - Excluding No Answer
100%
80%
60%
40%
20%
0%
Purchase Price
Vehicle Type
Fuel Economy
Horsepower
Safety
Seating Capacity
Reliability
Styling
Attributes
5
4
3
2
1


SUPPORT FOR TRANSPORT POLICIES - Q30
Total Valid Data - Excluding No Answer
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Expanding & Upgrading Roads
Road Tolls & Gas Taxes
Bike Lanes & Speed Controls
Emissions Testing & Standards
HOV & Transit Priorities
Improving Transit Service
Promoting Compact Communities
Encouraging Short Work Weeks
Policies
5
4
3
2
1


EVALUATION OF TRANSPORT PROBLEMS - Q33
Total Valid Data - Excluding No Answer
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Traffic Congestion
Traffic Noise
Poor local air quality
Accidents caused by bad drivers
High greenhouse gas emissions
Speeding drivers in neighborhoods
Problems
5
4
3
2
1