Comparison of Vehicle-Ownership Models

Dimitris Potoglou, Ph.D.
(Corresponding Author)
Delft University of Technology
OTB Research Institute for Housing, Urban and Mobility Studies
P.O. Box 5030
2600 GA Delft
The Netherlands
Tel. +31 15 2785833
Fax. +31 15 2784422
E-mail: d.potoglou@tudelft.nl

Yusak O. Susilo, Ph.D.
Delft University of Technology
OTB Research Institute for Housing, Urban and Mobility Studies
P.O. Box 5030
2600 GA Delft
The Netherlands
Tel. +31 15 2782714
E-mail: y.susilo@tudelft.nl

Word Count: Text (5,586) + 5 Tables (1,250) = 6,836

Paper submitted for presentation and publication at the 87th Annual Meeting of the Transportation Research Board, Washington, D.C., January 2008
July 2007

Abstract

Empirical studies on household car ownership have used two types of discrete choice modelling structures, the ordered and the unordered. In ordered structures such as the ordered logit and ordered probit models, the choice of the number of household-vehicles arises from a uni-dimensional latent index that reflects the propensity of a household to own vehicles. Unordered response models, on the other hand, are based on the random utility maximization principle, under which a household associates a utility value across different car ownership levels and chooses the one with the maximum utility. The multinomial logit and probit models are representatives of the unordered response models, but only the multinomial logit has been used extensively because of its simple structure and low computational requirements. Following Bhat and Pulugurta's (1998) comparative study between the multinomial logit and ordered logit models, consensus among researchers is still lacking and empirical studies have reported car ownership model based on the multinomial logit, ordered logit as well as ordered probit models. It is apparent that there is still an open question to be addressed: Which of the aforementioned models would reflect better households' car ownership choices? This paper provides an empirical comparison of multinomial logit, ordered logit and ordered probit car ownership models by introducing a number of formal evaluation measures and using three datasets; the 2001 National Household Travel Survey for the metropolitan area of Baltimore, the 2005 Dutch National Travel Survey and the 2000 Osaka Metropolitan Person Trip Data. Results show that the multinomial logit model is the one to be selected for modelling the level of household car ownership over ordered logit and ordered probit.

Keywords: car ownership, discrete choice models, automobile demand, multinomial logit, ordered logit, ordered probit

1. INTRODUCTION

Car ownership is a key feature of modern life that to a great extent influences travel behaviour and participation in out-of-home activities *(1)*. Increasing utilization patterns and high-levels of household car ownership have a direct effect on energy consumption and air quality levels at local and global scales *(2)*. It is therefore of interest to transport planners and policy makers at all levels of governance to use models that are capable of explaining the causal factors of car ownership, being sensitive to policy measures and forecast market shares of automobiles. Such models can be used as stand-alone systems or as part of integrated land-use and transportation models for simulating interactions between land-use and transportation *(3)*.

Models of car ownership analysis can be distinguished into two broad categories: disaggregate and aggregate. Disaggregate models focus on the household level, whereas aggregate models consider car ownership as an accumulation of household decisions at different geographical scales such the traffic analysis zone, region, state or country levels. Disaggregated models have dominated over aggregate models mainly because of their behavioural structure and improved capability in identifying causal relationships *(4)*. Moreover, disaggregate models have overcome deficiencies and limitations of aggregate models such as mutli-collinearity across explanatory variables, large standard errors of estimated parameters and aggregation bias. Furthermore, the extensive use of disaggregate models lies in their capability to conduct policy-sensitive analyses and compatibility with contemporary agent-based approaches in transportation modelling *(5)*.

Disaggregate models regard car ownership observations (i.e., 0, 1, 2, 3 or more cars) either as ordinal or nominal discrete variables, thus giving rise to two types of choice models, the ordered and unordered, respectively. Ordered response models assume that the choice of the number of household-vehicles arises from a uni-dimensional latent index reflecting the propensity of a household to own vehicles. On the other hand, unordered response models follow the random utility maximization axiom, in which a household associates a utility value across different ownership levels and chooses the one with the maximum utility.

A review of the recent literature on disaggregate car ownership models reveals that it is not yet clear whether ordered response or unordered response models are the most appropriate *(*see for example, *6)*. While Bhat and Pulugurta *(7)* argued that a multinomial logit model (MNL) model would be more appropriate for modelling car ownership over an ordered logit (ORL) model, subsequent empirical studies have developed ordered probit (ORP) or ORL models claiming the discrete, ordered nature of the dependent variable *(8,9)* as well as MNL models on the basis of their sound behavioural and theoretical nature *(10-12)*. In particular, comparing modelling results of MNL and ORL using several datasets, Bhat and Pulugurta *(7)* found substantial differences in the elasticities of exogenous variables across the choice probabilities of car ownership levels and identified misspecification problems associated with the ORL that could lead to incorrect and inaccurate forecasts. Also, Potoglou and Kanaroglou *(12)* found that a MNL model was a significantly improved model over the ORL through a likelihood-ratio test between the two models using data from the metropolitan area of Hamilton, Canada.

This paper offers a comprehensive comparison of car ownership models including MNL, ORL and ORP. Two research questions are addressed in the paper. First, what measures could be used in order to conduct a comparison between different discrete choice model structures? And second, which is the most appropriate among the MNL, ORL and ORP for modelling car ownership? We address these questions by developing a set of the aforementioned car ownership models using three sources of data; 2001 U.S. National Household Travel Survey (NHTS) of the Baltimore Metropolitan Area *(13)*, the 2005 Dutch National Travel Survey *(14)* and the 2000 Osaka Metropolitan Person Trip Data *(15)*. Following, we compute a number of comparison measures introduced in the econometrics literature.

The reminder of the paper is organized as follows. First, we provide a brief description on the structure and major assumptions of ordered and unordered models. Next, we introduce a set of evaluation measures to be used in the comparison of models. Following that, we present the data sets used and describe the construction of explanatory variables. Next, we discuss estimation results of each model and report on the findings of the comparative analysis as reflected by the evaluation measures. In the last section, we offer some concluding remarks.

2. ORDERED AND UNORDERED MODELS: THEORETICAL BACKGROUND

Ordered and unordered models require different techniques for their respective analysis *(16)*. We briefly present the main characteristics and assumptions of ordered and unordered models in the following subsections.

2.1 Ordered Response Models: ORL and ORP

Ordered response models assume that the observed number of household cars (dependent variable) is a discrete, ordinal variable that is mutually exclusive and collectively exhaustive. Hence, the observed number of cars per household is assumed to be *inherently ordered* implying that higher number of cars is ranked higher than the outcome associated with less number of cars. Specifically, the ordered response models assume that the number of vehicles of household n - denoted as (Y_n) - arises from a dimensional latent index y^*_n, as follows :

$$y^*_n = \sum_{n=1}^{K} \beta_k X_{nk} + \varepsilon_n \qquad [1]$$

where, y^*_n reflects the propensity of a household to own vehicles, β_k (to be estimated) is a set of parameters associated with a set of k explanatory variables (X_{nk}). The distribution of the error terms ε_n marks the difference between the ordered logit and the ordered probit models.

In general, the observed level of automobile ownership of a household (Y_n) equal to i (= 0, 1,…M) number of cars is assumed to be related to the latent auto ownership index as follows *(17)*:

$$Y_n = i \text{ if and only if } \mu_{i-1} < y^*_n < \mu_i,\ i = 0, 1,..., M \quad \mu_{-1} = -\infty, \mu_i = +\infty, \mu_0 = 0, \mu_{i-1} < \mu_i, \text{ for all i} \qquad [2]$$

where μ are cut-offs points for discriminating between successful automobile ownership levels on the underlying latent scale. Cut-off points are estimated along with the β coefficients in one-step procedure *(*see, *18)*. Hence, the probability of household n to own i number of cars is as follows:

$$P_n[Y = i] = P_{ni} = \Phi(\mu_i - \beta x_n) - \Phi(\mu_{i-1} - \beta x_n) \qquad [3]$$

where Φ(.) is the standard cumulative normal distribution, which gives rise to the ordered probit model. Alternatively, Φ(.) can be replaced with the Gumbel distribution, Λ(.), resulting into the ordered logit model. With regard to the selection of the cumulative distribution function and hence, the choice between ordered logit and probit models, Greene *(18)* pointed out that "it is difficult to justify the choice of one distribution over the other on theoretical grounds…in most applications, it seems not to make much difference".

A significant assumption in the estimation of ordered response models (ordered probit and logit) is that of *parallel slopes*. This assumption implies that the estimated coefficient of an explanatory variable affecting the probability of a household to own a number of cars would be equal for all outcomes (i.e., i = 1, 2, 3 or more cars) *(16)*. If the parallel slopes assumption is invalid and coefficients associated with a particular variable are different across different levels of car ownership, then the ordered response mechanism is no longer appropriate. In that case, the model should be estimated using an unordered response model. As Borooah *(16)* suggested "the validity of the parallel slopes assumption can be tested by estimating a multinomial logit model on the data…while the ORL model estimates K coefficients, the MNL model estimates K(M-1) parameters…if the LL_1 is the likelihood function value from the ORL model and LL_2 is the likelihood value from the MNL model, then one can compute 2*(LL_2-LL_1) and compare with χ^2(K(M-2))". This test is only suggestive implying that a "very large" χ^2 value would provide grounds for concern, a moderately "large" value would not *(16)*.

2.2 Unordered Response Models: MNL and MNP

In the case of the unordered response models, the information conveyed by the ordinal nature of the observed car household ownership is discarded. Hence, the probability of a household to own a given number of automobiles is based on the utility maximization principle, in which a household n associates a *utility* value to each car ownership level and chooses the one with the maximum utility. Assuming that a household is perfectly informed and the decision is completely rational and consistent, the random utility maximization framework implies that:

$$U_{i,n} = \beta_{i,n} X_n + \varepsilon_{i,n} \quad [4]$$

where $U_{i,n}$ represents the true utility of a household n owning i number of cars, $\beta_{i,n}$ is a vector of parameters to be estimated and $\varepsilon_{i,n}$ is the random component that captures the unobserved utility or uncertainty from the point of view of the observer/analyst. Again, if the error terms $\varepsilon_{i,n}$ in Equation 4 are identically and independently distributed (IID) with a Type I Extreme Value (or Gumbel) distribution, the probability of owning a number of cars i takes the form the MNL expressed as follows *(19)*:

$$P_n(i) = \frac{\exp(\beta_{i,n} x_n)}{\sum_{j \in C_n} \exp(\beta_{j,n} x_n)} \quad [5]$$

The computational simplicity of the MNL model has permitted its wide use in empirical studies of car ownership. Nevertheless, the MNL has an undesirable property known as the "independence of irrelevant alternatives" (IIA). The IIA property implies that the ratio of the probabilities of two alternatives is independent from any other available alternatives. Should this assumption is violated then the MNL is unsuitable for use in any application in which the random components of utility are correlated across alternatives and observations of choices *(20)*.

Alternatively, the assumption of error terms $\varepsilon_{i,n}$ being multivariate normal distributed with mean zero and covariance matrix Ω_n leads to the Multinomial Probit (MNP) model. As opposed to the MNL, the MNP model is more flexible because it relaxes the IIA assumption of IID errors and thus, allows for correlations of the error terms of different alternatives *(20)*. In the MNP model, the probability of a household choosing a certain number of cars i is given as follows:

$$P_n(i) = \int_{\varepsilon_{in}=-\infty}^{\infty} \int_{\varepsilon_{1n}=-\infty}^{\varepsilon_{in}+V_{in}-V_{1n}} \int_{\varepsilon_{2n}=-\infty}^{\varepsilon_{in}+V_{in}-V_{2n}} \cdots \int_{\varepsilon_{mn}=-\infty}^{\varepsilon_{in}+V_{in}-V_{mn}} \phi(\tilde{\varepsilon}_{i,n}) d\varepsilon_{m_n n} \ldots d\varepsilon_{2n} d\varepsilon_{1n} d\varepsilon_{in} \quad [6]$$

where $\tilde{\varepsilon}_{i,n}$ is a vector of error differences $\tilde{\varepsilon}_{i,n} = \langle \tilde{\varepsilon}_{n1i}, \ldots, \tilde{\varepsilon}_{nJi} \rangle$ over all alternatives except i, with density function $\phi(\tilde{\varepsilon}_{i,n})$.

As shown in Equation 6, the required integrations are computationally burdensome, especially when the number of alternatives increases. As a result, the use of the MNP model has been limited in empirical studies involving discrete choice models. Several simulation methods (see, *21)* have been introduced to overcome this problem, however, there remain considerations with regard to the applicability of the MNP model. Weeks *(22)* concluded that the estimation of the MNP entails several estimation, specification and identification issues. Also, Horowitz *(23)* argued that the MNP model involves problems such as the proliferation of random effects and parameters as well as specification testing. Proliferation can complicate forecasting as well as adds to the complexity of estimation because it requires a higher number of covariance matrix elements to be estimated. Regarding specification testing, Horowitz *(23)* claimed that "… a misspecified model can create non-IID random taste variation or additive random components in the error terms of the model". In such case, it may appear that a MNP model with random taste variation or non-IID additive random utility components fits the data set better that a MNL model, when the real problem is that the systematic component of the logit utility function is misspecified. As a result, the use of MNP for car ownership models may present several difficulties, while also estimation results may not be directly comparable with those of a MNL.

TABLE 1 Main characteristics of common ordered and unordered models

	Ordered Models		**Unordered Models**	
	ORL	**ORP**	**MNL**	**MNP**
Behavioural Context	No	No	Yes	Yes
Distribution of Error Terms	Gumbel	Normal	IID - Gumbel	Normal
Parallel Slopes Assumption	Yes	Yes	No	No
IIA Property	No	No	Yes	No
Computational Requirements	Low	Low	Low	High

3. EVALUATION MEASURES OF DISCRETE CHOICE MODELS

The likelihood ratio ρ^2 and the adjusted likelihood ratio indices $\bar{\rho}^2$ are the most common measures of goodness-of-fit used with discrete choice models and are given as follows *(20)*:

$$\rho^2 = 1 - \frac{LL(\hat{\beta})}{LL(C)} \tag{7}$$

$$\bar{\rho}^2 = 1 - \frac{LL(\hat{\beta}) - K}{LL(C)} \tag{8}$$

where $LL(\hat{\beta})$ and $LL(C)$ are the log-likelihood function values at convergence and sample shares, respectively. K is the number of parameters estimated in the model, excluding any constants as well as cut-off parameter values in the case of ordered models.

The difference [$LL(\hat{\beta}) - K$] is also known as the *Akaike Information Criterion* (AIC), which is another measure of goodness-of-fit of a model given as *(24)*:

$$AIC = -2LL(\hat{\beta}) + 2K \tag{9}$$

To correct for small samples Hurvich and Tsai *(25)* proposed the AIC_c criterion, which can be used regardless of sample size:

$$AIC_c = -2\,LL(\hat{\beta}) + 2K + \frac{2K(K+1)}{(N-K-1)} \tag{10}$$

where N is number of observations. AIC and AIC_c allow for a better comparison across models because they account for goodness-of-fit as well as include a penalty that is an increasing function of the number of parameters.

Another measure for model comparison is the *Bayesian Information Criterion* (BIC), also known as *Schwarz Criterion* (SIC) *(26)*:

$$BIC = -2\ LL(\hat{\beta}) + K \cdot \ln(N) \qquad [11]$$

The BIC can be used to compare models that are non-nested as is the case with MNL, ORL and ORP models. Given any estimated models derived from the same sample, the model with the lower value of AIC_c or BIC is the one to be preferred because this model is estimated to be the "closest" to the unknown true model. Compared with AIC, the BIC penalizes free parameters more strongly than the AIC.

Another consistent criterion as opposed to the non-consistent AIC_c and BIC is the *Hannan and Quinn Information Criterion* (HQIC) criterion *(27)*:

$$HQIC = -2\,LL(\hat{\beta}) + 2K \cdot \ln(\ln(N)) \qquad [12]$$

AIC_c, BIC and HQIC indices take into account model parsimony. That is, other things equal, given two models with equal log-likelihood values, the model with fewer parameters is better.

Finally, a comparison between two non-nested models may be performed with Ben-Akiva and Lerman's *(20)* adjusted log-likelihood ratio test, which determines if the adjusted likelihood ratio indices between two non-nested models are significantly different. Under the null hypothesis that model 1 is a better representation than model 2, the following holds asymptotically:

$$\Pr(\bar{\rho}_2^2 - \bar{\rho}_1^2 > z) \leq \Phi\left\{-\left[-2z\,LL(C) + (K_2 - K_1)\right]^{0.5}\right\},\ z > 0 \qquad [13]$$

where $\Phi(.)$ is the standard normal cumulative distribution function. As Bhat and Pulugurta *(7)* argued a small value of the probability in Equation 13 indicates that the difference z is statistically significant and the model with the higher value of adjusted likelihood ratio index is to be preferred.

Finally, another goodness-of-fit statistic is "the percent correctly predicted", which is calculated by identifying the highest probability for each decision-maker in the sample. The "the percent correctly predicted" is the portion of decision-makers in the sample for which the alternative with the highest probability and the actually chosen alternative are the same. However, as Train *(21)* argued, this contradicts with the notion of probability. A model provides enough information in identifying the probability of choosing each alternative, however, it is incapable of determining repeatedly the actual choice. Hence, this measure will not be considered further in the evaluation of models.

4. SOURCES OF DATA

The data sets used in this paper were obtained from three sources: the 2001 US NHTS *(13)*, the 2005 Dutch NTS *(14)* and the 2000 Osaka Metropolitan Person Trip Data *(15)*. The US NHTS has been conducted by the US Bureau of Transport Statistics (BTS) and the US Federal Highway Administration (FHWA). The objective of this survey is to gain a better understanding of travel behaviour in the US. The data are used by the US Department of Transportation to assess program initiatives, to review programs and policies, to study current mobility issues, and to plan for the future. In this paper, the analyses used data from the

Baltimore Metropolitan Area, which consist of 3,496 households, 8,744 persons, and 32,359 trips.

The Dutch NTS is a cross-sectional travel-diary of households and data collection has been conducted continuously by Statistics Netherlands since 1978. For each year up to 1993, the NTS recorded data for approximately 10,000 households, 20,000 individuals (and more than 80,000 journeys). During 1994 and 1995 the NTS was extended to include substantially more respondents and households each year and also to include children younger than 12, who were previously excluded from the survey *(14)*.

Finally, the Osaka metropolitan area person-trip survey is a conventional large-scale household travel survey conducted in the Osaka metropolitan area of Japan in 1980, 1990 and 2000, with sampling rates of 2.4% to 3.0%. The survey collected travel patterns as well as home and work locations of the respondents on the observed day. It also collected the socio-demographics and household characteristics *(15)*.

5. ESTIMATION RESULTS AND EVALUATION OF MODELS

Car ownership models were estimated considering four car ownership levels: zero, one, two and three or more cars. All estimations were performed using the program NLOGIT Ver. 3.0 *(28)* and final models are reported in Tables 2, 3 and 4. There are some key features to be addressed at this point. First, parameters of the MNL models were estimated for three alternatives (i.e., one, two and three or more cars), whereas the zero-cars alternative was considered as base alternative for identification purposes. Second, the parameters of the MNL models indicate propensity to own one, two or three or more vehicles in such way that positive values increase the probability and negative values decrease the probability of owning a particular number of cars. Similarly, the coefficients in the ordered response models indicate propensity to own more cars and thus, positive values of coefficients indicate propensity of households to own higher number of cars. Finally, it is worth noting that MNP models of car ownership did not reach convergence with any of the aforementioned datasets, and therefore MNP model estimated are not reported in this comparison.

With regard to the Baltimore dataset, the *single-family house* attribute was used as proxy for parking-space availability at the place of residence. The coefficient of this variable was significant in all models, except in the utility of owning one car. Thus, in the case of MNL model, households living in single family dwellings were more likely to own two and three-or-more cars. In the ORL and ORP models, the positive sign of the *Single Family House* attribute implied the increased propensity of households to own cars. This finding is in line with findings reported in Potoglou and Kanaroglou *(12)*, Bhat and Pulugurta *(7)* and Chu *(8)*. Furthermore, socio-economic variables such as number of workers, total household income and race of the head of the household had positive coefficients implying an increased probability of owning vehicles in the MNL as well as an increased likelihood of owning more vehicles in the case of ORL and ORP models, respectively. Furthermore, as shown in Table 3, parameter signs of the number of workers and income variables were consistent with estimation results using the 2005 Dutch NTS.

The *household life cycle* reflected the influence of the composition of the household on the number of cars owned. In all datasets, we used as reference class the household composition "single" (households with one adult only). As shown in Table 2, the MNL model captures non-linear effects between *household life cycle* and the probabilities of car ownership. Specifically, a household comprised of two adult members ("couple") is less likely to own one car and more likely to own two or three-or-more cars. It is demonstrated in this case that unordered models such as the MNL placed no restrictions on the effect of an explanatory variable across car ownership levels. This finding agrees with the theoretical claim of Bhat and Pulugurta *(7)*.

Also in Table 2, the pattern of switching parameter signs of the variable "couple" across car ownership levels cannot be captured in ORL and ORP models where the parameter value is restricted to a unique (positive) parameter. Similarly, the "couple with children" variable had no effect in the probability of owning one car, however, it did affect positively the probability of owning two-or-more vehicles. Furthermore, the s*ingle parent household* structure had no effect at all levels of car ownership; that was also the case of "retired" households and the probability of owning one car. For all other types of household structure, coefficients were statistically significant and their signs agreed across MNL, ORL and ORP models as well as with *a priori* expectations.

TABLE 2 Parameter estimates of the Baltimore Dataset

		MNL		**ORL**	**ORP**
Variable	**1**	**2**	**3 +**		
Constant	0.504 (1.94)	-2.695 (-8.00)	-6.746 (-12.52)	-0.374 (-2.32)	-0.191 (-2.09)
Type of Dwelling: Single Family House	0.275 (1.49)	1.118 (5.53)	1.778 (7.38)	0.763 (8.36)	0.452 (8.77)
Number of Workers	0.509 (4.13)	1.212 (8.65)	2.175 (13.77)	1.035 (17.15)	0.574 (17.15)
Household Income (>$30,000)			REFERENCE		
Household Income ($30,000 - $80,000)	1.683 (10.80)	2.401 (13.18)	3.178 (11.16)	1.413 (15.44)	0.806 (15.82)
Household Income (> $80,000)	2.013 (4.91)	3.708 (8.80)	4.659 (9.82)	1.906 (16.61)	1.100 (16.91)
Household Life Cycle: Single			REFERENCE		
Household Life Cycle: Couple	-0.573 (-2.46)	1.872 (6.91)	1.950 (4.73)	1.243 (10.41)	0.669 (9.84)
Household Life Cycle: Single Parent	-0.333 (-1.49)	-0.070 (-0.21)	-0.254 (-0.42)	-0.080 (-0.48)	- 0.053 (-0.55)
Household Life Cycle: Couple w. Children	-0.460 (-1.81)	2.082 (7.14)	2.350 (5.51)	1.383 (10.97)	0.751 (10.49)
Household Life Cycle: Retired	0.011 (0.06)	1.864 (8.07)	2.620 (6.73)	1.310 (11.83)	0.733 (11.60)
Race: Caucasian	1.066 (7.95)	1.636 (9.45)	2.007 (8.45)	0.939 (10.53)	0.531 (10.56)
Residential Density	-0.00025 (-6.62)	-0.00046 (-10.85)	-0.00063 (-11.63)	-0.00025 (-12.82)	-0.00015 (-13.02)
Threshold Parameters μ (for identification μ_0=0)					
μ_1				3.087 (52.76)	1.726 (53.49)
μ_2				6.263 (74.16)	3.520 (79.60)
Log-likelihood at convergence			-2949	- 3081	-3092

TABLE 3 Parameter estimates using the 2005 Dutch NTS Dataset

	MNL			ORL	ORP
Variable	**1**	**2**	**3 +**		
Constant	-1.051	-5.649	-9.090	-1.065	-0.590
	(-20.24)	(-47.67)	(-23.50)	(-24.36)	(-23.20)
Number of Workers	0.964	1.559	2.120	0.888	0.480
	(22.09)	(31.19)	(27.62)	(38.74)	(37.52)
Household Income (>€20,000)			REFERENCE		
Household Income (€20,000 - €60,000)	0.800	1.593	1.312	0.885	0.500
	(21.72)	(22.88)	(5.30)	(27.38)	(27.47)
Household Income (>€60,000)	1.693	3.269	3.475	1.791	1.020
	(15.76)	(26.61)	(12.94)	(39.26)	(39.48)
Household Life Cycle: Single			REFERENCE		
Household Life Cycle: Couple	1.073	2.757	1.534	1.256	0.700
	(20.83)	(28.14)	(5.27)	(31.29)	(31.06)
Household Life Cycle: Couple w. Children	1.425	3.596	1.904	1.722	0.951
	(18.04)	(31.64)	(6.23)	(39.12)	(38.64)
Household Life Cycle: Extended family	0.687	2.779	3.298	1.856	1.030
	(8.95)	(24.29)	(11.60)	(36.93)	(37.17)
Household Life Cycle: Retired	0.137	0.577	0.416	0.159	0.09
	(2.88)	(4.71)	(0.97)	(3.78)	(3.52)
Reside in very highly urbanised area			REFERENCE		
Reside in high urbanised area	0.631	1.026	1.550	0.616	0.350
	(12.36)	(12.95)	(6.35)	(14.79)	(14.91)
Reside in moderately urbanised area	0.904	1.557	2.199	0.911	0.520
	(16.25)	(18.85)	(9.01)	(21.01)	(21.23)
Reside in low urbanised area	1.195	1.987	3.049	1.168	0.670
	(20.63)	(23.73)	(12.81)	(26.933)	(27.22)
Reside in non-urbanised area	1.306	2.258	3.209	1.301	0.740
	(20.29)	(25.09)	(13.07)	(28.17)	(28.36)
Threshold Parameters μ (for identification μ_0=0)					
μ_1				3.870	2.19
				(153.13)	(169.49)
μ_2				7.145	3.95
				(162.78)	(185.95)
Log-likelihood at convergence			-22941	-23255	-23254

TABLE 4 Parameter estimates using the 2000 Osaka Metropolitan Area Dataset

Variable	MNL 1	MNL 2	MNL 3 +	ORL	ORP
Constant	-0.397	-1.541	-2.318	0.520	0.325
	(-14.53)	(-35.89)	(-35.08)	(21.77)	(23.93)
Household Life Cycle: Single	REFERENCE				
Household Life Cycle: Couple	0.912	0.252	-1.235	-0.356	-0.211
	(29.59)	(6.77)	(-18.74)	(-18.59)	(-18.30)
Household Life Cycle: Single parent	-0.871	-2.843	-4.388	-1.953	-1.154
	(-19.10)	(-25.00)	(-16.21)	(-45.65)	(-45.02)
Household Life Cycle: Couple w. Children	0.687	0.011	-1.373	-0.465	-0.272
	(34.88)	(0.46)	(-34.62)	(-35.52)	(-35.24)
Household Life Cycle: Extended family	1.164	1.092	0.835	0.299	0.190
	(81.37)	(69.87)	(48.35)	(36.01)	(39.82)
Household Life Cycle: Retired	-1.061	-2.195	-3.305	-1.898	-1.088
	(-85.54)	(-118.0)	(-105.86)	(-184.89)	(-186.24)
Reside in CBD area	REFERENCE				
Reside in mixed area	0.564	0.577	0.586	0.484	0.267
	(20.30)	(13.09)	(8.58)	(19.85)	(19.27)
Reside in autonomous area	1.458	3.212	4.200	2.631	1.499
	(41.19)	(66.16)	(59.87)	(101.21)	(100.93)
Reside in suburban area	1.270	2.011	2.362	1.452	0.824
	(46.46)	(46.72)	(35.61)	(60.97)	(60.95)
Reside in non-urbanised area	1.792	3.398	4.624	2.804	1.609
	(22.75)	(39.19)	(45.99)	(69.53)	(68.43)
Threshold Parameters μ (for identification μ_0=0)					
μ_1				2.375	1.400
				(531.86)	(555.86)
μ_2				3.877	2.265
				(659.11)	(710.88)
Log-likelihood at convergence			-350228	-355634	-355820

Finally, we specified a full set of alternative specific constants in the MNL models corresponding to one, two and three-or-more-cars options in order to capture the systematic influence of omitted variables, namely car ownership costs, in the utility functions of the model *(8)*.

5.1 Behavioural Comparisons across the American, Dutch and Japanese Datasets

Comparing the estimated coefficients across the Baltimore, Dutch and Japanese datasets, it is clear that estimates present some discrepancies, which may be explained by the contextual/cultural differences in the sampled population and also, by the different explanatory variables available in the dataset. For example, while in the Baltimore dataset (Table 2) we used residential density to measure the influence of land-use on car ownership, we used the degree of urbanisation in the Dutch (Table 3) dataset and urban-functions in the dataset of Osaka metropolitan area (Table 4). Moreover, unlike the other two datasets, the Osaka metropolitan area dataset did not include household income and information on the number of workers, which is a crucial variable for car ownership decisions.

Nevertheless, comparison of the datasets reveals interesting differences. For example, except for the extended family case, the Dutch and Osaka metropolitan respondents tend to have two cars rather than three or more cars, which is different than Baltimore respondents. Presumably this is because of the different needs for car usage. Dutch people tend to walk and to bike in their daily travel *(29)*, while the Osaka metropolitan area has a very dense and well developed public transport network *(15)* compared to Baltimore area. Yet, despite the differences in describing the land-use influences to the household car ownership, results in all datasets are in line with the hypothesis that higher densities or proximity to the activity locations reduces the probability of a household owning a car.

Furthermore, examining the influence of household structure to household car ownership, we found that - all else being equal - while Baltimore couples were less likely to own one car, Dutch and Osaka's couples were more likely to own one and three cars than zero cars. Also, single-parent households residing in Baltimore and Osaka had a higher probability of owning one car and lower probability of owning three cars. On the other hand, Dutch single parents were more likely to own one car and less likely to posses three cars. As shown in Tables 2, 3 and 4, couples with children also present differences across the dataset estimates. In particular, couples with children living in Baltimore were less likely to own one car and more likely to own three cars, as opposed to Dutch and Osaka's households that would be more likely to own one car. Interestingly, Baltimore's retirees had a higher probability to own cars as opposed to Osaka's where the negative coefficients imply that the probability of owning any number of cars would be less than owning no cars. Again, these discrepancies are because of differences in public transport availability and high usage of non-motorized modes across Baltimore, the Netherlands and Osaka. It is worth mentioning here that these discrepancies are also reflected in the estimates of the ORL and ORP models.

5.2 Comparison of MNL, ORL and ORP Models

All models performed well as indicated by the relatively high values of rho-squared. Also, likelihood ratio tests were used to test the null hypothesis that all parameters in each model - except the alternative specific constants - were zero. As shown in Table 5, chi-square values of the likelihood ratio indices in all models reject the null hypothesis and indicate that all models were statistically significant. The adjusted likelihood values lend support to the MNL model as the best for modelling car ownership because it exhibits the highest value of the index. Also, the values of AIC_c, BIC and HQIC indices clearly demonstrate the superiority of the MNL against the ORL and ORP. Finally, Ben-Akiva and Lerman's *(20)* adjusted log-likelihood ratio test was used for testing the non-nested hypothesis that the adjusted log-likelihood values of the MNL, ORL and ORP were statistically different. The last three rows of Table 5 present the upper bound of the probability of erroneous choosing the incorrect model, which however, had the highest adjusted likelihood index. In other words, this hypothesis states that the estimated differences in the adjusted likelihood ratio index values between the MNL, ORL and ORP models could have occurred by chance. The estimated probability values of the aforementioned hypothesis showed that the adjusted likelihood ratio index of the MNL is significantly different than those of the ORL and ORP models and therefore, the MNL is the preferred model. Similarly, the ORL model should be preferred as opposed to the ORP model in the Baltimore case, whereas there is no difference between the ORL and ORP models in the Dutch and Japanese datasets.

TABLE 5 Evaluation Summary of MNL, ORL, ORP Models

	2001 Baltimore Dataset			2005 Dutch NTS			2000 Osaka Metropolitan Area Dataset		
	MNL	ORL	ORP	MNL	ORL	ORP	MNL	ORL	ORP
Sample Size	3496	3496	3496	28436	28436	28,436	312632	312632	312632
Log-likelihood Function: No Coefficients, LL(0)	-4846.48			-39420.67			-433399.98		
at Constants only, LL(C)	-4576.60	-4576.60	-4576.60	-30127.51	-30127.51	-30127.51	-400341.90	-400341.90	-400341.90
at Convergence, $LL(\hat{\beta})$	-2949.082	-3080.880	-3092.183	-22941.60	-23254.61	-23254.3	-350228.36	-355634.10	-355820.20
Likelihood Ratio Index: $-2\,(LL(C) - LL(\hat{\beta}))\ (d.f.)$	3255 (30)	2991 (10)	2969 (10)	14372 (33)	13746 (12)	13746 (12)	100227 (27)	89416 (9)	89043 (9)
Rho-Squared (ρ^2)	0.356	0.327	0.324	0.239	0.228	0.228	0.125	0.112	0.111
Adjusted Rho-Squared ($\bar{\rho}^2$)	0.349	0.325	0.322	0.237	0.228	0.228	0.125	0.112	0.111
AIC_c	5958	6181	6204	45949	46533	46,531	700511	711286	711658
BIC	6142	6243	6265	46222	46632	46,621	700798	711382	711754
HQIC	6024	6203	6226	46037	46565	46,560	700594	711314	711686
Parallel Slopes Assumption $\chi^2\,(d.f.) = 2 * (LL_{ORL\ or\ ORP} - LL_{MNL}),\ p$		$\chi^2\,(20) = 263.5$, $p < 0.000$	$\chi^2\,(20) = 286.2$, $p < 0.000$		$\chi^2\,(12) = 625.4$, $p<0.000$	$\chi^2\,(12) = 625.4$, $p<0.000$		$\chi^2\,(9) = 10227.1$, $p<0.000$	$\chi^2\,(9) = 11183.68$, $p<0.000$
$\Phi(\bar{\rho}^2_{ORL} - \bar{\rho}^2_{ORP} > z)$	$\Phi(-5.24) = 8.029 \times 10^{-8}$			$\Phi(0) = 0.5$			$\Phi(-26.46) = 1.4 \times 10^{-154}$		
$\Phi(\bar{\rho}^2_{MNL} - \bar{\rho}^2_{ORP} > z)$	$\Phi(-16.34) = 2.56 \times 10^{-60}$			$\Phi(-22.93) = 1.16 \times 10^{-116}$			$\Phi(-99.12) = 0$		
$\Phi(\bar{\rho}^2_{MNL} - \bar{\rho}^2_{ORL} > z)$	$\Phi(-15.48) = 2.37 \times 10^{-54}$			$\Phi(-22.93) = 1.16 \times 10^{-116}$			$\Phi(-95.52) = 0$		

6. CONCLUSION

Car ownership is a key element in the study and simulation of urban systems. Hence, it is of interest to develop models of car ownership that are capable of explaining and predicting households' choices. To date, empirical studies have focused on two types of disaggregate models, the ordered and unordered. The first type refers to ORL and ORP whereas the second is mainly represented by the MNL. While a previous comparison suggested the MNL as more appropriate for car ownership modelling over an ORL *(7)*, subsequent empirical analyses have adopted both MNL and ORL as well as ORP models. Thus, it has not been clear, which of the aforementioned models would be more suitable for car ownership modelling.

The objective of this paper has been to evaluate the MNL, ORL and ORP models for car ownership based on a number of data fit measures. We approached this task by empirically studying household car ownership levels using as explanatory variables the household's life-cycle stage, income, race, type of dwelling and number of workers per household. In addition, we included residential density as a measure of urban form assuming that higher densities would discourage households to own more vehicles. The results of model estimations confirm previously published findings highlighting the importance of socio-demographic and economic characteristics of households as well as the type of dwelling and urban form in explaining household car ownership.

A key difference between ordered and unordered models is that unordered models, namely the MNL, are based on the utility maximization principle, whereas ordered models (ORL, ORP) are not. This difference makes the MNL more appealing over ordered models, because findings are based on a solid behavioural framework and not a single continuous propensity measure. Furthermore, a quick glance in the estimation of parameter reveals that the MNL model is more flexible because it allows for alternative-specific effects of explanatory variables across car ownership levels. On the other hand, ordered models are constrained to a unique coefficient per explanatory variable. Finally, comparison tests of data fit among the MNL, ORL and ORP models suggest that the MNL model should be the preferred model structure for household car ownership.

ACKNOWLEDGEMENTS

An earlier version of this paper was presented at the 9th NECTAR Conference held in Porto, Portugal in May 2007. The authors are grateful to Professors Yoram Shiftan of Technion-Israel Institute of Technology and Gerard de Jong of Institute of Transport Studies at University of Leeds for their constructive comments and suggestions. Also, the authors would like to thank Nam Seok Kim, Ph.D. Candidate at the OTB Research Institute of Delft University of Technology, for his help in identifying the Baltimore data for this analysis and the Kinki Branch Office of the Japanese Ministry of Land, Infrastructure and Transportation for providing Osaka's dataset.

REFERENCES

1. Ortuzar, J. D. and L. G. Willumsen *Modeling Transport.* John Wiley & Sons, Ltd, Chichester, 2001.

2. Fenger, J. Urban Air Quality. *Atmospheric Environment*, Vol. 33, 1999, pp. 4877-4900.

3. Miller, E. J., J. D. Hunt, J. E. Abraham and P. A. Salvini. Microsimulating Urban Systems. *Computers, Environment and Urban Systems*, Vol. 28, 2003, pp. 9-44.

4. Bunch, D. S. Automobile Demand and Type Choice In *Handbook of Transport Modeling*; Hensher, D. A. and Button, K. J., (Eds.), Pergamon, New York, 2000, pp 463-479.

5. Miller, E. J., D. S. Kriger and J. D. Hunt. *Integrated Urban Models for Simulation of Transit and Land-Use Policies*. University of Toronto, Joint Program in Transportation and DELCAN Corporation, 1998.

6. Potoglou, D. and P. S. Kanaroglou. Disaggregate Demand Analyses for Conventional and Alternative Fuelled Automobiles: A Review. *International Journal of Sustainable Transportation (forthcoming)*, 2007

7. Bhat, C. R. and V. Pulugurta. A Comparison of Two Alternative Behavioural Choice Mechanisms for Household Auto Ownership Decisions. *Transportation Research Part B*, Vol. 32, No. 1, 1998, pp. 61-75.

8. Chu, Y.-L. Automobile Ownership Analysis Using Ordered Probit Models. In *Transportation Research Record: Journal of the Transportation Research Board*, *No. 1805*, TRB, National Research Council, Washington, D.C., 2002, pp. 60-67.

9. Kim, H. S. and E. Kim. Effects of Public Transit on Automobile Ownership and Use in Households of the USA. *RURDS - The Applied Regional Science Conference (ARSC)*, Vol. 16, No. 3, 2004, pp. 245-262.

10. Ryan, J. M. and G. Han. Vehicle-Ownership Model Using Family Structure and Accessibility Application to Honolulu, Hawaii. In *Transportation Research Record: Journal of the Transportation Research Board*, *No. 1676*, TRB, National Research Council, Washington, D.C., 1999, pp. 1-10.

11. Purvis, C. L. Using 1990 Census Use Microdata Sample to Estimate Demographic and Automobile Ownership Models. In *Transportation Research Record: Journal of the Transportation Research Board*, *No. 1443*, TRB, National Research Council, Washington, D.C., 1994, pp. 21-29.

12. Potoglou, D. and P. S. Kanaroglou. Modelling Car Ownership in Urban Areas: A Case Study of Hamilton, Canada. *Journal of Transportation Geography (in press/ doi:101016/jjtrangeo200701006)*, 2007

13. NHTS. National Household Travel Survey: Federal Highway Administration http://nhts.ornl.gov/download.shtml. Accessed March 2007.

14. van Evert, H., W. Brog and E. Erl. Survey Design: The Past, the Present and the Future In *Travel Survey Methods - Quality and Future Directions*; Stopher, P. and Stecher, C., (Eds.), Elsevier, London, 2006.

15. Susilo, Y. O. and R. Kitamura. Structural Changes in Commuters' Daily Travel: The Case of Auto and Transit Commuters in the Osaka Metropolitan Area of Japan, 1980 through 2000. *Transportation Research A*, 2007 (forthcoming).

16. Borooah, V. K. *Logit and Probit.* Sage Publications, London, 2001.

17. McKelvey, R. D. and W. Zavoina. A Statistical Model for the Analysis for Ordinal Level Dependent Variables. *Journal of Mathematical Sociology*, Vol. 4, 1975, pp. 103-120.

18. Greene, W. H. *Econometric Analysis (4th Edition).* Prentice Hall, Englewood Cliffs, NJ, 2000.

19. McFadden, D. Conditional Logit Analysis of Qualitative Choice Behaviour In *Frontiers in Econometrics*; Zerembka, P., (Ed.), Academic Press, New York, 1973, pp 105-142.

20. Ben-Akiva, M. and S. R. Lerman *Discrete Choice Analysis: Theory and Application to Travel Demand.* MIT Press, Cambridge, 1985.

21. Train, K. *Discrete Choice with Simulations.* Cambridge University Press, Cambridge, 2002.

22. Weeks, M. The Multinomial Probit Model Revisited: A Discussion of Parameter Estimability, Identification and Specification Testing. *Journal of Economic Surveys*, Vol. 11, No. 3, 1997

23. Horowitz, J. L. Reconsidering the Multinomial Probit Model. *Transportation Research Part B*, Vol. 25, No. 6, 1980, pp. 433-438.

24. Akaike, H. 2nd International Symposium on Information Theory, Akademiai Kaido, Budapest, Hungary, 1973, pp 267-281.

25. Hurvich, C. M. and C. Tsai. Regression and Time Series Model Selection in Small Samples. *Biometrika*, Vol. 76, No. 2, 1989, pp. 297-307.

26. Schwarz, G. Estimating the Dimension of a Model. *Annals of Statistics*, Vol. 6, No. 2, 1978, pp. 461-464.

27. Hannan, E. and B. Quinn. The Determination of the Order of an Autoregression. *Journal of the Royal Society Series B (Methodological)*, Vol. 41, No. 2, 1979, pp. 190-195.

28. Greene, W. H. *Nlogit Ver. 3.* Econometric Software Inc., New York & Sydney, 2002.

29. Susilo, Y. O. and K. Maat. The Influence of Built Environment to the Trends in Commuting Journeys in the Netherlands. *Transportation (forthcoming)*, 2007