# Developing an Auto Ownership Model for Edmonton

**John Douglas Hunt***, Professor, University of Calgary
2500 University Drive N.W., Calgary, Alberta, Canada T2N 1N4
Phone: (403) 220-8783
Email: jdhunt@ucalgary.ca

&

**Alan T Brownlee**, Director, Evaluation and Monitoring
Transportation and Streets Department, City of Edmonton
9803 102A Avenue, Edmonton, Alberta, Canada T5J 3A3
Phone: (780) 496-1201
Email: alan.brownlee@edmonton.ca

*Presenter

Paper prepared for presentation

at the **Emerging Best Practices in Urban Transportation Planning (B)** Session

of the 2005 Annual Conference of the

Transportation Association of Canada

Calgary, Alberta

September, 2005

### Acknowledgements

The City of Edmonton funded the model development work described here along with the associated data collection. The preparation of this paper were funded from grants provided by the Natural Sciences and Engineering Council of Canada and the Social Sciences and Humanities Research Council of Canada as part of a Major Collaborative Research Initiative, and also by the Institute for Advanced Policy Research at the University of Calgary. Notwithstanding this funding support, the views expressed in this paper are entirely those of the authors and cannot be taken to indicate the positions of any of the sponsoring agencies. Any errors or omissions are also solely the responsibilities of the authors.

## Abstract

The City of Edmonton has an advanced form of aggregate, equilibrium model for transportation system planning and analysis. Auto ownership is a key input to the mode allocation component of this model – expressed as the average number of vehicles per person 16 or more years old in each of the model zones. This paper describes the development of an endogenous process for forecasting auto ownership, incorporating the influence of changes in travel conditions on the propensity to own autos, and thereby providing representation of an important positive feedback in the system. Observations of the numbers of autos owned by households across the entire Edmonton Region were used together with information on household characteristics and travel conditions at the zonal level from the model to develop a representation of relevant influences on auto ownership, including zonal income, household composition and mode-specific accessibilities. This representation was then incorporated within the larger model system and the entire system re-calibrated. The intention is that this paper will include presentation of the results of some example model runs demonstrating the changes in forecast outputs arising with this extension to the model system.



**Keywords:** Auto Ownership; Personal Transportation Demand Modelling; Edmonton


# 1. INTRODUCTION

## 1.1. Context

The City of Edmonton Transportation and Streets Department operates a transportation planning model covering the Edmonton Region on behalf of the City of Edmonton and the Alberta Infrastructure and Transportation Department. This model, called the Edmonton Transportation Analysis (ETA) Model, is implemented in EMME/2 and uses an aggregate, equilibrium framework, with just under 1100 zones and 11,000 links. It represents all household travel using 25 separate segments based on person category (grades K-9 children, grades 10-12 children, post-secondary students, working age adults and senior citizens) and movement type (home to compulsory, home to discretionary, non-home-based, etc). In each such segment, a separate multi-level nested logit model is used to represent travel choices (from the top of the nest down) concerning trip generation, destination choice, time of day choice, mode choice and peak spreading. The zone-to-zone flows for all segments are assigned to multi-modal networks and the system is iterated to convergence on the networks and at all choice levels (Hunt, 2003).

One of the important inputs to the ETA Model, to the mode split component in particular, is an indication of auto ownership and associated availability for travel for the households in each zone. The variable used is called '*AutoOwn*'. It is defined at the zonal level and is the ratio of number of autos owned by households living in the zone over the number of people 16 or more years old in zone. Autos in this case includes all private vehicles owned and available for use by household members in travel for household purposes, including passenger cars, vans and light (pick-up) trucks. The extent of potential auto users is represented using the number of persons 16 or more years old rather than the number of licensed drivers because the number of persons 16 or more years old in each zone is more directly available. It is already forecast as part of the modelling done in Edmonton, whereas the number of licensed drivers in each zone is not. Even if the number of licensed drivers were forecast, it would likely be based on a transformation of the number of people 16 or more years old in the zone, so the implication is that its use rather than the number of people 16 or more years old would add a step of forecasting work with no benefit in model accuracy

Before the work described here was done, the value for *AutoOwn* in each zone was determined exogenously, using census information for the base year and external forecasts for future years. It was recognized that this was an aspect of the ETA Model that needed improvement; that in reality the auto ownership decision is based on a number of factors, some of which were already present in the ETA Model and influencing its results, such as income and household composition, and some of which are themselves intermediate results of the ETA Model, including mode-specific accessibilities for auto, transit and walk. But with the *AutoOwn* values held static while the values associated with these other factors were allowed to vary, there was a strong potential for inconsistencies where the spatial distribution and related overall model-wide rates of auto ownership would diverge from expectations given the distributions for these other variables. Based on this recognition, work was undertaken to develop model of zonal auto ownership as expressed by *AutoOwn*. The mandate was to develop a model that would (a) not require any new variable values to be determined for each run of the entire ETA Model and (b) not disturb the existing model calibration that used observed base-year values for *AutoOwn*.

This paper describes the work done and results of this mandate, judged successes that can be used in a similar practical context elsewhere.

### 1.2. Structure of Paper

Section 2 outlines some basic considerations regarding the modelling of auto ownership in transportation demand models, covering the factors represented and modelling techniques used in the work described here. Section 3 describes the data used in the model estimations; and Section 4 presents the estimations and their results. Section 5 discusses the selection of the model from the estimations. Section 6 describes the implementation of the model into the larger ETA Model system. Section 7 presents a comparison of results from the ETA Model both with and without the auto ownership model, and Section 8 offers conclusions.

## 2. INFLUENCES ON AUTO OWNERSHIP

### 2.1. Factors

Some of the most prevalent influences on the auto ownership selection behaviour of households are:

- household income;
- auto purchase costs;

- auto operating costs;
- household composition, including the number of people in the household and the age, gender and driving license status of each of these people;
- auto-specific accessibilities; and
- other mode-specific accessibilities, including transit-specific and walk-specific accessiblities.

Ideally, the auto ownership model would represent the influences of all of these factors.

But auto purchase costs do not vary across households or their locations within the Edmonton Region generally, and are entirely exogenous to the ETA Model. This means that a single exogenously-determined adjustment of the auto ownership rates across all zones is all that is required. It would be no more challenging to forecast this single model-wide adjustment in the auto ownership rate directly than to first forecast the future auto purchase cost and then use a model to determine the corresponding future auto ownership rate. So the decision was made to exclude purchase costs from the model.

All of the other factors do vary across households and their locations. Data representing all of these are available for a cross-section of households from the 1994 Household Travel Survey and the associated ETA model networks, which meant that all of these other factors could be considered for the auto ownership model. The specific variables used to represent these factors are described below in Section 3, covering data.

## 2.2. Modelling Techniques

There are two basic modelling approaches available for representing household auto ownership.

One type is 'aggregate', where the dependent variable of the model function is an average, or representative, zonal-level, or even model-area-level, value for the rate of auto ownership. The independent variables are also zonal-level or model-area-level average or representative values. The form of the model function may be linear, but in many cases is non-linear, often following a logit, or S-shaped, curve where the ownership rates approach definite floors and plateaus with a much steeper transition for a particular range of values in-between. An example of this sort of S-shaped curve is shown in Chart 1.

The other type is 'disaggregate', where the dependent variable of the model function is a household-specific or person-specific value for the number of autos owned or corresponding probability distribution for each of different whole numbers of autos owned. The independent variables indicate conditions specific to a particular household or person. Typically, the model represents disaggregate choice behaviour based on random utility theory. Aggregate values, when they are required by other parts of the larger modelling system or as part of the evaluation of model outputs, are determined by adding the results across the full set of households or persons as required.

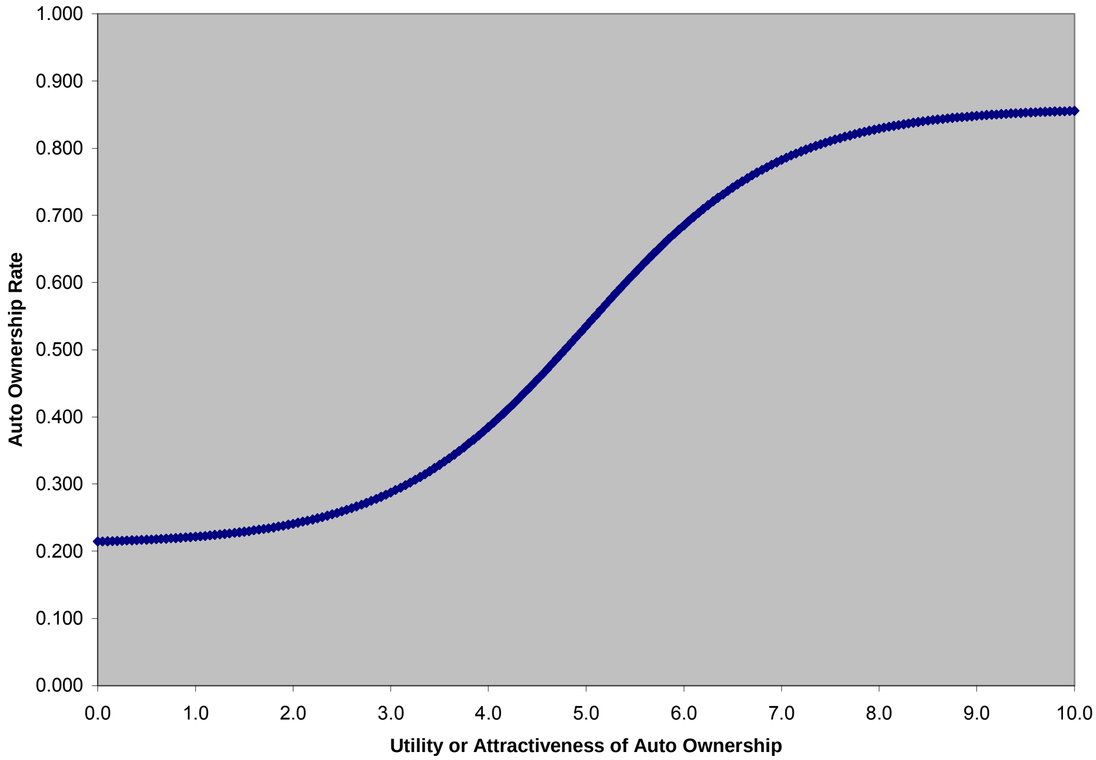


Chart 1: Example S-Curve for Model Form for Auto Ownership

The disaggregate approach is certainly the more sophisticated, and it is better able to capture representation of a wider range of the specific influences on auto ownership and the nature of the behavioural mechanisms involved. But it really makes sense to use it as part of a larger disaggregate treatment of travel demand generally. There was a strong desire in the work described here to avoid adding further data requirements with the addition of the auto ownership model; and the current set of available variables were all zonal-level values. This led to the decision to use an aggregate approach for the auto ownership model at this stage – as a somewhat interim measure – and then use a disaggregate approach at a later stage as part of a larger redevelopment of the entire transportation analysis model using a disaggregate treatment of travel demand generally.

## 3. DATA

### 3.1. Variables

As indicated above, part of the mandate was to avoid expanding the data requirements. Values already used in the ETA Model, at the level of the transport analysis zone (TAZ), and thus available for the development of the auto ownership model without any expansion of the data requirements, were as follows:

- *AutoOwn*: the ratio of the number of autos owned by households resident in the zone over the number of people 16 or more years old resident in the zone;

- *HHIncome*: the average annual before tax income per household resident in the zone, in units of thousands of dollars;
- *PopAll*: the number of people resident in the zone;
- *PopGsc*: the number of grade-school children resident in the zone;
- *PopSnr*: the number of people 65 or more years old resident in the zone;
- *PopWorkAuto*: the number of people resident in the zone who need their own autos at work as part of the work they do;
- *CUAuto*: the accessibility for home to other travel for just the auto mode;
- *CUTransit*: the accessibility for home to other travel for just the transit mode;
- *CUWalk*: the accessibility for home to other travel for just the walk mode.

*AutoOwn* is the representation of auto ownership used in the ETA Model. It is therefore the dependent variable for the auto ownership model. In order to avoid problems with the mathematics in some of the model forms considered (involving logarithm transformations or division operations) if *AutoOwn* is 0 it is reset to 0.001 and if it is 1 it is reset to 0.999. This resetting allows the consideration of the model form to proceed without any significant distortion to the model performance in any practical sense.

*HHIncome* can be used directly as an independent variable representing household income at the zonal level.

*CUAuto, CUTransit and CUWalk* are calculated using the logit destination choice model for the home to other segment in the ETA Model. Each is the logsum for the full set of destination zone alternatives for just the relevant mode, as follows:

$$CU_{m,i} = \ln \left( \Sigma_{z \in Z} \exp (U_{m,i,z}) \right)$$

which expands to:

$$CU_{m,i} = \ln \left( \Sigma_{z \in Z} \exp (\alpha_{travel} \cdot UTravel_{m,i,z} + \alpha_{attr} \cdot Attract_z) \right)$$

where:

| | | |
|---|---|---|
| $z$ | = | index representing model zones; |
| $Z$ | = | set of all model zones; |
| $m$ | = | index representing mode alternative; |
| $i$ | = | index representing origin model zone (for which the accessibility is being calculated); |
| $CU_{m,i}$ | = | logsum value that is the accessibility for zone i for mode m specifically; |
| $U_{m,i,z}$ | = | utility for zone z as a destination alternative from zone i using mode m; |
| $UTravel_{m,i,z}$ | = | utility travel from zone i to zone z using mode m; |
| $Attract_z$ | = | attraction to zone z as a destination; and |
| $\alpha_{travel}, \alpha_{attr}$ | = | model coefficients. |

Consistent with the random utility theory underpinning the destination choice model (Williams, 1977), the logsum $CU_{m,i}$ is the expected maximum utility for the full set of destination choice alternatives from the zone i. Thus, each of *CUAuto, CUTransit and*

*CUWalk* provides an indication of the magnitude of what can be reached and how easily it can be reached by each of the auto, transit and walk modes, respectively, consistent with the destination choice model for home to other travel calibrated for the Edmonton Region using observations of mode and destination choice behaviour collected in Edmonton. As such, they can be used directly as independent variables representing the mode-specific accessibilities for zones. They are the same as a form of accessibility measure identified by Wilson (1970) and often used as such. They also include representation of development density and auto operating costs as components influencing the length and cost of travel, so their use means that changes in these components will impact auto ownership.

In order to distinguish the impact of changes in *CUTransit* from the impact of not having any transit with walk access in a zone, an additional zonal variable *NoTransit* is defined for each zone, and if the zone has transit with walk access then *NoTransit* is set to '0' and *CUTransit* is unchanged whereas if the zone has no transit with walk access then *NoTransit* is set to '1' and *CUTransit* is set to '0'.

Independent variable representing aspects of household composition at the zonal level (with scale effects removed) can be developed from the numbers of people in different categories as follows:

*PropGsc = PopGsc / PopAll*

*PropSnr = PopSnr / PopAll*

*PropWorkAuto = PopWorkAuto / PopAll*

*PropGsc* can be used as an independent variable representing the overall proportion of grade-school children in households at the zonal level, and both *PropSnr* and *PropWorkAuto* similarly used as independent variables representing the overall proportions of seniors and those who need their own autos at work as part of the work they do, respectively.

### 3.2. Sample

Values for the variables identified above were available for the ETA Model base year for a total of 744 zones, after the zones with no population had been removed. This set of values for 744 zones was the sample used for the model estimation described in the next section below.

## 4. ESTIMATIONS

Models were estimated with the sample of observations for 744 zones described above. Three alternative model forms were considered in these estimations: linear, exponential and S-curve. For each of these forms, a model with the full set of independent variables was considered first. Then, as appropriate, a series of additional models with subsets of the independent variables were considered, based on the results of this first estimation and subsequent estimations.

Multiple linear regression was used to perform the estimations, with transformations of the variables used as required in order to consider the alternative model forms.

The model forms and the results for each are described below.

### 4.1. Linear

The linear form of the model is as follows:

$AutoOwn_i = \Sigma_{k \in K}\, \alpha_k \cdot \mathrm{Attribute}_{k,i}$

where:

| | | |
|---|---|---|
| k | = | index representing model attributes (the independent variables defined above in Section 3.1.); |
| K | = | set of all model attributes zones; |
| $\mathrm{Attribute}_{k,i}$ | = | value for attribute k for zone i; and |
| $\alpha_k$ | = | model coefficient associated with attribute k. |

The model estimation provides 'best fit' estimates for the values for the vector of $\alpha_k$ for all $k \in K$ along with standard linear regression statistics concerning the quality of the fit of the resulting estimated model to the data.

The estimation results for the full set of independent variables are shown in Table 1. A graphical comparison of the modelled values for $AutoOwn_i$ using these estimation results with the corresponding observed values is provided in Figure 1.

The R square value indicates a reasonably good fit for the model overall.

The coefficient estimates for all of the attributes have reasonably low standard errors in relative terms – resulting in t-statistics that are reasonably high in absolute magnitude and P-values that are reasonably low – with the exception of the one for *PropWorkAuto*, which has a t-statistic that is slightly less than 1 in absolute magnitude and a P-value that is almost 0.33.

Table 1: Estimation Results for Linear Model with All Independent Variables

| *Regression Fit Statistics* | |
|---|---|
| Multiple R | 0.873 |
| R Square | 0.761 |
| Adjusted R Square | 0.759 |
| Standard Error | 0.070 |
| Observations | 744 |

| *ANOVA* | *df* | *SS* | *MS* | *F* | *Significance F* |
|---|---|---|---|---|---|
| Regression | 8 | 11.5721 | 1.4465 | 293.0245 | 8.5E-223 |
| Residual | 735 | 3.6283 | 0.0049 | | |
| Total | 743 | 15.2004 | | | |

| *Attribute* | *Coefficient* | *Stnd Error* | *t Stat* | *P-value* | *Lower 95%* | *Upper 95%* |
|---|---|---|---|---|---|---|
| Constant | 0.4854 | 0.0232 | 20.8920 | 0.0000 | 0.4398 | 0.5311 |
| CUAuto | 0.0963 | 0.0067 | 14.4554 | 0.0000 | 0.0832 | 0.1093 |
| CUWalk | -0.0314 | 0.0023 | -13.4077 | 0.0000 | -0.0360 | -0.0268 |
| CUTransit | -0.0587 | 0.0071 | -8.2756 | 0.0000 | -0.0727 | -0.0448 |
| NoTransit | 0.0236 | 0.0140 | 1.6860 | 0.0922 | -0.0039 | 0.0511 |
| HHIncome | 1.509E-03 | 2.233E-04 | 6.7576 | 0.0000 | 0.0000 | 0.0000 |
| PropGrs | 0.1731 | 0.0405 | 4.2787 | 0.0000 | 0.0937 | 0.2525 |
| PropSnr | -0.0610 | 0.0282 | -2.1646 | 0.0307 | -0.1164 | -0.0057 |
| PropWorkAuto | -0.0463 | 0.0469 | -0.9875 | 0.3237 | -0.1385 | 0.0458 |

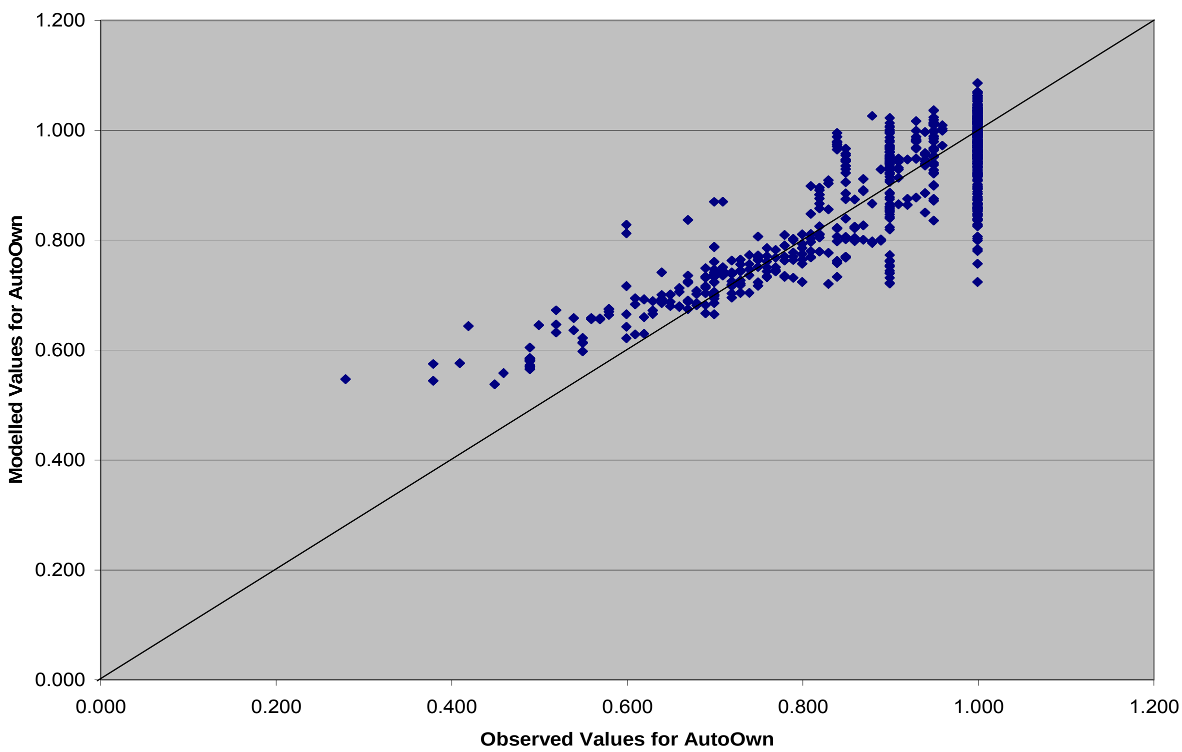


Figure 1: AutoOwn Observed Values Versus Modelled Values Using Linear Model with Estimation Results for All Independent Variables

The P-value for a coefficient estimate is a standard statistic provided in multiple linear regression. It is the probability that the estimated value or beyond (greater if positive, lesser if negative) would be obtained if the actual (or 'population') value of the coefficient is 0. If the P-value is reasonably large (like around 0.33) then it is fairly risky to reject the possibility that the corresponding attribute has no effect and the value of the estimated coefficient is merely the result of random effects. In practice – a high P-value casts a 'shadow' over the value of the estimated coefficient, suggesting it is comparatively imprecise and that the effect of the corresponding attribute is less clearly indicated and perhaps may not be substantial enough for the attribute to be included as one of the independent variables. The implication here is that the indication of the effect of *PropWorkAuto* (the proportion of people resident in the zone who need their own autos at work as part of the work they do) provided by the estimation results is comparatively less reliable, not 'significantly different from 0' and perhaps should be ignored – by re-doing the model estimation with *PropWorkAuto* removed. In the work described here, the model estimation was not re-done with *PropWorkAuto* removed; rather, the next model form was considered first with the full set of independent variables.

Returning to the estimation results for the linear model with all of the independent variables, the estimated coefficients for all of the attributes have signs (positive or negative) that are consistent with expectations – again, with the exception of the one for *PropWorkAuto*.

The sign for *CUAuto* is positive, consistent with increases in auto-specific accessibility making the use of auto more attractive and thus leading to more auto ownership. Auto-specific accessibility can be increased in the ETA Model by adding road capacity, upgrading road functional classes, improving road connectivity to large activity centres and by decreasing auto operating costs.

The signs for *CUWalk* and *CUTransit* are both negative, consistent with increases in these non-auto-specific accessibilities making the use of auto less attractive and thus leading to less auto ownership.

Walk-specific accessibility can be increased in the ETA Model by improving the connectivity provided by walking links and by bringing activities closer to households – reflecting increases in settlement density and/or the extent of mixed use in neighbourhoods. Thus, when an auto ownership model that includes *CUWalk* as an independent variable with a negative coefficient is added to the ETA Model, then the full ETA Model system indicates that an increase in settlement density results in less auto ownership and consequently less auto use - with the behavioural mechanism linking this cause and effect based on a mode-specific accessibility consistent with the model of destination choice behaviour observed in Edmonton.

Transit-specific accessibility can be increased in the ETA Model by improving the connectivity provided by walking links to and from transit, moving transit closer to activities generally, increasing the frequency and speed of transit services and reducing fares. Again, increases in settlement density act to decrease auto ownership and hence, decrease auto use in the ETA Model – with these acting via the impact on transit-specific accessibilities using a representation consistent with the model destination choice behaviour observed in Edmonton.

The sign for *NOTransit* is positive, consistent with a complete lack of walk-access transit resulting in a greater reliance on the auto and thus leading to more auto ownership.

The sign for *HHIncome* is positive, consistent with higher income households being able to afford more autos and thus displaying higher auto ownership.

The sign for *PropGrs* is positive – and comparatively large, consistent with households (or more precisely neighbourhoods) with more school-children relying much more on autos for their travel – which is presumably much more disparate and also much more family-group-oriented and chauffeuring-related than for others – leading to much higher auto ownership rates.

The sign for PropSnr is negative, consistent with seniors travelling less generally and relying less on autos in particular leading to lower auto ownership rates.

The sign for *PropWorkAuto* is negative, which is not what was expected beforehand; rather, it was thought beforehand that a greater proportion of people needing their own auto at work as part of the work they do would lead to more auto ownership, which would result in a positive value for the coefficient for *PropWorkAuto*. The negative sign obtained in this case was at first attributed to the imprecision of the estimate, as indicated by the high P-value. But this negative sign result for *PropWorkAuto* was found to be more systematic, and is considered further below in the discussion of some of the other estimation results.

Considering Figure 1: the comparison of observed versus modelled values shows a skewing pattern where the model over-estimates consistently when the auto ownership is relatively low, up to about 0.60 or a bit higher. This sort of bias in the model error implies that the model is systematically missing or distorting certain important effects.

### 4.2. Exponential

The exponential form of the model is as follows:

$AutoOwn_i = \exp ( \Sigma_{k \in K}\ \alpha_k \cdot \text{Attribute}_{k,i} )$

where:

| | | |
|---|---|---|
| k | = | index representing model attributes (the independent variables defined above in Section 3.1.); |
| K | = | set of all model attributes zones; |
| $\text{Attribute}_{k,i}$ | = | value for attribute k for zone i; and |
| $\alpha_k$ | = | model coefficient associated with attribute k. |

Again, the model estimation provides ‘best fit’ estimates for the values for the vector of $\alpha_k$ for all $k \in K$. Multiple linear regression is still used, but with a transformed version of the equation as follows:

$$\ln ( AutoOwn_i ) = \Sigma_{k \in K}\ \alpha_k \cdot \text{Attribute}_{k,i}$$

The estimation results for the full set of independent variables are shown in Table 2; and a graphical comparison of the modelled values for $AutoOwn_i$ using these estimation results with the corresponding observed values is provided in Figure 2.

The R square value is a bit lower than for the linear model, but it still indicates a reasonably good fit for the model overall.

The coefficient estimates for most of the attributes have reasonably low P-values that are reasonably low. The P-values for *PropWorkAuto* and *NoTransit* are the exceptions. The t-statistic for NoTransit is only 0.23, which is very small.

The estimated coefficients for all of the attributes have the same signs they had with the linear model; so they are all consistent with expectations – again, with the exception of the one for *PropWorkAuto*.

The comparison of observed versus modelled values shows the same sort of skewing pattern where the model over-estimates consistently when the auto ownership is relatively low, up to about 0.60.

Comparing the estimation results for the linear and exponential models: the linear model is better.

### 4.3. S-Curve

The logit, or S-curve, form of the model is as follows:

$$AutoOwn_i = \exp ( \Sigma_{k \in K}\ \alpha_k \cdot \text{Attribute}_{k,i} ) / ( 1 + \exp ( \Sigma_{k \in K}\ \alpha_k \cdot \text{Attribute}_{k,i} ) )$$

Once again, multiple linear regression is still used to get ‘best fit’ estimates for the values for the vector of $\alpha_k$ for all $k \in K$, but with a transformed version of the equation as follows:

$$\ln ( AutoOwn_i \ / ( 1 - AutoOwn_i ) ) = \Sigma_{k \in K}\ \alpha_k \cdot \text{Attribute}_{k,i}$$

The estimation results for the full set of independent variables are shown in Table 3; and a graphical comparison of the modelled values for $AutoOwn_i$ using these estimation results with the corresponding observed values is provided in Figure 3.

The R square value is lower than for the other two models considered above, indicating a fit that is still fairly good overall.

Table 2: Estimation Results for Exponential Model with All Independent Variables

| *Regression Fit Statistics* | |
|---|---|
| Multiple R | 0.860 |
| R Square | 0.739 |
| Adjusted R Square | 0.736 |
| Standard Error | 0.097 |
| Observations | 744 |

| *ANOVA* | *df* | *SS* | *MS* | *F* | *Significance F* |
|---|---|---|---|---|---|
| Regression | 8 | 19.6206 | 2.4526 | 260.5412 | 9.2E-209 |
| Residual | 735 | 6.9188 | 0.0094 | | |
| Total | 743 | 26.5394 | | | |

| *Attribute* | *Coefficient* | *Stnd Error* | *t Stat* | *P-value* | *Lower 95%* | *Upper 95%* |
|---|---|---|---|---|---|---|
| Constant | -0.6861 | 0.0321 | -21.3846 | 0.0000 | -0.7491 | -0.6232 |
| CUAuto | 0.1354 | 0.0092 | 14.7261 | 0.0000 | 0.1174 | 0.1535 |
| CUWalk | -0.0426 | 0.0032 | -13.1784 | 0.0000 | -0.0490 | -0.0363 |
| CUTransit | -0.0841 | 0.0098 | -8.5830 | 0.0000 | -0.1034 | -0.0649 |
| NoTransit | 0.0046 | 0.0193 | 0.2387 | 0.8114 | -0.0334 | 0.0426 |
| HHIncome | 2.096E-03 | 3.083E-04 | 6.7990 | 0.0000 | 0.0000 | 0.0000 |
| PropGrs | 0.2493 | 0.0559 | 4.4619 | 0.0000 | 0.1396 | 0.3589 |
| PropSnr | -0.0618 | 0.0389 | -1.5865 | 0.1130 | -0.1382 | 0.0147 |
| PropWorkAuto | -0.0358 | 0.0648 | -0.5525 | 0.5807 | -0.1630 | 0.0914 |

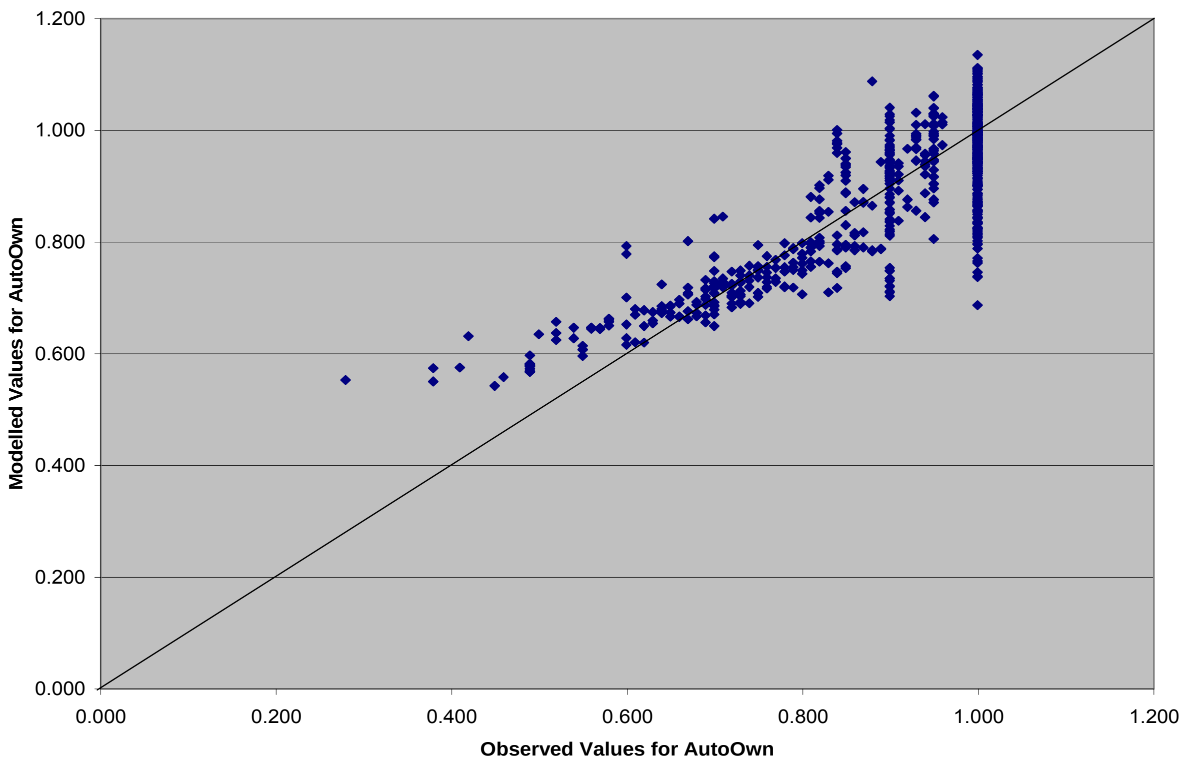


Figure 2: AutoOwn Observed Values Versus Modelled Values Using Exponential Model with Estimation Results for All Independent Variables

Table 3: Estimation Results for S-Curve Model with All Independent Variables

| *Regression Fit Statistics* | |
|---|---|
| Multiple R | 0.754 |
| R Square | 0.568 |
| Adjusted R Square | 0.563 |
| Standard Error | 1.813 |
| Observations | 744 |

| *ANOVA* | *df* | *SS* | *MS* | *F* | *Significance F* |
|---|---|---|---|---|---|
| Regression | 8 | 3176.6501 | 397.0813 | 120.8283 | 1.6E-128 |
| Residual | 735 | 2415.4499 | 3.2863 | | |
| Total | 743 | 5592.1000 | | | |

| *Attribute* | *Coefficient* | *Stnd Error* | *t Stat* | *P-value* | *Lower 95%* | *Upper 95%* |
|---|---|---|---|---|---|---|
| Constant | -1.7613 | 0.5995 | -2.9379 | 0.0034 | -2.9383 | -0.5843 |
| CUAuto | 1.1712 | 0.1718 | 6.8163 | 0.0000 | 0.8339 | 1.5085 |
| CUWalk | -0.5244 | 0.0605 | -8.6732 | 0.0000 | -0.6431 | -0.4057 |
| CUTransit | -0.2824 | 0.1832 | -1.5417 | 0.1236 | -0.6420 | 0.0772 |
| NoTransit | 1.7784 | 0.3615 | 4.9191 | 0.0000 | 1.0686 | 2.4881 |
| HHIncome | 7.347E-03 | 5.761E-03 | 1.2752 | 0.2026 | 0.0000 | 0.0000 |
| PropGrs | 4.8429 | 1.0438 | 4.6395 | 0.0000 | 2.7936 | 6.8921 |
| PropSnr | -1.3135 | 0.7275 | -1.8056 | 0.0714 | -2.7416 | 0.1146 |
| PropWorkAuto | -2.9576 | 1.2108 | -2.4426 | 0.0148 | -5.3347 | -0.5805 |

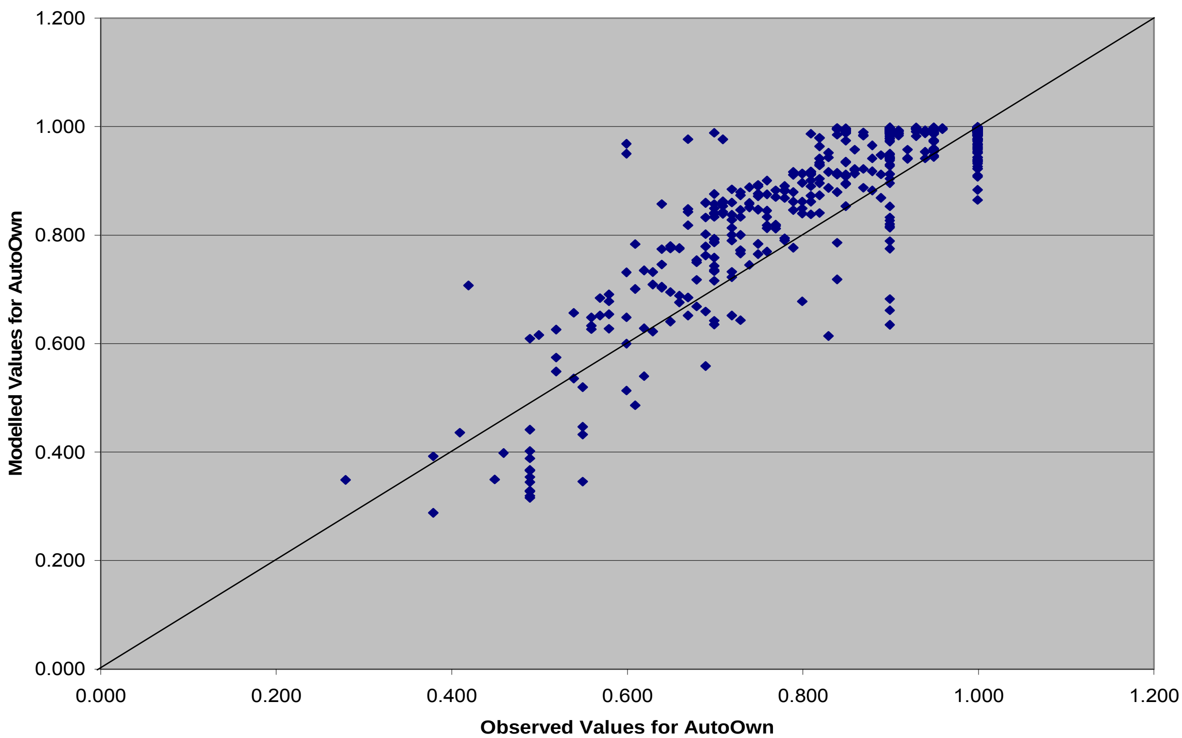


Figure 3: AutoOwn Observed Values Versus Modelled Values Using S-Curve Model with Estimation Results for All Independent Variables

The coefficient estimates for all of the attributes have low P-values, better than the situations with the other two models. The highest P-value is for *HHIncome*, which is perhaps a bit surprising given that income was found to have such a strong positive effect and the P-value was so low with the other two models. It may be that with the S-curve form the range of values over which income plays a strong role as an explanatory variable is reduced. The much lower P-value for *PropWorkAuto* and the somewhat higher P-value for *CUTransit* relative to the other models may also be at least in part due to similar sorts of changes in the range of values over which these attributes play strong roles.

The estimated coefficients for all of the attributes again have the same signs they had previously. The only sign inconsistent with initial expectations is the negative one for *PropWorkAuto*.

In this case the much lower P-value and corresponding higher t-statistic for *PropWorkAuto* contradicts the idea that the negative sign is merely a random effect arising with an imprecise estimate; rather, it suggests there is something more systematic that is acting. It may be that people in occupations where they are more likely to need their own auto at work for work purposes are less willing or able to have additional household autos. Some of the effect related to income with the other models may be emerging as something more linked to related occupations over the ranges enjoying greater emphasis with the S-curve form. In any case, the consistency of the negative sign for *PropWorkAuto* across all models, and its low P-value in this case in particular, suggests it is appropriate to leave it in the model. The implication is that the initial expectation of a positive sign for *PropWorkAuto* is now judged to be questionable, and possibly wrong.

In this case the comparison of observed versus modelled values in Figure 3 does not show any distinct skewing patterns.

## 5. SELECTED MODEL

The S-curve form of the model is selected for use in implementation. It did not get the highest R square value, but the comparison of observed versus modelled values did not show the same skewing pattern and all of the estimated coefficients had very low P-values. It would seem to do the best job of picking up relevant influences without systematic distortion even though it did not score the highest on the formal goodness-of-fit measure.

When all of the independent variables are included, all of the P-values are sufficiently low that it is appropriate to leave everything in. Thus, there is no need to re-do the model estimation; the S-curve model is adopted with all the independent variables included using the coefficient estimates indicated in Table 3.

## 6. IMPLEMENTATION

### 6.1. Pivot-Point Approach

The addition of the auto ownership model was not to alter the base year calibration of the ETA Model. Consequently, it was necessary to use a 'pivot-point' form of treatment, where the results of the auto ownership model are used to adjust the auto ownership for a zone off of the base year value. This is accomplished by adding a zone-specific correction term to the model-determined value for each zone so that the resulting summed value matches the corresponding base-year value in each zone, as follows:

$$AutoOwn^{O}_{i} = AutoOwn^{M}_{i} + AutoOwn^{C}_{i}$$

where:

$AutoOwn^{O}_{i}$ = observed value for auto ownership in zone i for the base year, which is the value input to the ETA Model for the base year;
$AutoOwn^{M}_{i}$ = model-determined value for auto ownership in zone i for the base year, the value of the dependent variable output from the S-curve model selected as described above in Section 5;
$AutoOwn^{C}_{i}$ = correction term for auto ownership in zone i.

The correction term is retained for each zone, and is added in the same way to the model-determined value in any model run, as follows:

$$AutoOwn^{OS}_{i} = AutoOwn^{MS}_{i} + AutoOwn^{C}_{i}$$

where:

$AutoOwn^{OS}_{i}$ = value for auto ownership in zone i for any scenario run of ETA Model, the value input to the ETA Model for the scenario;
$AutoOwn^{MS}_{i}$ = model-determined value for auto ownership in zone i for any scenario run of ETA Model, the value of the dependent variable output from the S-curve model selected as described above in Section 5.

Thus, changes in the model-determined value result in changes in the summed value, and the summed value is used as the input to the rest of the ETA Model.

### 6.2. Placement in ETA Model Operation

The auto ownership model is run as a pre-cursor to the component of the ETA Model that determines the mode split for all zone-to-zone interactions for each of the 25 market segments of person and trip type.

Mode split in the ETA Model is calculated simultaneously with time-of-day allocation and peak-spreading in a series of iterations seeking an equilibrium solution for all components of travel demand together with the mode-specific networks of transportation supply. In each iteration, the mode split / time-of-day / peak-spreading allocation is re-calculated in response to changes in supply conditions and in the demand to be allocated. With the auto ownership model run as a pre-cursor, the value for *AutoOwn* used in each zone is similarly updated in response to changes in supply conditions before it is used in the utility functions for the auto mode in the model split / time-of-day / peak-spreading component.

Because *CUAuto* and *AutoOwn* have positive influences on each other – that is, *CUAuto* increases when *AutoOwn* goes up, and *AutoOwn* increases when *CUAuto* goes up – the insertion of the auto ownership model into the calculations in the ETA Model adds a modest element of positive feedback to the overall iterative process. But this positive feedback is very small relative to the dampening effects in the process overall and there is no potential for convergence difficulties to be introduced with the addition of the auto ownership model specifically.

## 7. RESULTS

The outputs of two model runs with the same inputs will be compared, where one run is done without the auto ownership model and one is done with the auto ownership model implemented as described above. The results of this comparison will be included in the presentation at the conference in September.

## 8. CONCLUSIONS

The work described here was successfully able to use zonal data to develop an elastic auto ownership treatment without adding further data requirements as intended. The ETA Model is a better, more complete, model with this added response behaviour included. The model run times are a bit longer, but the model representation and results are more consistent.

The estimations done in the work and presented here indicated the S-curve form of the model to be the 'best' of those considered. Specifically, the S-curve model was found to be better than the linear and exponential forms, not in terms of the goodness-of-fit score specifically (the R square value was actually lower) but in terms of providing both (a) a more uniform spread in the model error, suggesting a less systematic bias, and (b) a set of individual parameter estimates with better statistics regarding the precision of the estimates, particularly for those with the highest standard errors.

All of the estimation results for the individual parameters were consistent with initial expectations regarding sign and general relative strength of influence, except for the parameter concerning the proportion of workers needing auto at work as part of work done, where the result was a negative sign rather than the positive one initially expected. The result for this particular exception was still accepted and used in the selected model – acknowledging that may be more complex influences acting. Thus, for the full set of individual parameter estimates, the indications regarding the influences of the attributes that were considered on auto ownership – and at the zonal level – are as follows:

- average household income has a moderate positive influence;
- the proportion of grade-school children has a strong positive influence;
- the proportion of people 65 or more years old has a moderate negative influence;
- the proportion of people who need their own autos at work as part of the work they do has a moderate negative influence;
- the accessibility for home to other travel for just the auto mode has a strong positive influence;
- the accessibility for home to other travel for just the transit mode has a moderate negative influence; and
- the availability of transit with walk access at the home end has a strong negative influence; and
- the accessibility for home to other travel for just the walk mode has a strong negative influence.

More specific indications of the influences of these attributes are provided by the estimated values for the parameters and their associated t-statistics, which are listed in the tables of results above, and Table 3 in particular.

The results obtained for both *CUWalk* and *CUTransit* have relevance with regard to some of the ongoing discussions concerning the influence of development density on auto use generally. Both *CUWalk* and *CUTransit* are highly correlated positively with development density. The indication in the results that increases in *CUWalk* and in *CUTransit* are related to decreases in auto ownership helps identify the causal-behavioural linkage between development density and auto use: increases in development density bring about increases in *CUWalk* and/or *CUTransit*, which means that people can get to more on foot and by transit, without their autos, resulting in less auto ownership as well as less auto use. The representation provided with *CUWalk* and *CUTransit* is 'richer' than the one provided with development density, because it is more directly consistent with the modelling of travel behaviour generally and because it allows the impacts of changes in other aspects of transit service and walking beyond just those related to density, such as fares and routing connectivities, to also be represented.

Even with the above, it is still difficult to separate the actual cause and effect sequence in car ownership changes in these estimation results based on cross-sectional data at a specific point in time. Households in zones where auto use is comparatively more attractive – for whatever reasons – may have lived there first and increased their rate of auto ownership over time in response to the comparatively greater attractiveness of auto use; or, households with preferences for auto use and greater auto ownership at first may have been led to locate more frequently in places where there is comparatively better support for, or even reliance on, auto use. But the temporal dynamics of these patterns are not at issue here. The auto ownership model as developed here merely seeks to identify the pattern in auto ownership that arises across space (in the model zones) in response to the related conditions. It does not seek to identify if a change in auto ownership in a zone is the result of new households with different auto ownership moving into the zone or if it is changes in auto ownership by the households already present in the zone – as intended, it only determines the change in response to the changes in the conditions related to auto ownership as identified in the statistical estimations.

The practical implications of this work are that the success of the approach used and the results provided point the way forward in a practical context where there is a desire for an auto ownership model providing endogenous auto ownership values for a larger transport modelling system. The methods used and the results could be used elsewhere.

## REFERENCES

Hunt JD, 2003, Modelling transportation policy impacts on mobility benefits and Kyoto-Protocol-related emissions. Built Environment. 29(1):48-65.

Williams, 1977, On the formation of travel demand models and economic evaluation measures of user benefit. Environment and Planning 9A:285-344.

Wilson AG, 1970, Entropy in Urban and Regional Modelling. Pion, London, UK.