A Comparative Analysis of Alternate Econometric Structures for Trip-Generation Models

Introduction

Trip-generation models have been traditionally estimated using linear-regression structures. This methodology does not recognize the non-negativity and integer nature of the trips. Relatively few studies have focused on only one type of trips and their conclusions are specific to the distribution patterns for that trip purpose. The practical benefits of count-data models for trip generation need to be fully established.

Data

The 2001 NHTS data are used for model estimations and 2009 NHTS data for validations. Samples used through screening:
- a sample of 86,063 households (728,104 trips) from the 2001 data
- a sample of 39,950 households (409,086 trips) from the 2001 data

Model Structures

Linear-regression (LL) model
\[ Y_i = \beta_0 + \beta_1 X_{1i} + \cdots + \beta_p X_{pi} + \epsilon_i \]
where \( i \) is the index for households, \( k \) is index for explanatory variables, \( Y \) is the observed number of trips, \( X \) is the value of explanatory variable for household \( i \), and \( \beta \) is a coefficient corresponding to the \( X \).
Log-linear (LL) model
\[ \ln(\lambda_i) = \beta_1 + \beta_2 X_{1i} + \cdots + \beta_p X_{pi} + \epsilon_i \]
(zero trip households set to 0.01 to avoid mathematical error)
Negative-binomial (NB) model
\[ \frac{\lambda_1}{\lambda_2} = \frac{(1 + \theta)}{\lambda_1} \cdot \frac{(1 + \theta)}{\lambda_2} \]
where \( \theta = \frac{\lambda_2}{\lambda_1} \)

Analysis Results

There are 10 explanatory variables used in modeling:
- 6 socio-economic descriptors (household size, number of workers, vehicle share, presence of children, household income, and housing tenure)
- 2 location descriptors (urban and census region)
- 2 temporal descriptors (day of the week and month of the year)

Log-linear (LL) model
\[ \ln(\lambda_i) = \beta_1 + \beta_2 X_{1i} + \cdots + \beta_p X_{pi} + \epsilon_i \]

Count-data models (NB and Ordered-probit) probabilities are directly calculated from the formulas:
- Continuous models (linear and Log-linear): used cut-off points (such as 0.5, 1.5, 2.5, an so on) to discretize the continuous probabilities For linear-regression model:

For log-linear model:
\[ \ln(P(Y = m | \lambda_i)) = \ln(\lambda_i) - m \cdot \ln(\lambda_i) - m \cdot \ln(\lambda_i) \]

where \( m \) is the estimated standard deviation of the error term.

Predicted Frequency-Distribution (aggregate validation)

- Average the calculated probabilities across all households for each trip level (0, 1, 2, 3, 15) to determine the predicted share of households.
- Calculation of mean absolute error (MAE): (observed share – predicted share) / 15
- Statistical test to compare model superiority

Conclusions

Predictive validations indicate that the ordered-probit models are able to replicate the trip generation pattern better than standard linear, log-linear and negative-binomial models for all three trip purposes. The negative-binomial model performs reasonably well in the case of the NHB trips which have a monotonically-decreasing distribution pattern. The negative-binomial and the log-linear models have comparable mean errors for aggregate predictions. The use of ordered-probit models is recommended as a substitute for the traditional linear-regression models.