Trip Generation Models for Infrequent Trips

JOSE MONZON, KONSTADINOS GOULIAS, AND RYUICHI KITAMURA

The adequacy of conventional linear regression models in trip generation analysis is examined in this study. Simulation experiments are conducted to determine whether model coefficients can be accurately estimated by least-squares estimation when the dependent variable is a nonnegative integer. Following this, nonlinear, two-stage model systems are estimated by using an empirical data set to examine whether more elaborate representation of the decision process underlying trip generation will lead to improved prediction. The results of this study indicate that linear regression models of trip generation offer consistent coefficient estimates and accurate predictions, and improved performance may not be obtained by adopting more complex model systems.

The most frequently used statistical methods for trip generation analysis are the least-squares estimation of linear regression models and trip rate analysis based on cross-classification of households on the basis of a few grouping variables. Both methods draw on principles that are relatively easy to understand, and models can be estimated using commonly available statistical software packages. However, these methods involve certain assumptions and limitations that need to be well understood for valid formulation of trip generation models.

Sample size requirements usually limit the number of grouping variables that can be used in cross-classification analysis, leaving least-squares regression the standard method used whenever an adequate data set and statistical package are available. Linear regression analysis is based on the assumption that the dependent variable (number of trips) is an untruncated continuous variable. Also important is the typical assumption that one model structure explains the entire range of trip generation behavior.

These assumptions may not be entirely satisfied in typical trip generation analyses. The dependent variable in this case is a nonnegative discrete variable, not an untruncated continuous variable. Trip generation behavior may result from a two-stage decision process in which a decision to make trips on a given day is made first; then, given that trips will be made at all, the number of trips is determined. This can be most typically seen in trip generation by purpose (e.g., the number of shopping trips on a given day) or by mode (e.g., the number of transit trips).

The question that naturally arises is whether linear least-squares regression can be used successfully in trip generation analysis when its assumptions are not satisfied. The latter is especially the case when models are formulated for infrequent trips whose observed frequencies are zero for many behavioral units. The dependent variable will then be heavily truncated and the underlying decision process may involve more than one stage that cannot be adequately represented by a single model.

The objective of this paper is to shed light on the following two questions.

1. Is the linear regression method suited for trip generation analysis in which the dependent variable (number of trips) is a nonnegative integer rather than an untruncated continuous variable?
2. Can a single regression model capture trip generation behavior that may involve multistage decision processes?

The first question is examined through simulation experiments in which the number of trips made by an individual has a Poisson distribution. In the simulation, discrete numbers are generated from Poisson distributions as the number of trips generated, the parameters of the distributions are estimated by least-squares regression, and the accuracy of the parameter estimates is examined against the true values used in the simulation to generate the data. Timmermans (*1*) offers a comprehensive discussion of trip generation analysis by examining a set of alternative trip generation model formulations, including Poisson regression models, and testing their goodness-of-fit empirically. The emphasis in this study is on the extent of estimation errors that result from the application of linear regression models to Poisson data (linear regression models are misspecified in this case).

The second question is examined by estimating two models on empirical data and comparing their relative fits. The first model is a regular linear regression model. The second is based on the assumption that the trip generation process consists of two stages: in the first stage the decision is made whether to make trips of a given type; then in the second stage the number of trips is determined.

The rest of this paper is organized as follows. Trip generation models used in this study are briefly described. The results of simulation experiments are presented after that together with the discussion on whether linear regression models can be successfully used with discrete and truncated dependent variables. The next two sections offer a description of the data used to address the second question, the results of the empirical analysis, and a comparison of the two models. The final section summarizes the study.

J. Monzon, Department of Civil Engineering, University of California at Davis, Davis, Calif. 95616. Current affiliation: Planning Department, General Highway Directorate, Guatemala City, Guatemala. K. Goulias and R. Kitamura, Department of Civil Engineering, University of California at Davis, Davis, Calif. 95616.

TRIP GENERATION METHODS

Cross-classification analysis of trip generation is based on the premise that each group of households, defined in terms of a set of grouping variables, has an average trip rate that remains stable over time. The grouping variables are categorical, and groups of households can be defined by combinations of their categories. An important advantage of this straightforward method is its capability to represent the interaction effect of the classification variables, that is, systematic variation in trip rates that is uniquely associated with a particular combination of categories.

For example, let S and T be the grouping variables, and let s and t be categories of S and T, respectively. In the case of household trip generation, variable S may represent household size and T, the number of cars available to the household. Let the set of values S be assumed to be {1, 2, 3, 4, 5, 6 or more} and that for T be {0, 1, 2, 3 or more}. Let the mean trip rate, $Y(st)$, of the group of households with $S = s$ and $T = t$ be

$$Y(st) = \mu + V(s) + W(t) + Z(st)$$

where

μ = grand mean,
$V(s)$ and $W(t)$ = the effects of category s of variable S and category t of variable T, respectively, and
$Z(st)$ = interaction effect of category s and category t.

$V(s)$ and $W(t)$ represent the effects that are attributable to S and T, respectively, whereas effect $Z(st)$ is the contribution of the particular combination of categories. The statistical significance of these models can be tested by analysis of variance (ANOVA), available in most statistical packages.

In a linear regression model, the expected number of trips made by household i is represented as

$$Y_i = \beta_0 + \beta_1 H_i + \beta_2 A_i$$

where

β = model coefficient,
H_i = number of persons in household i, and
A_i = number of cars available to household i.

In this formulation, β_1 represents the average number of trips generated per household member, and β_2 the average number of trips per automobile. The number of trips is linearly related to the explanatory variables, and no interaction effect is assumed in this formulation.

Interaction effects can be represented in a linear regression model by introducing terms representing combinations of categories. Possible nonlinear effects of an explanatory variable can also be included in a linear model by using nonlinear transformation of the variable (including a step function represented by a set of dummy variables). Although it is limited to the case in which the model is linear in terms of its coefficients, the least-squares method can be used in a variety of cases involving nonlinear relations or interaction effects.

A critical limitation of the least-squares approach to trip generation may stem from the assumption that the random variation in the dependent variable can be represented by a random error term that has a continuous, untruncated distribution. The dependent variable of trip generation analysis, the number of trips, is a nonnegative integer. Ideally this variable can be modeled by using a discrete distribution, such as a Poisson or negative binomial distribution (*1,2*). Application of the least-squares method, therefore, assumes that this discrete distribution can be replaced by a continuous distribution.

Problems arise when the expected number of trips is close to 0. For example, suppose the expected value is 0.2 trip. Then possible values that the error term may assume are −0.2, 0.8, 1.8, 2.8, and so on. The error distribution is truncated at −0.2 with the probability mass associated with this error value equaling the probability that no trip will be made. If the number of trips has a Poisson distribution with a mean of 0.2, this probability will be 0.819 and the distribution of the error term will be heavily skewed.

The validity and usefulness of the least-squares estimation and resulting trip generation models may be severely limited when there are many zeroes in the observation. This situation arises when models are formulated by purpose or by mode. Another example is the case in which models are specified at the person level rather than at the household level. In these instances, the probability is much higher that no trip of a given type will be generated by a given behavioral unit. The effect of error truncation may become significant, and the quality of estimated model coefficients and test statistics may deteriorate. This problem exists in addition to the more obvious problem of producing negative values as predicted numbers of trips. The possible extent of this problem is discussed later by using simulation examples.

As a second example, it may not be possible to properly capture travel behavior with one linear model. Trip generation behavior may be a result of a two-stage decision process in which a decision is first made to make, or not to make, trips of a given type at all on a given day; then, given trips will be made, and the number of trips is chosen in the second stage. If this is the case, it is probable that the decision to make trips at all is governed by a different causal mechanism than is the choice of the number of trips. For example, consider the case in which a transit trip generation model is developed at the household level. The primary determinants of the first-stage decision may include household car ownership and the number of nondrivers, whereas the second-stage decision may be described as a function of the number of household members and number of workers.

In the analysis of this study, this possible two-stage decision mechanism is represented by a system of two models: a binary probit model that represents the decision to make a trip of a given type at all, and a linear regression model applied to the number of trips, given that trips are made. Formally, the model system can be presented as

$$A_i = \alpha' X_i + u_i$$

$$Y_i = 0 \quad \text{if } A_i \leq 0$$

$$Y_i = \beta' Z_i + v_i \quad \text{if } A_i > 0$$

where

A_i = latent variable underlying the binary choice,
Y_i = number of trips made,
α' and β' = coefficient vectors,
X_i and Z_i = vectors of explanatory variables,
u_i and v_i = normal random error terms, and u_i is assumed to have a unit variance.

The two vectors of explanatory variables, X_i and Z_i, may contain the same variables. The binary choice probability is given as

$$\Pr[Y_i > 0] = \Phi(-\alpha' X_i)$$

where the lefthand side is the probability that trips will be made at all, and Φ on the righthand side is the standard cumulative normal distribution function.

The number of trips, Y_i, is defined to be 0 if $A_i \leq 0$. Given that trips are made at all ($A_i > 0$), the expected number of trips is $\beta' Z_i$. The unconditional expected number of trips can be obtained as

$$E[Y_i] = E[Y_i \mid Y_i = 0]\,Pr[Y_i = 0] + E[Y_i \mid Y_i > 0]\,Pr[Y_i > 0]$$

$$= E[Y_i \mid Y_i > 0]\Pr[Y_i > 0]$$

$$= \beta' Z_i \Phi(-\alpha' X_i).$$

The model system can be estimated simultaneously using the maximum likelihood method. Use of this method, however, requires the development of a computer code to estimate the coefficients. Alternatively, the model system can be estimated equation by equation using easily available binary probit and linear regression codes. A problem arises when the error term of the probit trip choice model (u_i) and that of the linear trip generation model (v_i) are correlated. Possible biases in coefficient estimates are avoided in this study by introducing a correction term into the linear regression model. Further discussions of this method are given elsewhere (*3–7*).

LINEAR REGRESSION ON SIMULATED POISSON TRIP DATA

The question of whether the least-squares regression approach will produce adequate model coefficients and test statistics is addressed in this section. Simulated data sets are generated assuming that trip generation is a Poisson process, the values of the parameters used to generate the data are then estimated by least-squares regression, and the quality of the parameter estimates is examined.

Trip generation is simulated as follows. For each case simulated, the expected number of trips is assumed to be

$$m_i = \beta_0 + \beta_{1i} X_{1i} + \beta_{2i} X_{2i}$$

where

m_i = expected number of trips for case i,
X_{1i} and X_{2i} = independent variables, and
β_0, β_1, and β_2 = model parameters to be estimated later by least-squares regression.

In the simulation, X_{1i} and X_{2i} are assumed to be 0–1 binary variables. Therefore, each case has one of the following four possible expected values: β_0, $\beta_0 + \beta_1$, $\beta_0 + \beta_2$, $\beta_0 + \beta_1 + \beta_2$. The number of trips, Y_i, is simulated using the following Poisson probability:

$$\Pr[Y_i = n] = \exp(-m_i) m_i^n / n! \quad n = 0, 1, 2, \ldots$$

where m_i is the expected number of trips for case i as defined above.

In each simulation run, cases are evenly divided into four groups, each having fixed values of the X's (and therefore one of the above four expected values). Consequently, dependent variable values in the data set come from four Poisson distributions. One hundred cases are generated for each group, and least-squares regression is applied to the resulting 400 cases in each simulation run.

A total of 10 simulation runs are performed for each combination of parameter values. The results of the simulation experiments with ordinary least-squares estimation are summarized in Table 1 for the eight sets of parameter values examined in this study.

The simulation experiment offers evidence that least-squares

TABLE 1 ORDINARY LEAST-SQUARES ESTIMATES OF THE PARAMETERS OF SIMULATED POISSON TRIP GENERATION

	Theoretical Values	Simulation Results*	Estimated S.E.*	True S.E.
Mean Y	.550	.547		
Constant	.100	.109	.121	.040
β_1	.300	.285	.139	.083
β_2	.600	.593	.139	.056
R^2	.170	.165		
Mean Y	1.000	0.980		
Constant	.100	.126	.084	.062
β_1	.600	.553	.097	.136
β_2	1.200	1.160	.097	.091
R^2	.310	.310		
Mean Y	1.900	1.889		
Constant	.100	.115	.083	.072
β_1	1.200	1.204	.097	.165
β_2	2.400	2.385	.097	.133
R^2	.486	.498		
Mean Y	2.800	2.771		
Constant	.100	.098	.140	.064
β_1	1.800	1.801	.162	.188
β_2	3.600	3.547	.162	.126
R^2	.591	.604		
Mean Y	3.450	3.337		
Constant	3.000	3.217	.168	.115
β_1	.300	.052	.194	.122
β_2	.600	.179	.194	.242
R^2	.032	.007		
Mean Y	3.900	3.847		
Constant	3.000	2.956	.166	.154
β_1	.600	.481	.192	.282
β_2	1.200	1.303	.192	.389
R^2	.103	.123		
Mean Y	4.800	4.797		
Constant	3.000	3.016	.188	.169
β_1	1.200	1.266	.217	.163
β_2	2.400	2.297	.217	.151
R^2	.273	.270		
Mean Y	5.700	5.580		
Constant	3.000	3.183	.194	.088
β_1	1.800	1.260	.225	.113
β_2	3.600	3.541	.225	.184
R^2	.415	.414		

*Average of 10 Simulation runs.

Note: The parameter values (constant, $\beta 1$, $\beta 2$) used to simulate data are shown under "Theoretical Values", and the ordinary least squares estimates of the parameters are shown under "Simulation Results".

regression yields adequate estimates of trip generation parameters even when the actual generation process is not compatible with its assumptions. The regression method performs very well when the theoretical R^2 (defined as the ratio of the systematic variation of the mean number of trips to the total theoretical variance) is higher than 0.10. The only exception is the case with a large constant (3.0) combined with small slope coefficients ($\beta_1 = 0.3$, $\beta_2 = 0.6$), for which the theoretical R^2 is only 0.032. Other than this exceptional case, the least-squares estimates adequately account for variations in trip generation as indicated by the R^2 values that are close to the theoretically expected values and the parameter estimates whose averages accurately replicate the true values used to generate the simulated data.

The estimated standard errors of the coefficient estimates, however, do not accurately represent the true standard errors obtained by evaluating the standard deviations of coefficient estimates from 10 repeated simulation runs. To examine whether this is due to heteroscedasticity (a variation in the variance of random errors across cases), weighted least-squares estimation was performed using as the weight the inverse of the square root of the predicted number of trips obtained by ordinary least-squares estimation. This weight was theoretically derived from the fact that the variance of a Poisson-distributed random variable equals its expectation.

Weighted least-squares estimation offered some improvement in estimated standard errors, although this improvement was at the cost of significantly diminished accuracy of coefficient estimates. The divergence between the estimated and true coefficient values was so large that it was only appropriate to conclude that the weighted least-squares procedure was not suitable for trip generation analysis when the underlying processes are composite Poisson processes with relatively small means (ranging from 0.1 to 7 trips). Although the reason for the poor performance is still undetermined, the parameters of trip generation processes may be accurately estimated by ordinary least-squares regression when the systematic variation in the data is reasonably high (with an R^2 of, say, 0.1 or higher).

TWO-STAGE TRIP GENERATION MODELS

Data Set

In the remainder of this paper the adequacy of linear trip generation models is examined by applying an alternative model formulation to empirical data. The conventional linear regression models and two-stage models described earlier are estimated and their relative performance is studied. The intent of the effort is to infer the validity of conventional linear models and the value of more elaborate models. Note that, unlike the simulation analysis above, the true behavioral mechanism is not known in this empirical analysis. The validity of the alternative models is therefore evaluated in the study on the basis of its statistical fit.

The results of the 1980 Southeastern Michigan Transportation Authority survey are used in the estimation of two-stage trip generation models. This standard home interview survey file contains demographic and socioeconomic attributes of the household and its members and records of all trips made by each household member (5 years old and over) on the survey day, including trips made by nonmechanized modes. The person, rather than the household, is used as the unit of analysis in this study. All individuals at least 16 years of age are included in the study sample. This particular cutoff age is selected because individuals can qualify to be licensed to drive at this age and become active users of the automotive transportation system.

A wide range of variables is considered in the model development to best capture trip generation behavior using the two types of models. These variables include age, sex, occupation, car availability, household composition, life cycle stage, income, residence county, residence area type, and day of the week (Table 2). The age and sex of an individual are known to influence trip generation significantly (*8, 9*) and therefore are included in this analysis. In addition, detailed occupation categories are used in the model development with the anticipation that variations in lifestyles can be captured by them.

Past studies also indicate that household structure influences trip generation behavior even when the model is formulated at the individual level. For example, a study shows that various measures of individual mobility vary significantly and meaningfully across subgroups defined by life cycle stages (*10*). Household structure is represented in this study by the number of household members by age and sex and by a set of five life cycle stages as defined in the data file.

Because the models are formulated at the individual level, car availability, rather than car ownership, is used to explain trip generation. The following four levels of car availability are defined according to the license-holding status of the individual and car ownership of the household:

Always: the individual holds a driver's license, and the number of cars available to the household equals or exceeds the number of adults in the household;

Usually: the individual holds a driver's license and at least one car is available to the household, but the number of cars available is less than the number of adults in the household;

Sometimes: the individual does not hold a driver's license but at least one car is available to the household; and

Never: no car is available to the household.

Combined household income is classified in the data file into 11 categories. In the analysis these categories are combined into four income classes as shown in Table 2.

The land use type and density variables are introduced to account for the possibility that trip generation is influenced by the availability of opportunities around the home base. The residence county variables are introduced in the belief that differences in lifestyles that are not reflected in the household and person-attribute variables in the data file can be captured by these variables. Note, however, that the notion that trip generation depends on residence area contradicts the commonly held belief that trip generation of a household or individual of given characteristics is invariant across areas.

Estimation Results

The final model forms and estimation results are summarized in Table 3. The dependent variable is the total number of person trips generated by an adult household member. All regression models are estimated using weighted least squares with the weight defined as $\theta(|Y|)^\tau$ where θ and τ are estimated

TABLE 2 VARIABLES USED IN MODEL FORMULATION

VARIABLE	DEFINITION
Age and Sex	
AGE:16-30	1 if the age is between 16 and 30; 0 otherwise
AGE:31-50	1 if the age is between 31 and 50
AGE:51-64	1 if the age is between 51 and 64
AGE:65+	1 if the age is 65 or over
MALE	1 if male
FEMALE	1 if female
Occupation	
PRO/TECH	1 if professional or technical
FARM	1 if farmer, farm manager, farm laborer, or farm foreman
LABORER	1 if non-farm laborer
MANAGER	1 if manager, official, or owner of a business
CLERICAL	1 if clerical and similar worker
SALES	1 if sales worker
CRAFTSMAN	1 if craftsman, foreman, and similar worker
OPERATOR	1 if equipment operator or motor vehicle operator
HHLDWORKER	1 if private household worker, maid, butler, etc.
SERVICE	1 if service worker
MILITARY	1 if in military
OTHER	1 if other worker
Car Availability	
ALWAYS	1 if the individual has a driver's license and the number of cars is no less than the number of adults in the household
USUALLY	1 if the individual has a driver's license and the number of cars is less than the number of adults in the household
SOMETIMES	1 if the individual does not have a driver's license and the household has at least one car available
NEVER	1 if no car is available to the household
Household Structure	
NADULTS	Number of adults (≥ 18 years old) in the household
NCHILD:0-4	Number of children of 0 to 4 years old
NCHILD:5-15	Number of children of 5 to 15 years old
NCHILD:16-18	Number of children of 16 to 18 years old
NMALES	Number of males in the household
NFEMALES	Number of females in the household
Household Lifecycle Stage	
NOCHLD-YNG	1 if head of household less than 35 years of age, and no children in the household less than 18 years of age
NOCHLD-MID	1 if head of household 35 years of age or older, but less than 65 years of age, no children in the household
NOCHLD-OLD	1 if head of household 65 years of age or older, no children in the household less than 18 years of age
PRESCHOOL	1 if the youngest child in the household is less than 6 years of age, for head of household of any age
SCHOOLAGE	1 if the youngest child in the household is 6 years of age or older, for head of household of any age
Household Income	
LOW	1 if household annual income is less than $10,000
MID-LOW	1 if household annual income is between $10,000 and $20,999
MID-HIGH	1 if household annual income is between $21,000 and $34,999
HIGH	1 if household annual income is $35,000 or more
Residence County	
DETROIT	1 if residence zone is in Detroit
WAYNE	1 if residence zone is in Wayne County
OAKLAND	1 if residence zone is in Oakland County
MACOMB	1 if residence zone is in Macomb County
WASHTENAW	1 if residence zone is in Washtenaw County
MONROE	1 if residence zone is in Monroe County
STCLAIR	1 if residence zone is in St. Clair County
LIVINGSTON	1 if residence zone is in Livingston County
Residence Area Type	
COMMERCIAL	1 if 10 or more employees per acre of usable land
HIDENSITY	1 if less than 10 employees and more than 5 dwelling units per acre of usable land
MIDDENSITY	1 if less than 10 employees and from 0.5 to 5.0 dwelling units per acre of usable land
LOWDENSITY	1 if less than 10 employees and less than 0.5 dwelling units per acre of usable land
Day of Week	
MONDAY	1 if Monday
TUESDAY	1 if Tuesday
WEDNESDAY	1 if Wednesday
THURSDAY	1 if Thursday
FRIDAY	1 if Friday

TABLE 3 TWO MODELS OF TOTAL PERSON-TRIP GENERATION

	Conventional Linear Model (WLS)		Two-Stage Model System: Probit Trip Choice (ML)		Two-Stage Model System: Conditional Trip Generatn (WLS)	
	β	t	β	t	β	t
AGE:31-50	-.113	-1.33	-.302	-5.23		
AGE:51-64	-.608	-6.51	-.608	-10.24		
AGE:65+	-.820	-7.34	-.753	-10.68		
MALE			.326	6.80	-.411	-5.30
PRO/TECH	.661	5.78	.290	3.36		
LABORER					-.398	-2.68
MANAGER	.927	5.38	.453	3.29	.305	1.79
CLERICAL	.476	3.37	.391	3.71	-.232	-1.68
SALES	.381	2.25	.122	1.03		
CRAFTSMAN			.238	1.91	-.388	-2.66
SERVICE	.693	3.68	.469	3.34		
OTHER					-.385	-2.39
ALWAYS	.912	9.26	.206	3.92	.368	4.14
USUALLY	.439	4.58				
NEVER					-.392	-3.46
NADULTS			-.058	-1.84	.109	2.98
NCHILD:0-4			-.039	-.44		
NCHILD:5-15	.133	3.09			.151	3.82
NCHILD:16-18			.205	3.17	-.236	-2.75
NMALES			-.117	-2.67		
NFEMALES	.076	1.55				
NOCHLD-YNG			-.299	-5.15	.284	2.81
SCHOOLAGE	.332	3.77	.237	4.06		
LOW	-.248	-2.78	-.436	-6.63		
MID-LOW			-.117	-2.01		
HIGH			.124	1.65		
WAYNE	-.158	-1.10	-.215	-4.18		
OAKLAND	.349	2.49			.291	3.01
MACOMB	.203	1.28			.195	1.66
WASHTENAW	.601	3.14			.842	4.95
HIDENSITY	-.188	-1.97			-.352	-3.86
MIDDENSITY	.182	2.18			.114	1.37
MONDAY	-.203	-2.25			-.155	-1.62
TUESDAY			.166	3.09	-.154	-1.74
WEDNESDAY	-.148	-1.58				
FRIDAY	.098	1.00	.108	1.77	.053	.51
Correction Term*					.485	3.97
Constant	2.231		1.172		2.834	
R^2	.138				.079	
F (df)	34.07 (24,5109)				15.85 (21,3884)	
-2[L(β)-L(0)] (df)			2594.2 (24)			
-2[L(β)-L(C)] (df)			1040.0 (23)			
N	5134		5077		3906	

*Introduced to correct for possible biases due to the correlation between the error term of the probit choice model and that of the conditional trip generation model.

WLS: Weighted least squares regression
ML: Maximum likelihood estimation
df: degrees of freedom
L(0): Log-likelihood with all coefficients constrained to 0
L(c): Log-likelihood with the constant term alone
L(β): Log-likelihood with no constraints
N: Sample size

-2[L(β)-L(0)] and -2[L(β)-L(0)] have chi-square distributions with indicated degrees of freedom, respectively. The former can be used to test the collective significance of all model coefficients, and the latter to test the significance of the model coefficients excluding the constant term.

by regressing the squared residual on the predicted number of trips (unlike the simulation analysis above, practically no differences emerged in this case between the ordinary least-squares and weighted least-squares estimation results).

Columns 1 and 2 of Table 3 present the estimated model coefficients and *t*-statistics of the conventional linear regression model. The coefficients of the age variables show the well-established relationship that trip generation declines as age increases. The results also suggest that white-collar workers tend to make more trips, that the presence of school-age children increases the adult members' trip generation, and that low-income families make fewer trips. The car availability variables are highly significant, indicating that trip generation increases with car availability.

The two sets of variables that are not normally included in trip generation models, residence county and day of the week, are both significant. The day-of-the-week variables suggest that trip generation is suppressed on Mondays. The coefficient for Friday trip generation is positive, although insignificant. This finding is consistent with earlier results that trip generation increases toward the end of the week (*11*), but statistically is not as conclusive.

The set of residence county variables suggests that residents of suburban counties tend to make more trips. This area-specific effect is in addition to those represented by the income variables or by the land-use type variables, the latter of which indicate that residents in the area with 0.5 to 5 dwelling units per acre make more trips. Although it is not possible to pinpoint the reasons for the significance of the county variables, it is conceivable that these variables act as proxies for unobserved and geographically correlated factors such as ethnic backgrounds.

The probit trip choice model includes a set of variables that is similar to that of the conventional model. However, the effect of income variables is more pronounced, whereas car availability variables are less dominant in the probit choice model. The sex variable is significant in the probit model and indicates that a man makes trips on any given day more frequently than does an equally situated woman. Importantly, no land use type variable is present and only one residence county variable is included in the model. This result suggests that the choice of whether to make trips at all does not vary substantially by geographical area.

Columns 7 and 8 of Table 3 show the conditional trip generation model that accompanies the probit trip choice model. Quite notable is the result that the age variables and income variables, both significant in the probit choice model, are excluded from the conditional trip generation model because of their insignificance. On the other hand, the land use type variables, which are not included in the probit model, are included in the conditional trip generation model.

It is also notable that when a variable is included in both models, its coefficient values tend to contradict each other. For example, the sex variable (MALE) is positive and significant in the probit model, but negative and significant in the conditional trip generation model. These values imply that women have a higher probability of not making any trips on a given day, but given that they make trips at all, they tend to make more trips than men.

The two-stage model system thus offers indications that the choice of making trips at all and the determination of the number of trips are influenced by overlapping but different sets of factors. New behavioral insights are offered by the model system. However, the indications are not so clear-cut as to reject the conventional linear model as an inferior formulation. In fact, the coefficients of the linear model are consistent with those of the two equations in the two-stage model system when viewed collectively. The same conclusions have been obtained from a similar analysis of shopping trips.

TABLE 4 PREDICTION RESULTS

Model	Total Trips	Shopping Trips
Linear	0.1297	0.0276
Two-Stage	0.1253	0.0268

The explanatory powers of the two alternative model formulations are evaluated by examining the correlation between the observation and prediction. The correlation is estimated by regressing the observed number of trips on the predicted number of trips. The results, summarized in Table 4 in terms of R^2, indicate that the two formulations have virtually identical fits to the observation, with the two-stage models showing slightly inferior fits for both total person trips and shopping trips. The conventional linear regression models are capable of accounting for as many variations in trip generation as are the more elaborate two-stage model systems.

CONCLUSION

The adequacy of conventional linear regression models in trip generation analysis has been the subject of this study. The following two issues have been addressed as possible factors that may invalidate linear regression analysis: (a) the incompatibility between the continuous, untruncated error term of a linear regression model and the discrete and nonnegative number of trips generated by a household or individual and (b) the possibility of a two-stage decision mechanism in which the choice of making trips at all is first made, and then the number of trips is determined given that trips are made.

Simulation experiments were conducted to address the first issue. In the simulation, trips were generated assuming Poisson distributions. Although the resulting error distributions were heavily truncated, the analysis indicated that model parameters can be consistently estimated and the expected number of trips can be forecast accurately by using the linear model and ordinary least-squares estimation method. The estimated standard errors of model coefficients were biased. The analysis indicated that weighted least-squares could not be applied to the simulated data to solve this problem because of the inaccurate coefficient estimates that the method produced. Further research is needed to identify the reason for the poor performance of weighted least-squares regression.

Two-stage model systems were estimated by using an empirical data set and then compared with linear regression models. The results indicated that the choice of making trips at all and the determination of the number of trips to make are influenced by overlapping, but different, sets of factors. However, the linear regression models offered essentially the same characterization of trip generation behavior as the two-stage models. Furthermore, the explanatory powers of the two alternative model formulations were found to be identical. The two-stage models provided some additional behavioral insights, but failed to show any improvement in fit despite their complex model structure, which involves an increased number of parameters and an elaborate estimation procedure.

The results of this study have indicated that linear regression models of trip generation offer consistent coefficient estimates and produce as accurate predictions as a more complex two-stage model system. The ordinary least-squares estimation is appropriate for generation models of infrequent trips for which the assumptions underlying the estimation method are unlikely to hold. Improvement in trip generation analysis may not be obtained by adopting more complex model systems.

REFERENCES

1. G. V. F. Timmermans. *A Comparative Analysis of Disaggregate Trip Generation Models*. Master's thesis. Northwestern University, Evanston, Ill., 1982.
2. S. R. Lerman and S. L. Gonzalez. Poisson Regression Analysis Under Alternate Sampling Strategies. *Transportation Science*, Vol. 14, 1980, pp. 346–364.
3. J. J. Heckman. Sample Selection Bias as a Specification Error. *Econometrica*, Vol. 47, 1979, pp. 153–161.
4. G. S. Maddala. *Limited-Dependent and Qualitative Variables in Econometrics*. Cambridge University Press, Cambridge, England, 1983.
5. R. Kitamura and P. H. L. Bovy. Analysis of Attrition Biases and Trip Reporting Errors for Panel Data. *Transportation Research A*, Vol. 21A, 1987, pp. 287–302.
6. D. A. Hensher, P. O. Bernard, N. C. Smith, and F. W. Milthrope. Modelling the Dynamics of Car Ownership and Use: A Methodological and Empirical Synthesis. Presented at the 5th International Conference on Travel Behavior, Aix-en-Provence, France, 1987.
7. F. L. Mannering. Selectivity Bias in Models of Discrete and Continuous Choice: An Empirical Analysis. *Transportation Research Record 1085*, TRB, National Research Council, Washington, D.C., 1985, pp. 58–62.
8. H. S. Levinson. Urban Travel Characteristics. In *Transportation and Traffic Engineering Handbook* (J. E. Baerwald, ed.), Prentice Hall, Englewood Cliffs, N.J., pp. 138–206.
9. R. Kitamura. Lifestyle and Travel Demand. In *Special Report 220: A Look Ahead: Year 2020*, TRB, National Research Council, Washington, D.C., 1988, pp. 149–189.
10. L. P. Kostyniuk and R. Kitamura. Household Lifecycle: Predictor of Travel Expenditure. In *Behavioural Research for Transport Policy*, VNU Science Press, Utrecht, The Netherlands, 1986, pp. 343–362.
11. R. Kitamura and T. van der Hoorn. Regularity and Irreversibility of Weekly Travel Behavior. *Transportation*, Vol. 14, 1987, pp. 227–251.