Formulation of Trip Generation Models Using Panel Data

RYUICHI KITAMURA

This study addresses the question of whether conventional models based on cross-sectional data alone account for trip generation adequately. Alternative model formulations are examined using a panel data set to determine whether elements associated with past time points should be considered in trip generation analysis and, if so, which elements should be introduced into the model. The analysis shows that estimated model coefficients and *t*-statistics differ substantially depending on model specification, and that allowing for serial correlation and incorporating a lagged dependent variable both significantly improve the model's fit. The results indicate that serial correlation should be incorporated whenever feasible for efficient estimation and improved fit and that it is safer to ignore state dependence (dependence of observed trip generation on that from a previous time point) and incorporate serial correlation, than to ignore serial correlation and incorporate state dependence. This significance of serial correlation, which presumably is due to omitted variables that are longitudinally correlated, suggests that important determinants of trip generation lie outside the set of variables that have traditionally been considered in travel behavior analysis.

Suppose a panel data set consisting of observations of the same behavioral units at several points in time is available for developing models of travel behavior. The modeling effort in this case will be more complex than with a cross-sectional data set because of the wider range of variables available—that is, measurements of behavior and contributing factors from previous observation periods. In addition, one must select a model formulation that best represents the behavior from several classes of models that can be developed with panel data.

Choosing a particular model formulation quite often implies adopting a particular behavioral hypothesis. One focus of model specification effort using panel data is how to represent dynamic characteristics of observed behavior. Two major hypotheses are (1) that behavior is static (or contemporaneous): namely, behavior at time t can be explained by factors observed at time t; and (2) that behavior is dynamic and cannot be fully explained by contemporaneously observed factors. If the former is the case, cross-sectional data suffice. If the latter is the case one may hypothesize further: (2a) behavior can be described adequately by introducing explanatory variables from past observation points into the model; (2b) past measurements of the behavior itself must be introduced into the model; and (2c) serial correlation in unobserved elements needs to be taken into account.

As is often the case in modeling travel behavior, no theory exists that dictates a priori which model formulation should be used. The selection of a model formulation is empirical, at least until such a theory can be constructed. In this context, the implications of recent observations of dynamic aspects of travel behavior, such as response lags, habit persistence, and history dependence, are important (*1–4*). These results suggest that dynamic models represent travel behavior more accurately and meaningfully. Then how best can the dynamism be captured in a quantitative model of travel behavior?

This study addresses the question of whether incorporating elements observed in past time points is worthy in trip generation analysis and, if so, which elements should be introduced into the model. The models examined include conventional cross-sectional model, lagged dependent variable model (which is formulated using a dependent variable from a previous time point as an explanatory variable), lagged independent variable model, and cross-sectional and lagged dependent variable models with serially correlated errors. The objective of this study is to examine these alternative model formulations and to determine the most relevant representation of trip generation behavior.

The weekly total numbers of person trips, travel time expenditures, and number of social-recreation trips, as reported by household members, are used as the dependent variables of this study. Models for these variables should ideally be formulated within a simultaneous equations framework together with car ownership models. This is not done in this study. Rather, this analysis is conducted to determine the model formulation to be used in such a modeling effort, which is to follow the present study as an extension of the model system reported elsewhere (*3, 5, 6*). This study considers only linear models.

The rest of this paper is organized as follows. The alternative model formulations considered are described in the next section. Then the data set and estimation procedures used in the study are described briefly, and the results of a preliminary analysis of trip generation using a cross-sectional model are presented. The relative importance of lagged independent variables and lagged dependent variables is next evaluated. Following this, the com-

Department of Civil Engineering, University of California at Davis, Davis, Calif. 95616.

peting hypotheses of serial correlation and state dependence are examined. The latter hypothesis underlies the use of the lagged dependent variable model. A summary and conclusions of the study are last.

ALTERNATIVE MODEL FORMULATIONS

This section describes the alternative model formulations considered in this study. The *cross-sectional model* has the form

$$Y(t) = \beta' X(t) + \varepsilon(t) \quad t = 1, \ldots, T \tag{1}$$

where $Y(t)$ is a scalar measure of observed behavior at time t, β is a vector of coefficients, $X(t)$ is a vector of explanatory variables at time t that are uncorrelated with $\varepsilon(t)$, $\varepsilon(t)$ is a random error term, and T is the number of equi-spaced observation points. In this formulation, observed behavior at time t is related to contributing factors also observed at t. Obviously, this model can be estimated using cross-sectional data. The assumption underlying the formulation is that Y changes immediately in response to a change in X and that the value of Y does not depend on the past history of either X or Y itself.

Alternatively, the behavioral relation of interest may be formulated as

$$\begin{aligned} Y(t) = \beta' X(t) + \tau' X(t-1) + \cdots \\ + \mu' X(t-s) + \varepsilon(t) \\ t = s+1, \ldots, T \end{aligned} \tag{2}$$

where β, τ, and μ are coefficient vectors and s is a positive integer. In this formulation behavior is assumed to be history-dependent; that is, behavior at time t is assumed to be a function of contributing variables measured at $t - s$ through t. A change in X is not fully reflected in Y immediately; Y will change gradually over time and reaches a new equilibrium after s periods, provided that no further change takes place in X. Accordingly, the model depicts behavior with response lags. This model is called the *lagged independent variables model* in this study.

Another possible formulation is

$$Y(t) = \theta Y(t-1) + \beta' X(t) + \varepsilon(t) \quad t = 2, \ldots, T \tag{3}$$

where θ is a scalar coefficient and $|\theta| < 1$. This model is known as the *lagged dependent variable model* and is often referred to as the *dynamic model.*

The assumption that $Y(t)$ depends on $Y(t - 1)$ implies that $Y(t)$ is a function of the entire history of X and ε. Applying the above recursive relation repeatedly, the following is obtained;

$$\begin{aligned} Y(t) = \beta' X(t) + \theta\beta' X(t-1) + \theta^2\beta' X(t-2) + \cdots \\ + \varepsilon(t) + \theta\varepsilon(t-1) + \theta^2\varepsilon(t-2) + \cdots \end{aligned} \tag{4}$$

Accordingly, the lagged dependent variable model exhibits behavior similar to that of the lagged independent variables model. The former model is advantageous when only a short panel (with small T) is available; if the specification of Equation 3 is correct the model can be estimated with observations from as few as two time points, even though $Y(t)$ depends on the entire history. The model of Equation 4 is a special case of the model of Equation 2 with $s = -\infty$, except for the difference in the error structure.

The lagged dependent and lagged independent models are both special cases of the general formulation

$$\begin{aligned} Y(t) - \theta_1 Y(t-1) - \theta_2 Y(t-2) - \cdots \\ = \beta' X(t) + \beta_1' X(t-1) + \beta_2' X(t-2) + \cdots + \varepsilon(t) \end{aligned} \tag{5}$$

which is referred to as a transfer function model in time-series analysis (*7*). Discussions of this class of models can be found, for example, in Griliches (*8*).

The error term of each model may be assumed to be serially correlated; that is,

$$\varepsilon(t) = r\varepsilon(t-1) + u(t) \tag{6}$$

where r is the coefficient of serial correlation and $u(t)$ is independent of ε. This is a convenient scheme to adopt when unobserved variables do not change their values frequently and are longitudinally correlated.

DATA SET AND INITIAL EXPLORATION

The observations from waves 1, 3, 5, and 7 of the Dutch National Mobility Panel survey are used in this study (these waves will be referred to hereafter as periods 1, 2, 3, and 4, respectively). The four surveys were conducted in the spring of 1984, 1985, 1986, and 1987, respectively. In each survey a weekly travel diary was collected from each member of the sample household who was at least 12 years old. The aim, format, and characteristics of this panel data set are discussed elsewhere (*9–12*).

The dependent variables of the analysis in this study are weekly measures of trip generation and travel time expenditure. These variables are used without any transformation because a weekly total is likely to lead to residuals with desirable properties because of the law of large numbers. Previous model estimation results using the same data set and weekly measurements agree with this expectation.

Heteroscedasticity is accounted for by applying the weight,

$$W_i = a(1 + |\hat{Y}_i|)^b \tag{7}$$

where i refers to the household. Parameters a and b are obtained by regressing the squared residual from ordinary least squares (OLS) estimation on the OLS prediction. This weight has been selected after examining several other formulations. Models that assume serial correlation are estimated by regressing $Y(t) - rY(t - 1)$ on $X(t) - rX(t - 1)$ and iterating on r. A convergence is obtained quickly, usually within five iterations, with the relative change in r being reduced to less than 1 percent of its absolute value.

A set of explanatory variables was selected after an examination of a wide range of variables, including the number of household members (≥12 years old) who kept a travel diary, number of workers, number of drivers, number of children by age group, household income, car ownership, education, a rough indicator of transit service levels, and their transformations. The variables that appear in the models presented in this paper are defined in Table 1.

Table 2 shows weighted least squares (WLS) estimates of the coefficients of a cross-sectional total person trip generation model obtained from wave-5 observations ($t = 3$). The sample of this model estimation consists of those 829 households that participated in waves 1, 3, and 5 surveys. Note that some variables are multiplied by the number of diary keepers (NDIARIES) so that their coefficients represent the difference in trip generation per diary keeper.

The most dominant factor is the number of diary keepers in the household (NDIARIES). The estimated coefficient value indicates that a diary keeper on average will add approximately 20 trips to weekly household trip generation. The number of added trips will on average be 3.71 trips fewer if this person is not a driver and 1.85 trips more if the person is employed. The three variables representing the number of children by age group are all significant and indicate the increased household travel needs resulting from the presence of children.

The education level of a household is defined in this study as that of the person who has the highest education in the household. The two dummy variables used in the model to represent household education are both significant and show that the number of reported person trips

TABLE 2 CROSS-SECTIONAL MODEL OF WEEKLY HOUSEHOLD PERSON TRIP GENERATION (WLS, $t = 3$)

R^2	.586	
F	89.2	
df	(13,815)	
	β	t
NDIARIES	20.150	9.63
NONDRIVERS	-3.714	-3.46
NWORKERS	1.853	1.86
CHLD0-6	3.537	3.48
CHLD7-11	3.588	3.35
CHLD12-17	5.796	3.71
HIEDUC*NDIARIES	1.275	2.08
LOEDUC*NDIARIES	-2.646	-3.37
√INCOME*NDIARIES	-.026	-.12
ONECAR*NDIARIES	.861	.94
TWOCAR*NDIARIES	-.074	-.06
BOV-HIGH*NDIARIES	-.940	-1.32
FOR-LOW*NDIARIES	-1.125	-1.06
Constant	3.006	
N	829	

TABLE 1 DEFINITION OF THE VARIABLES USED IN THE MODELS OF THIS STUDY

Variable	Definition
NTRIPS	Total no. of person trips reported by household members during the survey week
NTRNTRIPS	Total no. of transit trip segments reported by household members during the survey week
NDIARIES	No. of household members (≥ 12 years old) who filled out diaries
NDRIVERS	No. of licensed drivers in household
NONDRIVERS	No. of individuals of at least 12 years old who did not hold a driver's license
NWORKERS	No. of employed household members
CHLD0-6	No. of children between 0 and 6 years old
CHLD7-11	No. of children between 7 and 11 years old
CHLD12-17	No. of children between 12 and 17 years old
√INCOME	Square-root of annual household income divided by 100
LOEDUC	1 if the household member with the highest education had completed only primary school; 0 otherwise
HIEDUC	If at least one person in household has a college degree; 0 otherwise
ONECAR	1 if the household owns one car; 0 otherwise
TWOCAR	1 if the household owns two or more cars; 0 otherwise
BOV-HIGH	1 if the household resides in a metropolitan area with highly developed transit systems; 0 otherwise
BOS-MEDIUM	1 if the household resides in a community which is served by rail; 0 otherwise
FOR-LOW	1 if the household resides in a community which is not served by rail; 0 otherwise

increases with household education. The result, however, may be a reflection of reporting errors as well as of genuine variations in trip making. An earlier analysis of panel attrition using the same data set (*11*) suggests that households with lower education tended to underreport their trip making and tended to drop out of the panel. It is likely that the coefficients of these education variables represent in part the magnitude of underreporting.

Household income (√ INCOME) and transit service level indicators (BOV-HIGH and FOR-LOW) are all insignificant. Car ownership can be considered to be more closely associated with household mobility than is income itself, because it reflects long-term mobility choice. The estimation result shown in Table 2 indicates that car ownership, which is represented by two dummy variables (ONECAR, TWOCAR), is insignificant. This result is consistent with the earlier indication (*5*) that household person trip generation in the Dutch Panel data set is statistically independent of household car ownership. This is due in part to the inclusion of walking and bicycle trips. Nonetheless, its contradiction with the commonly held belief that car ownership is a major determinant of person trip generation is noteworthy.

In summary, the analysis of this section indicates that trip generation as depicted by this cross-sectional model is primarily a function of demographic characteristics of the household. Car ownership and transit service levels have little influence on the total number of person trips generated by household members over the period of a week.

SIGNIFICANCE OF LAGGED VARIABLES

The model development effort has considered a wide range variables and their transformations that extend beyond those typically used in trip generation models. The resulting model has a relatively high R^2 value of 0.586. Nevertheless, a question remains whether the set of variables used in the model adequately captures systematic variations in trip generation, and whether all pertinent elements are included in the model. This section is concerned with the possibility that observed variables from past time points, or lagged variables, influence trip generation.

If response lags, habit persistence, history dependence, and other longitudinal relationships significantly affect trip generation behavior, cross-sectional models such as the one discussed in the previous section would be a misspecification of the behavioral relationship. Response lags can be represented by introducing lagged independent variables—that is, explanatory variables observed at previous time points—into the model. Habit persistence and history dependence may be captured by the use of lagged dependent variables. The relative contributions of lagged dependent variables and lagged independent variables to the model's explanatory power are the focus of this section.

The main questions of this section are whether lagged variables significantly contribute to the model's explanatory power and, if so, whether a lagged dependent variable contributes more than do lagged independent variables. Suppose that the effect of a lagged dependent variable is due to the effect of independent variables from previous time points reflected in the lagged dependent variable. A special case of this is shown by Equation 4 derived from Equation 3 through Koych transformation. If this in fact is the case, then the fit of a model with lagged independent variables, but without a lagged dependent variable, should be approximately equal to that of a lagged dependent variable model. This question is examined by estimating both lagged independent and lagged dependent variable models.

The same set of 829 households that are in the data files for periods 1 through 3 is used to estimate models for period 3 trip generation ($Y(3)$), and the set of 645 households that are in the data files of periods 1 through 4 is used to estimate models of $Y(4)$. The results of WLS estimation are summarized in Table 3.

The same set of explanatory variables as in the cross-sectional model of Table 2, but from a previous period, is sequentially added to the cross-sectional model. The R^2 values shown in the table clearly indicate that the introduction of the X vector from each previous period does contribute to the model's goodness-of-fit. It is also clear, however, that the marginal improvement in R^2 is small and tends to decline as more sets are added. Based on the pseudo F-statistics presented in the table, the lagged independent variable vectors are mostly not significant (at $\alpha = 0.05$); the only exception is $X(t-1)$ for $t = 3$.

The goodness-of-fit of lagged dependent variable models is presented at the bottom of Table 3. As noted earlier, the models are estimated through the WLS method iterated on r. The difference in the goodness-of-fit is substantial between the models with lagged X vectors and those with a lagged Y. For example, the lagged independent variable model estimated for the third period ($t = 3$) with $X(1)$ through $X(3)$ has an R^2 of 0.605, while the lagged dependent variable model with $X(3)$ and $Y(2)$ has an R^2 of 0.770. Similar differences between the two sets of models can be observed for $t = 4$. A further inspection of the estimation results indicated that the lagged Y is the most significant explanatory variable of the lagged dependent variable models estimated here. The results of this analysis offer an indication that the improvement in the model's fit realized by a lagged dependent variable is not merely a reflection of the effects of independent variables from previous periods.

This point is further examined by repeating the analysis for weekly total travel time expenditure and number of social-recreation trips using data from the third period ($t = 3$). The same estimation procedure was applied to these dependent variables. The results summarized in Table 4 agree with those obtained for total person trip generation; the X vectors from past periods are generally insignificant, and lagged dependent variables account for much larger portions of the total variations than do lagged independent variables.

Importantly, the difference in the goodness-of-fit between the lagged independent model and lagged dependent model is more pronounced with social trip generation. The

TABLE 3 IMPROVEMENT IN THE MODEL'S FIT DUE TO LAGGED INDEPENDENT AND LAGGED DEPENDENT VARIABLES IN TOTAL TRIP GENERATION MODELS

Explanatory Variables		t = 3	t = 4
X(t)	R^2	.586	.613
	F	89.2	77.4
	df	(13,815)	(13,631)
X(t), X(t-1)	R^2	.599	.622
	F	46.4	39.4
	df	(26,802)	(26,618)
Significance of X(t-1)	F	2.09	1.14
	df	(13,802)	(13,618)
X(t), X(t-1), X(t-2)	R^2	.605	.633
	F	31.1	26.9
	df	(39,789)	(39,605)
Significance of X(t-2)	F	.83	1.24
	df	(13,789)	(13,605)
X(t), X(t-1), X(t-2) X(t-3)	R^2		.643
	F		20.7
	df		(52,592)
Significance of X(t-3)	F		1.28
	df		(13,592)
Y(t-1), X(t)	R^2	.770	.776
	F	191.93	160.42
	df	(14,814)	(14,630)
Significance of Y(t-1)	t	-6.5	-9.8
Serial Correlation	r	.780	.805
N		829	645

Note: Each X vector contains the 13 explanatory variables shown in Table 2. All models are estimated using the weighted least squares (WLS) method with the weight shown in Eqn (7). To enable comparison across the models, the R^2 and F values shown have been recomputed using predictions by the respective models. The F values are therefore approximate and do not exactly follow F distributions. The prediction by the lagged dependent variable models was obtained using the residual from the previous period.

lagged independent variable model with $X(1)$ through $X(3)$ but without a lagged Y has a small R^2 of 0.299. The lagged dependent variable model with $X(3)$ and $Y(2)$, on the other hand, has an R^2 of 0.487. The difference in the R^2 values is extremely large, with the lagged dependent variable model accounting for 63 percent more variation than does the lagged independent variable model. The corresponding percentages are only 27 percent for total trip generation (t = 3) and 33 percent for travel time expenditure. The relative explanatory power of lagged dependent variables varies from variable to variable.

The social-recreation trip generation models, for which the effect of the lagged dependent variable is most substantial, have much smaller R^2's. If this observation can be generalized, the relative contribution of a lagged dependent variable increases when the other explanatory variables do not capture much of the variation in the dependent variable. In other words, as the fraction of unaccounted variations increases, so does the contribution of the lagged dependent variable, presumably because this variable reflects idiosyncrasy.

The findings thus far obtained offer valuable insight, but are subject to a limitation that arises from the nature of the variables used. Some of the independent variables in this analysis are longitudinally multicollinear because their values change only infrequently. Consequently, vectors of lagged independent variables do not offer much additional information and therefore do not substantially improve the model's fit. Unfortunately, it is not possible to determine, on the basis of the available data, whether this in fact is the major reason for the significance of the lagged dependent variables and the insignificance of the lagged independent variables, as evidenced in Tables 3 and 4. For this reason, lagged independent variable models are excluded from consideration in the analyses presented in the rest of this paper.

TABLE 4 IMPROVEMENT IN THE MODEL'S FIT DUE TO LAGGED INDEPENDENT AND LAGGED DEPENDENT VARIABLES IN TOTAL TRIP TIME EXPENDITURE AND SOCIAL-RECREATIONAL TRIP GENERATION MODELS

Explanatory Variables		Travel Time	Social Recreation
X(t)	R^2	.424	.283
	F	50.1	36.0
	df	(12,816)	(9,819)
X(t), X(t-1)	R^2	.441	.293
	F	26.6	18.7
	df	(24,804)	(18,810)
Significance of X(t-1)	F	2.15	1.28
	df	(12,804)	(9,810)
X(t), X(t-1), X(t-2)	R^2	.448	.299
	F	17.9	12.7
	df	(36,792)	(27,801)
Significance of X(t-2)	F	.73	.74
	df	(12,792)	(9,801)
Y(t-1), X(t)	R^2	.598	.487
	F	93.63	78.032
	df	(13,815)	(10,818)
Significance of Y(t-1)	t	-6.8	-9.3
Serial Correlation	r	.676	.697
N		829	829

STATE DEPENDENCE VERSUS SERIAL CORRELATION

The analysis of this section focuses on the possibility that the significance of the lagged dependent variables shown in the previous section is due to omitted variables that are longitudinally correlated. A competing hypothesis is that observed travel behavior is dependent on the past behavior itself and a lagged dependent variable is its true determinant. If the effect of longitudinally correlated, omitted variables can be represented by serially correlated errors, then these two hypotheses constitute a panel analysis version of serial correlation versus "state dependence."

To address the question, the following two additional models for total person trip generation are estimated: cross-sectional model with serially correlated errors (Model 2) and lagged dependent variable model assuming no serial correlation (Model 3). These models are summarized in Table 5 together with the cross-sectional model with no serial correlation (Model 1), which was given in Table 2, and the lagged dependent variable model with serially correlated errors (Model 4) whose summary statistics were given in Table 3.

It is evident from Table 5 that coefficient estimates and estimated *t*-statistics vary substantially depending on the model specification. For example, the coefficient of number of workers (NWORKERS) is positive in both cross-sectional and lagged dependent variable models without serial correlation (Models 1 and 3), while it is negative in the models where serial correlation is assumed (Models 2 and 4). The coefficient of number of children between 7 and 11 years old (CHLD7-11) is not significant, while the coefficient of income ($\sqrt{\text{INCOME}}$) is positive and more significant in Models 2 and 4 with serial correlation. The significance of the household education variables (HIEDUC and LOEDUC) differs drastically between the two cross-sectional models (Models 1 and 2).

In general the coefficients of the lagged dependent variable model without serial correlation (Model 3) are smaller in their absolute values than those of the cross-sectional model without serial correlation (Model 1). Such regularity, however, cannot be found between the two lagged dependent variable models (Models 2 and 4). Many of the explanatory variables in the cross-sectional model (Model 1) become insignificant when a lagged independent variable is introduced (Model 3), but this is not observed between Models 2 and 4 that assume serial correlation.

Most importantly, the fraction of variance explained is virtually the same between the two models with serial correlation with R^2's of 0.763 (Model 2) and 0.770 (Model 4). The coefficient estimates also show similarity. The R^2 of Model 3, on the other hand, shows that the lagged dependent variable by itself does not account for as much variation as does the serially correlated error (Model 2). The results suggest that trip generation is not state dependent, and the apparent significance of the lagged dependent variable is due to serially correlated errors, which in turn are presumably due to longitudinally correlated, omitted variables.

TABLE 5 COMPARISON OF DIFFERENT MODEL FORMULATIONS FOR TOTAL NUMBER OF PERSON TRIPS PER WEEK

	Cross-Sectional				Lagged Dependent			
	Model 1 (r = 0)		Model 2 (r > 0)		Model 3 (r = 0)		Model 4 (r > 0)	
R^2	.586		.763		.718		.770	
F	89.2		189.0		150.7		182.05	
df	(13,815)		(14,814)		(14,814)		(15,813)	
	β	t	β	t	β	t	β	t
r			.695				.780	
NTRIPS(t-1)					.525	19.41	-.171	-6.49
NDIARIES(t)	20.150	9.63	23.470	11.68	9.137	5.08	24.536	11.95
NONDRIVERS(t)	-3.714	-3.46	-2.316	-4.04	-2.574	-2.87	-1.823	-3.25
NWORKERS(t)	1.853	1.86	-2.428	-2.22	.254	.31	-3.139	-2.87
CHLD0-6(t)	3.537	3.48	1.961	1.19	2.342	2.77	1.529	.85
CHLD7-11(t)	3.588	3.35	.219	.13	1.696	1.91	-.152	-.08
CHLD12-17(t)	5.796	3.71	2.216	1.31	4.601	3.62	1.883	1.09
HIEDUC*NDIARIES(t)	1.275	2.08	-.471	-.59	.159	.32	-.685	-.81
LOEDUC*NDIARIES(t)	-2.646	-3.37	-.172	-.20	-.854	-1.30	.086	.10
√INCOME*NDIARIES(t)	-.026	-.12	.277	1.59	.132	.77	.271	1.65
ONECAR*NDIARIES(t)	.861	.94	.534	.72	.617	.81	.791	1.12
TWOCAR*NDIARIES(t)	-.074	-.06	-.690	-1.23	-.197	-.21	-.462	-.88
GOV-HIGH*NDIARIES(t)	-.940	-1.32	-2.408	-1.73	-.155	-.27	-3.470	-2.14
FOR-LOW*NDIARIES(t)	-1.125	-1.06	-1.868	-.90	-.736	-.84	-2.335	-.96
Constant	3.006		.436		.583		1.314	
N	829		829		829		829	

Note: In order to facilitate comparison across the models, the R^2 and F values have been reevaluated by computing predicted values of Y using WLS coefficient estimates. Residuals from t = 2 are used to compute predicted values of the models with serially correlated errors.

This point is further examined by estimating another model of the form,

$$Y(t) = \theta Y(t-1) + \beta_0' X(t) + \beta_1' X(t-1) + u(t) \tag{8}$$

which is shown below to be equivalent to Model 2 when certain conditions are met. Suppose Model 2 of Table 5 represents the true relationship; that is,

$$Y(t) = \beta' X(t) + \varepsilon(t) \qquad \varepsilon(t) = r\varepsilon(t-1) + u(t) \tag{9}$$

where r is the coefficient of serial correlation. Then

$$Y(t) = rY(t-1) + \beta' X(t) - r\beta' X(t-1) + u(t) \tag{10}$$

Therefore, the coefficient vectors of Equations 8 and 10 are related as

$$r = \theta, \quad \beta_0 = \beta, \quad \text{and} \quad \beta_1 = -r\beta_0 = -\theta\beta_0 \tag{11a–c}$$

if Model 2, or Equation 9, holds true. Hendry and Mizon (*13*) suggest the use of Equation 11c to discriminate between serial correlation and state dependence (also see *14*). The results of estimation are summarized in Table 6.

The coefficient of the lagged dependent variable (θ) of Equation 8 is 0.649, which is very close to the estimated serial correlation coefficient (r) of 0.695 (Model 2). The R^2 values vary only slightly between the two models (0.763 and 0.771). The results again favor the conclusion that trip generation is not state dependent and that the apparent state dependency is due to serially correlated errors. Nevertheless, it is important to note that some elements of β are substantially different from the corresponding elements of β_0, and not all elements of $\beta_1 + r\beta$ and $\beta_1 + \theta\beta_0$ diminish, which should be the case if Equations 11a through 11c hold. In addition, the lagged dependent variable of Model 4 is significant, as the t-statistic shown in the table indicates.

In conclusion, the analysis of this section has shown that the model's explanatory power increases substantially by incorporating serially correlated errors. It also has been shown that a lagged dependent variable also adds to the model's explanatory power but to a much lesser extent. The analysis has indicated that trip generation is both serially correlated and state dependent, but that the former plays a more dominant role.

A model with serial correlation should be used whenever possible. This can be concluded from the discrepancy in

TABLE 6 COMPARISON OF THE COEFFICIENT ESTIMATES OF TWO FORMULATIONS OF CROSS-SECTIONAL MODEL WITH SERIALLY CORRELATED ERRORS

	Model 2	Eqn (8)					
R^2	.763	.771					
F	189.0	106.5					
d.f.	(14,814)	(26,802)					
	β	β_0	β_1	$-\beta\mu$	$\beta_1+\mu\beta$	$-\theta\beta_0$	$\beta_1+\theta\beta_0$
NDIARIES	23.470	22.50	-17.00	-16.31	-.69	-15.24	-1.76
NONDRIVERS	-2.316	-1.47	.24	1.61	-1.37	1.50	-1.27
NWORKERS	-2.428	-3.45	4.44	1.69	2.75	1.58	2.87
CHLD0-6	1.961	-.70	2.66	-1.36	4.02	-1.27	3.93
CHLD7-11	.219	-3.47	4.33	-.15	4.49	-.14	4.48
CHLD12-17	2.216	-.79	1.58	-1.54	3.12	-1.44	3.01
HIEDUC*NDIARIES	-.471	-1.10	1.09	.33	.77	.31	.79
LOEDUC*NDIARIES	-.172	-.43	-.28	.12	-.40	.11	-.39
√INCOME*NDIARIES	.277	.25	-.24	-.19	-.05	-.18	-.06
ONECAR*NDIARIES	.534	2.03	-.61	-.37	-.24	-.35	-.26
TWOCAR*NDIARIES	-.690	1.26	-.89	.48	-1.37	.45	-1.34
BOV-HIGH*NDIARIES	-2.408	-4.61	4.18	1.67	2.51	1.56	2.62
FOR-LOW*NDIARIES	-1.868	-1.03	-.16	1.30	-1.46	1.21	-1.37
NTRIPS(t-1) (θ)		.649					
r	.695						

the coefficient estimates as well as difference in the R^2 values between Model 1 and Model 2 of Table 5. The estimation results of this study also indicate that state dependence, although statistically significant, may be ignored as far as the model's fit is concerned. The similarity of the coefficient estimates between Model 2 and Model 4, but the distinct coefficient estimates of Model 3, and the R^2 values of these models, all suggest that it is safer to ignore state dependence and incorporate serial correlation, than to ignore serial correlation and incorporate state dependence.

The significant serial correlation is due presumably to omitted variables that are longitudinally correlated. Considering that an extensive set of demographic and socioeconomic variables was considered in the model development of this study, it may be concluded that the omitted variables are not likely among the variables typically collected in transportation surveys.

CONCLUSION

The trip generation analysis of this study has shown that allowing for serial correlation and incorporating a lagged dependent variable significantly improve the model's fit. The improvement that a lagged dependent variable offers cannot be attributed to the effects of independent variables from previous time points. Serial correlation, however, contributes more substantially to the model's fit than does a lagged dependent variable. Trip generation is both serially correlated and state dependent, but the former plays a more dominant role.

The exploration of alternative model specifications has shown that estimated model coefficients and *t*-statistics differ substantially depending on model specification. Serial correlation should be incorporated whenever feasible for efficient estimation and improved fit. The results of this study indicate that it is safer to ignore state dependence and incorporate serial correlation than to ignore serial correlation and incorporate state dependence.

The significant serial correlation is due presumably to omitted variables that are longitudinally correlated. Considering that an extensive set of variables was considered in the model development of this study, it is not likely that the omitted variables are among those typically used in travel demand models and conventionally contained in travel survey data. Some determinants of travel behavior appear to lie outside the set of variables that have traditionally been considered in travel behavior analysis.

Finally, the contribution of serially correlated errors and lagged dependent variables to the predictive performance of a model has not yet been examined. The work on this subject is in its early stages, and only limited results have been reported (*15*). The many waves of weekly trip diary data now available from the Dutch National Mobility Panel data set offer a unique opportunity for further examination of this important issue.

ACKNOWLEDGMENTS

This research was performed while the author was at Bureau Goudappel Coffeng (BGC), Deventer, the Netherlands, in July through September 1987. For the support he received, the author is grateful to the BGC staff—in particular, to Jacqueline Visser, who prepared the data file used in this study. The valuable comments by Toon van der Hoorn on an earlier version of this paper are also gratefully acknowledged. The funding was provided by Rijkswaterstaat, Dienst Verkeerskunde, Dutch Ministry of Transport and Public Works.

REFERENCES

1. P. B. Goodwin. A Panel Analysis of Changes in Car Ownership and Bus Use. *Traffic Engineering and Control*, 27, October 1986, pp. 519–525.
2. P. B. Goodwin. *Family Changes and Public Transport Use 1984–1987: A Dynamic Analysis Using Panel Data.* Report prepared for the Project Bureau of the Netherlands Ministry of Transport, Bureau Goudappel Coffeng, Deventer, the Netherlands, 1987.
3. R. Kitamura. Determinants of Household Car Ownership and Utilization: A Panel Analysis. Presented at the International Conference on Travel Behavior, Aix-en-Provence, France, October 1987.
4. R. Kitamura and T. van der Hoorn. Regularity and Irreversibility of Weekly Travel Behavior. *Transportation*, 14, 1987, pp. 227–251.
5. R. Kitamura. A Panel Analysis of Household Car Ownership and Mobility. *Proceedings of the Japan Society of Civil Engineers*, No. 383/IV-7, 1987, pp. 13–27.
6. R. Kitamura. A Dynamic Model System of Household Car Ownership Trip Generation and Modal Split: Model Development and Simulation Experiment. In *Proceedings of the 14th Conference of the Australian Road Research Board, Part 3*, ARRB, Vermont South, Victoria, Australia, 1989, pp. 96–111.
7. G. E. Box and G. M. Jenkins. *Time Series Analysis, Forecasting, and Control.* Holden-Day, San Francisco, 1976.
8. Z. Griliches. Distributed Lags: A Survey. *Econometrica*, 35, 1967, pp. 16–47.
9. J. M. Golob, L. J. M. Schreurs, and J. G. Smit. The Design and Policy Applications of a Panel for Studying Changes in Mobility Over Time. In *Behavioural Research for Transport Policy*, VNU Press, Utrecht, the Netherlands, 1986, pp. 81–95.
10. T. F. Golob and H. Meurs. Biases in Response Over Time in a Seven-Day Travel Diary. *Transportation*, 13, 1986, pp. 163–181.
11. R. Kitamura and P. H. L. Bovy. Analysis of Attrition Biases and Trip Reporting Errors for Panel Data. *Transportation Research A*, 21A, 1987, pp. 287–302.
12. H. Meurs, P. van de Mede, J. Visser, and L. van Wissen. *Analysis of Panel Data.* Bureau Goudappel Coffeng, Deventer, the Netherlands, 1987.
13. D. F. Hendry and G. E. Mizon. Serial Correlation as a Convenient Simplification, Not a Nuisance: A Comment on a Study of the Demand for Money by the Bank of England. *Economic Journal*, 88, 1978, pp. 549–563.
14. G. S. Maddala. Recent Developments in the Econometrics of Panel Data Analysis. *Transportation Research A*, 21A, 1987, pp. 303–326.
15. T. Van der Hoorn and R. Kitamura. Evaluation of the Predictive Accuracy of Cross-sectional and Dynamic Trip Generation Models Using Panel Data. Paper presented at the 66th Annual Meeting of the Transportation Research Board, Washington, D.C., January 1987.

Publication of this paper sponsored by Committee on Traveler Behavior and Values.