UCI-ITS-WP-95-2

Models of Dynamic Commuter Behavior Using Longitudinal Data

UCI-ITS-WP-95-2

Mandar Khanal
Will Recker

Department of Civil and Environmental Engineering and
Institute of Transportation Studies
University of California, Irvine

January 1995

Institute of Transportation Studies
University of California, Irvine
Irvine, CA 92697-3600, U.S.A.
http://www.its.uci.edu

Prepared for the 74th Annual Meeting of the Transportation Research Board, Washington, D.C., January 1995.

MODELS OF DYNAMIC COMMUTER BEHAVIOR USING LONGITUDINAL DATA

M. Khanal and W. W. Recker

University of California, Irvine, CA 92717

Department of Civil and Environmental Engineering

and

Institute of Transportation Studies

University of California, Irvine, CA 92717

Abstract

The majority of demand models used at present are based on cross-sectional data. Behavior, however, is temporally related. Using three waves of a panel data of 2200 commuters in Southern California, this paper conducts a comparative analysis of three types of models: an ordered probit model, a two-period joint choice probit model, and a two-period dynamic beta-logistic model. The choice behavior modeled is the choice between driving alone and sharing a ride for the work commute. Prediction tests both on a hold-out sample as well as on a forecast sample were conducted. With the hold-out sample, all three models performed similarly. With the forecast sample, however, the beta-logistic model performed better than the other two models in aggregate predictions, and approximately the same as the joint choice probit model in disaggregate predictions, while both of these models performed better than the ordered probit model.

1. INTRODUCTION

The level and extent of demand for a transportation service, including the determinants of the demand, can be meaningfully analyzed only by incorporating their evolution over time. Conversely, analytical tools in use at the present time are, in the majority of cases, based on data observed at a single point in time. Although time series on aggregate measures of various aspects of transportation have been used for trend projections, and cross-sectional data for different time periods pertaining to the same system have been collected and used in the past, observations on the same set of persons over an extended period of time have only infrequently been collected and used in transportation behavior analysis. To the extent that temporal dependences and questions arising from heterogeneity of behavior are important, behavioral modeling in the transportation planning field is not adequately developed.

Problems of state dependence and heterogeneity in behavioral modeling have been recognized in various disciplines, including labor economics (Heckman and Willis, 1977; Heckman, 1981b; and Davies, Crouchley and Pickles, 1982), urban and regional planning (Clark and Huff, 1977; Davies, 1984; Davies and Crouchley, 1984; Davies and Pickles, 1985; Dunn and Wrigley, 1985; and Dunn, Reader and Wrigley, 1987), and highway safety (Bates and Neyman, 1951). Applications more directly related to transportation decision making include the works of Goodwin (1989), Mahmassani and Chang (1986), Mahmassani (1990), Johnson and Hensher (1982), Daganzo and Sheffi (1982), Uncles (1987), Kitamura and Bunch (1990), Hensher and Le Plastrier (1985), and Smith, Hensher and Wrigley (1986), all of which are concerned with discrete choice-making processes.

Existing discrete-time models for recurrent choice analysis generally can be classified into two major categories: "error distribution models" and "probability distribution models." In the former category of models the effect of such random variations as those rooted in heterogeneity in behavior are removed by integrating the probability expression over the domain of the random component, requiring *a priori* assumptions about the distributional properties of the random component. Examples of error distribution models are provided in Heckman (1981a) and Kitamura and Bunch (1990).

Kitamura and Bunch (1990) present a dynamic analysis of household car ownership in which the effects both of heterogeneity in behavior among households and of past behavior on current behavior are accounted for explicitly. Examination of hypotheses on state dependence and heterogeneity has been accomplished through alternate specifications of the error terms. They test two specifications for the structure of error: a components-of-variance error scheme and a one-factor error scheme. Using the two schemes to control for heterogeneity, they estimate eight different variants of the model. With the components-of variance scheme, the conclusion was that heterogeneity is not significant but state dependence is; conversely, with the one-factor scheme both heterogeneity and state dependence were found to be statistically significant.

The Heckman and Kitamura and Bunch approaches belong to a class of dynamic models in which the heterogeneity of behavior is removed by integrating out the error term from the model. Apart from this explicit recognition of heterogeneity and the treatment used to control for it, reliance is largely on techniques traditional to discrete choice modeling. Probability distribution models are, however, from a different class. Instead of focusing on the error term in the utility or latent value function of the alternatives in the choice models, this class of models is predicated on distributional assumptions about the probability of choice itself. Closed solutions are obtained for a certain class of distributions of the choice probability, with differences in the measured variables introduced in the model through the parameters of the distribution of the choice probability.

To account for heterogeneity or variation in choice behavior between two alternatives at a given point in time, the choice probability, p, can be assumed to be beta-distributed.

The beta distribution has the density function:

$$f(p) = \frac{p^{a-1}(1-p)^{b-1}}{B(a,b)} \qquad a,b>0 \qquad (1)$$

where

$$B(a,b)=\int_{0}^{1} t^{a-1}(1-t)^{b-1}dt=\frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}, \text{ and} \tag{2}$$

$\Gamma(\,)$ is the gamma function.

There are three reasons for the choice of beta distribution to represent the distribution of probability of choice. First is that its domain is the interval (0,1). Second is that it is parsimonious with only two parameters, a and b. Third is that the shape of the distribution is flexible. When $a>1$ and $b>1$ the distribution is unimodal. The shape is U-shaped when $a<1$ and $b<1$. It is J shaped when $a>1$ and $b<1$, and is reversed J-shaped when $a<1$ and $b>1$. The shape is rectangular when both a and b are equal to 1. With regard to symmetry: when $a=b$ it is symmetric; it is negatively skewed when $a>b$ and positively skewed when $a<b$.

With the specification of the distribution of probability, the effect of unobserved factors on the probability of choice can be removed by integrating the probability function over its domain. Mean and variance can be thus obtained. An especially attractive feature of the beta distribution in this regard is that these integrations reduce to gamma functions, computations for which can be further reduced by using the recurrence relation $\Gamma(x+1) = x\Gamma(x)$.

If it can be assumed that the error terms in the two time periods are uncorrelated, the choice over the two periods can be modeled as a product of two probability functions, each beta-distributed. But the error terms are correlated due to person specific unobserved factors. Therefore, the choice probabilities for the different time periods are also related. If this relationship between choice probabilities can be specified, then the sequence of choice over time can also be expressed in terms of a single beta distribution function.

Specification of such a relationship will allow for the measurement of state dependence of choice at one time period on choice at some preceding time point. This state dependence is the dependence between propensities of choice, since the dependence measured is between probabilities of choice and not between dummy variables measuring the selection or rejection of the alternatives or states. The association between propensities of choice measures the degree of attachment or the strength of attraction to an alternative. In a dummy variable, this measure of the strength of attachment is lost and all that is recorded is whether or not the person chooses a particular alternative.

Approaches have been adopted by other researchers toward extending the one period static Beta-Logistic Models to model choice behavior over two or more periods. In a study of residential mobility histories from birth of 633 individuals residing in County Borough of Leeds in England in 1973, Davies (1984) examined both heterogeneity and state dependence problems. The heterogeneity problem is addressed using the beta distribution approach, while the state dependence, or dynamic part, is handled by assuming that all probabilities are a logistically-scaled function of some reference probability. Denoting the reference period as r, this means that the probability for some time t is given by:

$$P_{qt} = \frac{\gamma_{qt}\left(P_{qr}\right)}{\left(1 - P_{qr}\right) + \gamma_{qt}\left(P_{qr}\right)} \tag{3}$$

where,

γ_{qt} = the logistic scaling factor for individual q at time t.

The logistic scaling factor used by Davies is given by:

$$\gamma_{qt} = e^{\left(X_{qr} - X_{qt}\right)' \beta}. \tag{4}$$

The vector X_{qt} is the vector of exogenous variables associated with decision maker q and the time t. For computational simplification, the likelihood function is expressed in terms of probability for the reference period using a Maclaurin series expansion of the logistic scaling function; the logarithm of the likelihood is then maximized to compute the required parameter estimates.

The effects included in the study were: duration of stay in years, age in years beyond age 21, and time in years measured from year 1920. Age effects and the need for time-varying variables were shown to exist, while the evidence of duration-of-stay effects was not found to be conclusive. Though the work by Davies represents a valuable extension of the beta-logistic approach and an important step toward estimating truly dynamic models, the following caveat about the methodology adopted appears to be in order.

By expressing the posterior probability of choice P_{qt} as a deterministic function of the reference period probability P_{qr}, the covariance between two period error components, expressed by equation (6), is reduced to zero in the Beta-Logistic Model generalization adopted by Davies. No theoretical justification to using the logistic scaling factor is given, other than the fact that it allows the posterior probability value to be limited between zero and unity. The generalization is also not tested empirically. Additionally, the values of all exogenous variables are assumed to be zero for the reference period. The distribution of probability is thus defined only by constants in the beta parameters.

Smith *et al.* (1986) used the methodology developed by Davies to analyze the decision of households relative to whether to replace or to keep their car in each year of a four year period between 1981 and 1984 in Sydney, Australia. Two groups of variables were used in the model: "slow" and "fast." The "slow" set of variables are those that are unchanging or are changing only slowly and are used in the specification of the beta distribution parameters. The "fast" variables are those whose values vary during the course of the study period. These variables are used in the specification of the scaling factor in the logistic scaling function.

The main problem with the paper by Smith *et al.* is the division of exogenous variables into the "slow" and "fast" groups. The authors do not present either a theoretical or an empirical reasoning of key methodological decisions taken in their study. The extension of Beta-Logistic models to longitudinal analysis appears to be unjustifiable, and thus the conclusions obtained from such extensions are suspect.

The research methodology adopted in this paper is to test the validity of the generalization adopted in the two papers described above. Since a theoretical test of the generalization is not readily available, the methodology adopted is to conduct empirical tests. The empirical tests are conducted in the context of the dynamics of the choice of mode for daily commute among full-time workers in the southern California South Coast Air Basin.

2. DATA ON COMMUTING BEHAVIOR

The data used in this study relate to the commuting behavior of full-time employees in the South Coast Air Basin in California. They are from a panel (instituted in 1990) of approximately 2,200 commuters. A detailed description of the panel study is provided in Uhlaner and Kim (1992). Approximately half of the commuters in the panel work in the Irvine Business Complex (IBC) located near John Wayne Airport in Orange County. The other half are from different areas from around the contiguous urbanized four-county area comprising Los Angeles, San Bernardino, Orange, and Riverside in Southern California. Businesses included in the survey are all relatively large, with 100 or more employees. The survey was conducted by the Institute of Transportation Studies, the University of California, Irvine, as part of a University of California Transportation Center project on commuter behavior. The first wave consists of 2,189 commuters. The study uses Waves 1, 5, and 8 of this panel, which encompasses a period beginning February, 1990 to February, 1993. The time periods during which the surveys were mailed were: February and March of 1990 for Wave 1, July of 1991 for Wave 5, and the middle of February of 1993 for Wave 8.

A brief description of the questionnaire used in the first wave is provided below. The questionnaire has six parts, Part A through Part F. The first part, Part A, is short and has questions regarding the day, the distance traveled, the travel time length, the congestion level, whether or not a freeway was used on the respondent's last working day, and the mode used for the commute. Following this preliminary section, there are three parts, Part B through Part D, based on the mode used. The modes are: bus; shared ride in a car, truck, or van; and drive alone, including motorcycle. Respondents are also allowed to pick a fourth mode, 'other'. However, only 1.1% of the total number of respondents in the first wave indicated this choice as their mode of travel. (Since the 'other' mode is not restricted to one specific mode, there is no separate mode-specific part devoted to the 'other' mode.) Part E contains a set of questions on the respondent's attitude toward various issues regarding mobility and congestion, including whether or not the respondent rode the bus any time during the previous two weeks, and questions regarding the respondent's knowledge of employer incentives for non-solo modes. The last part of the questionnaire (Part F) has socio-economic and personal questions related to mobility.

The mode-specific sections of the questionnaire have questions that can be classified broadly into three groups. The first group of questions are, as the name suggests literally, mode-specific. For example, for bus mode some of the questions in the first group are whether the service type was express or local, whether or not a bus pass was used, and whether or not bus use involved transfers. The second group of questions are of a more general nature. Some examples of the second group are: the length of walk from the bus stop or from the drop-off point or parking spot to the work site, whether stops were made in the trip from home-to-work or from work-to-home, and the frequency of use of the mode during last two weeks. The third group of questions in each of the three mode-specific parts are related to the availability to the respondent of the two alternate modes not used on the day of the survey.

The questionnaire used in the fifth wave is slightly different from that used in the first. The changes in the newer questionnaire reflect improvements in the questionnaire design. There are also some new questions about changes in respondent's job and employment type and location. The number of response categories in some variables such as occupation type and income, are also different in the two waves. The categories in such cases were redefined by combining two or more classes into one class. Details of such modifications follow later in the paper. Overall, there are many common travel-related survey questions to allow for a meaningful dynamic analysis of travel behavior. The eighth wave questionnaire is similar to the questionnaire for Wave 5.

Choice data are available for three modes: bus, shared ride, and drive alone. Of the 2189 respondents in the first wave only 1.7% were found to use bus, while for the shared ride and the

Uncles, M.D., 1987, "A beta-logistic model of mode choice: goodness of fit and intertemporal dependence". *Transportation Research B* **21B No 3**, 195-205.

Wachs, M., 1991, "Transportation Demand Management: Policy Implications of Recent Behavioral Research", *Journal of Planning Literature*, **Vol. 5, No. 4**, 333-341.

Willson, R.W. and D.C. Shoup, 1990, "Parking subsidies and travel choices: Assessing the evidence", *Transportation*, **17**, 141-157.

drive alone modes the market shares were 17.4% and 79.9%, respectively. Because of this predominance of the shared ride and the drive alone modes, the bus mode is excluded from the analysis reported here. The two modes selected for analysis will be abbreviated as SR for shared ride and DA for the drive alone modes.

3. MODEL SPECIFICATION

Denote the probability of choice of SR for wave i as p_i. The probability of choice of DA is thus $1-p_i$. With two time periods (Waves 1 and 5) and two choices per time period (DA and SR) the four possible state transitions are: SR→SR, SR→DA, DA→SR, and DA→DA. The probabilities of choice for different time periods in each of the four paths can not be expected to be independent of each other due to correlation among unobserved factors in the different time periods. If this correlation can be removed, the probability function for each of the four possible paths can be specified very simply, and the probability functions, in the order listed above, can be represented as: p_1p_5, $p_1(1-p_5)$, $(1-p_1)p_5$, and $(1-p_1)(1-p_5)$.

For individual q, at time t, let the probability of sharing a ride either exclusively or occasionally be P_{qt}. Since only two modes are considered, the probability of driving-alone is thus equal to $(1-P_{qt})$. Let there be a total of T time periods and define an indicator variable f_{qt} which takes a value of zero if the individual drives alone in period t and 1 if he or she does not. Then the likelihood function for the probability of the individual choosing the exhibited sequence can be expressed as a conditional function of the parameters of the probability processβ, and the individual-specific error terms e_q and e_{qt}, as shown in the following equation.

$$L_q\left[j^*/\beta, e_q, e_{qt}\right] = \prod_{t=1,5} P_{qt}^{f_{qt}}\left(1-P_{qt}\right)^{1-f_{qt}}, \tag{5}$$

where,

j* is the exhibited sequence of choice.

The probability of choice during a given time period is specified, following Davies (1984), as a logistic function of the probability for the reference time period r. The reference period is taken to be the first period, which in this case is Wave 1. The subsequent period probability is assumed to be a function of the reference period probability as shown in Equation (3). The logistic scaling factor γ_{qt} is expressed as an exponential function of the difference between the exogenous variables for the two periods and also of the levels of the variables in the first period, similar to, but not quite the same as, the function shown in Equation (4).

The likelihood function expressed by Equation (5) is conditional. It is subscripted by q to signify the conditioning on individual-specific error terms. The individual-specific effects are removed by integrating the product of L_q and the density function of choice probability over its domain, resulting in

$$L = \int_0^1 L_q f(p) dp \tag{6}$$

Making use of the assumption that p is beta distributed, the expression for the conditional likelihood (CL) function for each of the four possible paths outlined earlier can be reduced to the following simple form:

$$\begin{aligned}
CL_{SS} &= p_1 p_5 \\
CL_{SD} &= p_1 (1 - p_5) \\
CL_{DS} &= (1 - p_1) p_5 \\
CL_{DD} &= (1 - p_1)(1 - p_5)
\end{aligned}$$

The unconditional likelihood (UL) function for the four alternative paths are as shown below:

$$UL_{SS} = \int_0^1 \frac{\gamma_5 p^{a+1} (1-p)^{b-1}}{(1 - p + \gamma_5 p) B(a,b)} dp \tag{7a}$$

$$UL_{SD} = \int_0^1 \frac{p^a (1-p)^b}{(1 - p + \gamma_5 p) B(a,b)} dp \tag{7b}$$

$$UL_{DS} = \int_0^1 \frac{\gamma_5 p^a (1-p)^b}{(1-p+\gamma_5 p)B(a,b)} dp \tag{7c}$$

$$UL_{DD} = \int_0^1 \frac{p^{a-1} (1-p)^{b+1}}{(1-p+\gamma_5 p)B(a,b)} dp \tag{7d}$$

B(a,b) in Equations (20) denotes, as before, the beta function with parameters "a" and "b."

4. MODEL ESTIMATION

A review of current literature (refer to Cervero and Griesenbeck (1988), Feeney (1989), Ferguson (1990), Small(1983), Teal (1987), Wachs (1991), and Willson and Shoup (1990), among others) on ride-sharing and on the effects of incentives for high-occupancy and disincentives against single-occupancy vehicles in the commute to work helped identify the following groups of variables as potential determinants of the choice between the two modes: *accessibility, mobility, socio-economic, dynamic, incentive, and disincentive* variables. The individual variables available from the survey and belonging to any one of the above groups are listed in Table 1 (the variable symbols are shown in bold). For each variable two symbols are shown, one corresponding to level of the variable in Wave 1 and the other corresponding to difference in the variable values between Waves 1 and 5 (Wave 5 minus Wave 1).

The mean of the distance-to-work for the 681 cases with non-missing values is 16.5 miles and the median is 13 miles. Of these, more than half (376) report having a change in value between Waves 1 and 5, with a mean change of about .7 miles. (With a standard error of .35, the estimate of the mean is almost equal to zero, at the 5% level.) Mean vehicle ownership for this sample is 2.41; 234 households reported a change in vehicle ownership between the two waves.

Over 65 percent of the commutes use a freeway in either wave. Of the 456 commuters who used a freeway 213 had an HOV lane available in Wave 1. This breakdown for Wave 5 was 255 of 478 commuters who used a freeway. Only 3.3 percent of the commuters reported traffic as being *very heavy* in their commutes in Wave 1, although almost 20 percent commuters reported traffic as being *heavy* in the same wave.

TABLE 1. Variable Descriptions

VARIABLE GROUP	VARIABLE DESCRIPTION	VARIABLE NAME	
ACCESSIBILITY	Natural logarithm of home to work distance	**DIST**	**DDISTWRK**
	Use of freeway during commute	**W1FRWYUZ**	**DFRWY**
	Existence of an HOV lane on any freeway used	**W1HOVLAN**	**DHOVLANE.**
MOBILITY	Number of vehicles owned by household	**W1VEH**	**DVEH**
SOCIO-ECONOMIC	Household size	**W1HHSIZE**	**DHHSIZE**
	Children under 15 in the household	**W1CHLD15**	**DCHLD15**
	Workers in the household	**W1WORKER**	**DWORKER**
	Household income	**W1HHINCM**	**DINCOME**
	Education level of the respondent	**W1EDCTN**	
	Size of employment site in Wave 5	**W5EMPSIZ**	
DYNAMIC	Changed employers during the last six months:	**CHGEMP6M**	
	Changed employment since January 1990	**CHGEMJ90**	
	Changed job during last 6 months	**CHGEML6M**	
	Changed residence since January 1990	**CHGRSJ90**	
	Bought vehicles during the last 6 months	**BUYVEH6M**	
	Sold vehicles during the last 6 months	**SLDVEH6M**	
	Moving residence within the next year	**MVNNXTYR**	
RS INCENTIVE	Reserved parking for ride-sharing	**RESERVPK**	**DRESERV**
	Cost subsidy for ride-sharing	**CSTSBSDY**	**DCSTSB**
	Guaranteed ride home	**GRNTDRID**	**DGRNTD**
DA DISINCENTIVE	Traffic was very heavy during home to work trip	**W1CONGVH**	
	Traffic was heavy during home to work trip	**W1CONGH.**	

Wave 1 and Wave 5 data were merged and a subset that included respondents common to both waves was extracted from the merged set. This resulted in a file with 1062 cases. The attrition rate was about 40 percent (not all of the 2189 returned first wave surveys were complete). Brownstone and Chu (1992) have reported that the attrition seen between the two waves, Waves 1 and 5, are not ignorable. However, their results indicate that attrition bias is not serious for the data set or the models used in their analysis. No correction is made for attrition bias in the models presented here.

The removal of erroneous cases during the "data cleaning" phase of the research resulted in a file size of 1025 cases. Of these, 75 percent were randomly sampled to form a file for model development. The remaining 25 percent were set aside as a hold-out sample to be used in evaluating predictions after the model building stage.

The GAUSS-386 software package from Apteck Systems Inc.(1991) was used to estimate the parameters of the likelihood function system shown in Equations (7). Unlike previous efforts, no series approximation of the logistic scaling function was done in the estimation of the model shown in Equation (7) above; the model system was directly used in a maximum likelihood estimation of the variable coefficients. The modeling effort for the dynamic models described in this section start with a comprehensive explanatory variable set that was identified in an analysis with conventional probit models that is described in Section 5 below. The variable set used in the model shown in Table 2 below is the most comprehensive set of variables determined in that analysis. The model shown in Table 2 is denoted as the full model. As can be seen from the table not all variables have commonly accepted significance levels. However, the variables are retained in the analysis because the primary focus of this research is in testing the structure of dynamic Beta-Logistic Models rather than in finding a individual variables with satistically significant influence on mode choice. Also, since the model's effectiveness will be determined on the basis of prediction success rate, the inclusion of some variables with lower than normally accepted significance levels is not believed to lead to erroneous conclusions regarding the predictive capabilities of the developed model.

TABLE 2. Full Model Coefficient Estimates

MODEL I	COEFFICIENT ESTIMATES					
	γ PARAMETERS		BETA FUNCTION - A PARAMETERS		BETA FUNCTION - B PARAMETERS	
PARAMETERS	EST.	t-STAT.	EST.	t-STAT.	EST	t-STAT.
CONSTANT	1.79	1.24	-2.36	-0.10	5.18	4.05
DVEH	-0.05	-0.10	-1.86	-0.16	1.05	2.47
DHHSIZE	0.28	0.46	2.79	0.26	0.39	0.98
DDISTWRK	0.20	0.44	0.77	0.06	-0.60	-1.21
DINCOME	0.07	0.28	1.65	0.65	-0.14	-0.86
DWORKER	-0.24	-0.50	-0.70	-0.09	-1.10	-2.57
DCHLD15	0.09	0.12	-4.63	-0.31	-0.54	-0.95
DRESERV	0.61	1.03	3.21	0.25	-0.10	-0.20
DCSTSB	-0.09	-0.10	3.41	0.26	0.50	0.63
DGRNTD	1.54	2.67	-2.36	-0.13	0.17	0.35
W1VEH	0.05	0.11	-13.63	-0.75	0.13	0.44
W1HHSIZE	-0.30	-0.54	6.25	0.41	0.80	1.82
DIST	-0.52	-1.28	-3.36	-0.45	-1.64	-4.21
W1HHINCM	-0.08	-0.73	0.670	0.36	0.03	0.28
W1WORKER	0.25	0.41	0.52	0.05	-1.22	-2.60
W1CHLD15	0.50	0.80	-5.41	-0.34	-0.91	-1.82
RESERVPK	-0.80	-1.30	8.39	0.36	-0.92	-1.91
CSTSBSDY	0.31	0.33	-8.32	-0.02	-0.34	-0.37
GRNTDRID	2.37	2.45	4.05	0.30	-1.01	-1.38

AUXILIARY STATISTICS	
-L.L.(# of param.)	433.87(57)

Next, the results from successively restricting the full model are presented. Model II (shown in Table 3) is nested in model I -- all of the difference variables in the γ term as well as in the beta functions, a and b, have been constrained to zero in model II. The results shown in Table 3 indicate that this restriction does not appear to be valid based on a likelihood ratio test between models I and II.

TABLE 3. Restricted Model II Coefficient Estimates

MODEL II	COEFFICIENT ESTIMATES					
	γ PARAMETERS		BETA FUNCTION - A PARAMETERS		BETA FUNCTION - B PARAMETERS	
PARAMETERS	EST.	t-STAT.	EST.	t-STAT.	EST.	t-STAT.
CONSTANT	2.11	1.82	0.76	0.000	4.39	4.34
DVEH	-	-	-	-	-	-
DHHSIZE	-	-	-	-	-	-
DDISTWRK	-	-	-	-	-	-
DINCOME	-	-	-	-	-	-
DWORKER	-	-	-	-	-	-
DCHLD15	-	-	-	-	-	-
DRESERV	-	-	-	-	-	-
DCSTSB	-	-	-	-	-	-
DGRNTD	-	-	-	-	-	-
W1VEH	0.09	0.27	-14.64	-0.49	0.06	0.31
W1HHSIZE	-0.54	-1.17	4.21	1.08	0.39	1.40
DIST	-0.48	-1.38	-4.43	-0.37	-1.32	-4.47
W1HHINCM	-0.08	-0.86	-1.14	-0.73	0.01	0.12
W1WORKER	0.48	1.07	2.32	0.16	-0.72	-1.99
W1CHLD15	0.66	1.21	-3.68	-0.55	-0.48	-1.38
RESERVPK	-0.83	-1.72	11.45	0.002	-0.98	-2.63
CSTSBSDY	0.82	1.08	-8.06	-0.25	-0.11	-0.15
GRNTDRID	1.59	1.93	15.71	0.47	-0.75	-1.21

AUXILIARY STATISTICS	
-L.L.(# of param.)	467.15(30)
L.R.χ^2(d.f.)	66.55(27)
Restrictions on Model	I
Restricted Group of variables	Differences in Functions γ, A, & B

In model III (shown in Table 4), only the differences in the beta functions, a andb, have been reduced to zero. This specification is somewhat analogous to the specifications reported in the literature. However, as evidenced by the likelihood ratio test between models I and III, the restrictions are not justified statistically. Thus, difference variables can not be ignored in beta-logistic dynamic variables.

TABLE 4. Restricted Model III Coefficient Estimates

MODEL III	COEFFICIENT ESTIMATES					
	γ PARAMETERS		BETA FUNCTION - A PARAMETERS		BETA FUNCTION - B PARAMETERS	
PARAMETERS	EST.	t-STAT.	EST.	t-STAT.	EST.	t-STAT.
CONSTANT	0.94	0.71	3.11	0.88	4.19	4.017
DVEH	-0.74	-1.80	-	-	-	-
DHHSIZE	0.19	0.44	-	-	-	-
DDISTWRK	0.59	1.46	-	-	-	-
DINCOME	0.21	1.08	-	-	-	-
DWORKER	0.36	0.86	-	-	-	-
DCHLD15	0.31	0.56	-	-	-	-
DRESERV	0.67	1.392	-	-	-	-
DCSTSB	-0.23	-0.35	-	-	-	-
DGRNTD	1.43	2.92	-	-	-	-
W1VEH	-0.09	-0.22	-8.72	-1.83	-0.11	-0.44
W1HHSIZE	-0.50	-0.94	2.11	1.80	0.41	1.45
DIST	-0.39	-1.00	-0.80	-1.64	-1.24	-4.16
W1HHINCM	-0.03	-0.31	0.14	0.78	0.04	0.52
W1WORKER	0.62	1.17	1.30	1.07	-0.62	-1.71
W1CHLD15	0.75	1.28	-2.51	-1.58	-0.53	-1.47
RESERVPK	-0.53	-0.92	0.52	0.47	-0.82	-2.08
CSTSBSDY	0.30	0.34	-8.17	-1.32	-0.34	-0.44
GRNTDRID	2.21	2.49	6.98	1.46	-0.55	-0.84

AUXILIARY STATISTICS	
-L.L.(# of param.)	454.65(39)
L.R.χ^2(d.f.)	41.55(18)
Restrictions on Model	I
Restricted Group of variables	Differences in Functions A & B

Model IV (shown in Table 5) has level variables in the γ function restricted to zero. These restrictions are based on the premise that variables measuring levels are not appropriate in the γ function since, purportedly, the need for including the γ function in the dynamic model is to capture the dynamic behavior of probability of choice and levels do not capture dynamic effects. This hypothesis appears to be valid based on a likelihood ratio test between models IV and I.

TABLE 5. Model IV Coefficient Estimates

MODEL IV	COEFFICIENT ESTIMATES					
	γ PARAMETERS		BETA FUNCTION - A PARAMETERS		BETA FUNCTION - B PARAMETERS	
PARAMETERS	EST.	t-STAT.	EST.	t-STAT.	EST.	t-STAT.
CONSTANT	0.06	0.16	-2.89	-0.12	4.27	4.48
DVEH	-0.09	0.23	-1.86	-0.17	1.01	2.60
DHHSIZE	0.54	1.07	2.93	0.27	0.52	1.40
DDISTWRK	0.34	0.82	0.92	0.08	-0.51	-1.06
DINCOME	0.08	0.426	1.67	0.67	-0.14	-0.90
DWORKER	-0.46	-1.16	-0.72	-0.11	-1.17	-2.88
DCHLD15	-0.23	-0.36	-4.75	-0.32	-0.67	-1.22
DRESERV	1.02	2.20	3.37	0.28	0.06	0.13
DCSTSB	-0.37	-0.54	3.38	0.26	0.38	0.53
DGRNTD	0.85	1.62	-2.79	-0.16	-0.16	-0.34
W1VEH	-	-	-13.81	-0.76	0.13	0.52
W1HHSIZE	-	-	6.43	0.43	0.93	2.61
DIST	-	-	-3.28	-0.42	-1.38	-4.78
W1HHINCM	-	-	0.65	0.36	0.05	0.59
W1WORKER	-	-	0.53	0.06	-1.29	-3.26
W1CHLD15	-	-	-5.52	-0.37	-1.08	-2.60
RESERVPK	-	-	8.74	0.36	-0.58	-1.39
CSTSBSDY	-	-	-8.15	-0.02	-0.45	-0.54
GRNTDRID	-	-	3.77	0.28	-1.81	-2.45

AUXILIARY STATISTICS	
-L.L.(# of param.)	440.44(48)
L.R.χ^2(d.f.)	6.57(9)
Restrictions on Model	I
Restricted Group of variables	Levels in Function γ

Model V (shown in Table 6)is nested in model IV; the difference variables used in model IV have been dropped. A likelihood ratio test reveals that, statistically, theγ function needs only the constant term.

TABLE 6. Restricted Model V Coefficient Estimates

MODEL V	COEFFICIENT ESTIMATES					
	γ PARAMETERS		BETA FUNCTION - A PARAMETERS		BETA FUNCTION - B PARAMETERS	
PARAMETERS	EST.	t-STAT.	EST.	t-STAT.	EST.	t-STAT.
CONSTANT	0.35	1.62	-4.29	-0.33	4.14	4.49
DVEH			-1.49	-0.20	0.99	2.81
DHHSIZE			2.71	0.43	0.29	1.01
DDISTWRK			0.82	0.10	-0.65	-1.45
DINCOME			1.62	0.87	-0.17	-1.22
DWORKER			-0.31	-0.06	-0.92	-2.53
DCHLD15			-4.32	-0.46	-0.55	-1.21
DRESERV			3.53	0.45	-0.37	-0.90
DCSTSB			2.12	0.35	0.45	0.78
DGRNTD			-4.01	-0.39	-0.62	-1.47
W1VEH			-12.74	-1.43	0.14	0.56
W1HHSIZE			5.39	0.65	0.88	2.62
DIST			-2.95	-0.70	-1.34	-4.88
W1HHINCM			0.69	0.58	0.05	0.71
W1WORKER			1.26	0.21	-1.22	-3.20
W1CHLD15			-4.39	-0.51	-1.01	-2.53
RESERVPK			8.94	0.65	-0.47	-1.15
CSTSBSDY			-5.52	-0.31	-0.46	-0.58
GRNTDRID			3.43	0.44	-1.80	-2.63

AUXILIARY STATISTICS	
-L.L.(# of param.)	447.27(39)
L.R.χ^2(d.f.)	6.83(9)
Restrictions on Model	IV
Restricted Group of variables	DIFFERENCES IN γ FUNCTION

A further restricted model is estimated to check the very need of the γ function in the model system. Such a restriction is shown in model VI (shown in Table 7), in which the γ function has been removed entirely from the specification. This restriction appears to be justified statistically. The empirical results obtained so far indicate that the parameterization of the beta function employed here captures the effects of heterogeneity as well as non-stationarity. The need for the logistic function is thus obviated. Thus the generalization of the Beta-Logistic Model, adopted by Davies (1984), does not have any empirical basis in the longitudinal mode choice behavior exhibited by the data set used in this study.

TABLE 7. Restricted Model VI Coefficient Estimates

MODEL VI	COEFFICIENT ESTIMATES					
	γ PARAMETERS		BETA FUNCTION - A PARAMETERS		BETA FUNCTION - B PARAMETERS	
PARAMETERS	EST.	t-STAT.	EST.	t-STAT.	EST.	t-STAT.
CONSTANT			-4.33	-0.33	3.90	4.44
DVEH			-1.50	-0.20	0.98	2.77
DHHSIZE			2.70	0.43	0.27	0.92
DDISTWRK			0.83	0.11	-0.64	-1.43
DINCOME			1.62	0.88	-0.17	-1.19
DWORKER			-0.30	-0.06	-0.90	-2.48
DCHLD15			-4.32	-0.46	-0.52	-1.16
DRESERV			3.54	0.46	-0.38	-0.93
DCSTSB			2.16	0.35	0.46	0.80
DGRNTD			-4.00	-0.40	-0.60	-1.40
W1VEH			-12.77	-1.46	0.13	0.53
W1HHSIZE			5.42	0.66	0.84	2.53
DIST			-2.95	-0.70	-1.31	-4.91
W1HHINCM			0.70	0.57	0.06	0.76
W1WORKER			1.25	0.21	-1.17	-3.13
W1CHLD15			-4.42	-0.52	-0.97	-2.44
RESERVPK			8.96	0.65	-0.46	-1.13
CSTSBSDY			-5.40	-0.31	-0.47	-0.57
GRNTDRID			3.44	0.45	-1.75	-2.49

AUXILIARY STATISTICS	
-L.L.(# of param.)	448.63(38)
L.R.χ^2(d.f.)	1.36(1)
Restrictions on Model	V
Restricted Group of variables	CONSTANT in γ FUNCTION

The validity of the beta-logistic models developed above was evaluated on the basis of its predictive performance. Both the hold-out sample as well as a new sample using data from Waves 5 and 8 were used for predictions. The predictive performance of these models was also tested against those of more conventional models. The conventional models used are probit type models. Their development is described in the following section.

5 CONVENTIONAL MODELS

Ordered Probit

Ordered probit models differ from binary probit models primarily in the specification of the utility function and in the decision rule used in making a choice among alternatives. Specifically, utility functions are not defined separately for each of the alternatives, only one utility function is defined for the choice process. The different choices a user can make is assumed to be represented by an ordinal variable. In other words, it is assumed that the alternatives in the choice set can be put in an order. The ordinal dependent variable is further assumed to be related to an unobserved latent utility measure depending on the value of the latent utility. Various regions of the domain of the latent utility measure are assumed to have a one-to-one correspondence with the ordinal dependent variable, as explained in the next paragraph.

Denote the latent utility, the representative utility, and the unobserved utility by U, V, and ε, respectively. U is assumed to be related to V and ε as shown below:

$$U_i = V_i + \varepsilon_i .$$

Note that the variables have only one subscript which denotes the individual making the choice decision. The error term ε_i is assumed to be unit normally distributed. Define a variable Z_i which denotes an ordinal dependent variable with the number of outcomes equal to the number of alternatives in the choice set. Assume that the number of choices are three corresponding to shared ride, sometimes shared ride and drive alone modes. The dependent variable $Z_i$is related to the utility measure through a set of thresholds or constants, α's, as shown below:

Z_i = shared ride,	if	$U_i \le \alpha_1$
Z_i = sometimes shared ride,	if	$\alpha_1 < U_i \le \alpha_2$
Z_i = drive alone,	if	$\alpha_2 < U_i$.

Expressing the representative utility, V_i, in terms of independent variables and a set of parameters, $x_i'\beta$, the probabilities of choosing each of the three modes are as shown below:

$P[\text{shared ride}] = P[U_i \leq \alpha_1] = P[x_i'\beta + \varepsilon_i \leq \alpha_1] = P[\varepsilon_i \leq \alpha_1 - x_i'\beta] = \Phi(\alpha_1 - x_i'\beta)$,

$P[\text{sometimes shared ride}] = P[\alpha_1 < x_i'\beta + \varepsilon_i \leq \alpha_2] = \Phi(\alpha_2 - x_i'\beta) - \Phi(\alpha_1 - x_i'\beta)$, and

$P[\text{drive alone}] = P[\alpha_2 < x_i'\beta + \varepsilon_i] = 1 - \Phi(\alpha_2 - x_i'\beta)$.

Conditional Probit

The other type of conventional model that was developed for comparative purposes modeled the two period joint choice as a product of a conditional and a marginal choice. To help simplify the estimation, the choices were collapsed to two: *always drive-alone* (DA), and *always or sometimes shared-ride* (SR), allowing the estimation to be in the form of two binary probit models. The condition is the choice made in Wave 1.

6. COMPARATIVE MODEL ESTIMATION RESULTS

In the ordered probit model the dependent variable has three categories. The dependent variable is created from information obtained regarding respondents' choice of mode both on the day of the survey as well as on other days during the past two weeks. Using this information, commuters in the sample are classified into one of the following three groups: *ride-sharing*, *sometimes ride-sharing*, and *driving-alone*. The ordered probit models capture the category choice in Wave 5 based on explanatory variable values in Wave 5 as well as in Wave 1.

The results of the estimation of two ordered probit models that were developed from a series of specifications with various subsets of available explanatory variables are shown in Table 8. The variable set used was based on a priori expectations about determinants of travel behavior, that was obtained from a review of past studies, as explained in Section 4. Variables depicting levels in Wave 1 as well as differences in levels between Waves 1 and 5 were included in the analysis. It was deemed necessary to include the difference variables since one past study on dynamic beta-logistic models had used difference variables as determinants of dynamic behavior. The first model, labeled *Unrestricted Model*, uses the full set of variables. A sequence of other models which were nested in some model preceding it (i.e., which had successively more comprehensive subsets of variables which were restricted from the specification) were estimated and the likelihood ratio statistics for each pair of such models compared to determine if the models were significantly different from each other. On the basis of these tests the specification labeled *Restricted Model* has a set of explanatory variables that can not be further restricted without significantly changing the estimation results

TABLE 8. Ordered Probit Model Estimation Results

VARIABLE	FULL MODEL	RESTRICTED MODEL
	Coefficient Est. (t-Score)	Coefficient Est. (t-Score)
DIST	-0.38(-3.74)	-0.36(-4.66)
DDISTWRK	-0.36(-2.26)	-0.34(-2.36)
W1FRWYUZ	-0.24(-1.21)	
W5FRWYUZ	0.37(1.71)	
W1HOVLAN	0.13.(.75)	
W5HOVLAN	-0.30(-1.76)	
W1VEH	0.18(2.4)	
DVEH	0.12(1.24)	
W1HHSIZE	0.20(1.43)	0.31(2.38)
DHHSIZE	0.06(0.53)	0.12(1.14)
W1CHLD15	-0.31(-1.86)	-0.44(-2.88)
DCHLD15	-0.23(-1.35)	-0.29(-1.85)
W1WORKER	-0.53(-3.47)	-0.51(-3.62)
DWORKER	-0.28(-2.09)	-0.28(-2.28)
W1HHINCM	0.03(1.1)	0.03(1.34)
DINCOME	-0.05(-0.99)	-0.03(-0.73)
W5EMPSIZ	-0.04(-1.08)	
CHGEMJ90	0.17(0.71)	0.20(0.84)
CHGEML6M	0.75(1.84)	0.85(2.14)
BUYVEH6M	0.19(1.15)	0.25(1.50)
MVNNXTYR	0.18(1.21)	
RESERVPK	0.08(0.45)	0.05(0.31)
CSTSBSDY	-0.22(-0.9)	-0.19(-0.80)
GRNTDRID	-0.49(-2.36)	-0.55(-2.86)
RESERVP5	-0.10(-0.53)	-0.11(-0.67)
CSTSBSD5	0.01(0.05)	0.001(0.00)
GRNTDRD5	-0.34(-2.23)	-0.36(-2.53)
THRESHOLD 1	-2.71(-7.55)	-2.68(-8.34)
THRESHOLD 2	-1.77(-5.03)	-1.76(-5.65)
-2L.L.(#ofParameters)	804.09(27)	823.50(19)
% correctly predicted	70.4	69.9

Although they use time-changing explanatory variables, the ordered probit models described above are essentially static models since only one period choice behavior is modeled. They can not be used to model the two-period choice spanning Waves 1 and 5 dynamically, since the four alternatives that result in a two-period-two-alternative choice have no natural order in their propensities to be modeled using the ordered probit formulation. An alternative approach is to treat the dynamic choice as a conditional choice problem.

If the choice alternatives are formulated as being between always driving-alone (DA) and sometimes ride-sharing or always ride-sharing (SR), the choice process for the two waves viewed collectively consists of four paths: SR→SR, SR→DA, DA→SR, and DA→DA. To avoid the computational complexity of a multinomial probit formulation the four-alternative choice problem is transferred to a two-alternative conditional choice problem, with the condition being the choice made in Wave 1. Thus, based on whether the choice made in Wave 1 is DA or SR, two sets of probit models are estimated. The results of the estimation are shown in Table 9.

TABLE 9. Conditional Probit Model Estimation Results

VARIABLE	PROBABILITY CONDITIONED ON	
	SR in Wave 1 Coefficient Est. (t-Score)	**DA in Wave 1** Coefficient Est. (t-Score)
CONSTANT	0.63(1.41)	1.76(6.6)
DIST	-0.30(-1.96)	-0.25(-2.51)
DDISTWRK	-0.33(-1.21)	-0.11(-0.81)
RESERVPK	0.25(0.87)	0.05(0.27)
CSTSBSDY	-0.06(-0.12)	-0.15(-0.48)
GRNTDRID	-1.16(-2.62)	-0.01(-0.06)
RESERVP5	-0.24(-0.82)	-0.08(-0.41)
CSTSBSD5	0.18(0.45)	0.08(0.32)
GRNTDRD5	-0.22(-0.81)	-0.54(-3.05)
-2L.L.(#of Parameters)	177.87(8)	415.83(8)
% correctly predicted	67.7	82.8

Models conditioned on choice of DA in Wave 1 were estimated first. Starting with a large set of potentially significant variables, groups of variables were successively restricted and a most parsimonious model was obtained. A model conditioned on choice of SR in Wave 1 could not, however, be estimated using the specification used for the most parsimonious DA-conditioned model obtained as explained above. The specification shown in Table 3 had to be used instead since it worked for both conditional choices. The same specification for both conditional choices were used so that the predictions from both models could be combined for comparison with predictions that will be described in the next section.

Finally, Table 10 shows a beta-logistic model that has been modeled with the specification used in the conditional choice models of Table 9. This was done so that the prediction from the dynamic beta-logistic model could be compared with the prediction from the conventional joint choice probit model.

TABLE 10

MODEL VII	COEFFICIENT ESTIMATES			
	BETA FUNCTION - A PARAMETERS		BETA FUNCTION - B PARAMETERS	
PARAMETERS	EST.	t-stat.	EST.	t-stat.
CONSTANT	-2.80	-2.34	1.70	2.68
DDISTWRK	0.58	1.01	0.02	0.06
DRESERV	-1.89	-2.82	-1.01	-2.27
DCSTSB	-0.21	-0.27	-0.26	-0.40
DGRNTD	2.50	3.34	1.03	1.80
DIST	0.56	1.60	-0.47	-2.05
RESERVPK	-1.97	-3.17	-1.28	-3.19
CSTSBSDY	0.13	0.12	-0.37	-0.38
GRNTDRID	1.78	1.91	-1.01	-1.09
AUXILIARY STATISTICS				
-L.L.(# of param.)	598.27(18)			

7 MODEL VALIDATION RESULTS

The models (i.e., as specified in Tables 8, 9, and 10) developed and estimated in the previous section were validated in two ways. First, model results were tested against the hold-out sample comprised of Wave 1 and Wave 5 data. The second method of validation used data from Waves 5 and 8 and involved applying the model calibrated on the period encompassing Waves 1 and 5 to predict behavior in the succeeding time interval between Waves 5 and 8. (Recall that the time periods during which the surveys were mailed were: February and March of 1990 for Wave 1, July of 1991 for Wave 5, and the middle of February of 1993 for Wave 8. Since, the time between Waves 1 and 5 is approximately the same as that between Waves 5 and 8, it was deemed reasonable to use the models calibrated on Waves 1 and 5 to predict the choice between the latter two waves.)

There are 255 cases in the hold-out sample, comprising 25% of the total number of cases in the data set. Of the 255 cases, only 228 cases remained when a selection criterion of only solo drivers or car-poolers was applied, excluding respondents using other modes from the analysis. The prediction results associated with the ordered probit model for the choice among three alternatives for Wave 5 are shown in Table 11. When applied to the hold-out sample, the model produced an overall successful prediction rate of about 94% in terms of aggregate shares and about 70% in terms of disaggregate predictions. The difference between propensities of choice and actual assignment of choice based on the highest probability of choice can be seen clearly from the table in the difference in predictions in the aggregate shares and disaggregate classifications. For example, for the RS alternative, only 1 out of 20 was correctly predicted. This represents an error of 95%. However, for the same alternative, the deviation in prediction of aggregate shares is only about 3%. This phenomenon is repeated for the two remaining model types as well, as will be seen later in this paper.

TABLE 11. Ordered Probit Model Validation Results - Holdout Sample

Wave 5 Modal Choice	RS	SRS	DA	TOTAL
Observed Number	20	36	133	189
Predicted Number	1	1	131	133
% Correct	5.00	2.78	98.5	70.37
% Incorrect	95.00	97.22	1.50	**29.63**
Observed Shares(%)	10.58	19.05	70.37	
Predicted Shares(%)	7.64	18.98	73.38	
Predicted-Observed Shares	-2.94	-0.07	3.01	
$\frac{\textbf{Deviation in Shares}}{\textbf{Observed Shares}}(\%)$	-27.79	-0.37	4.28	
Absolute Sum of Deviations(%)				**6.02**

The numbers shown in bold are comparable measures for the aggregate and disaggregate predictions. For example, the overall incorrect prediction rate for the disaggregate prediction is about 30%. For the aggregate shares prediction, the overall discrepancy in prediction is only 6.02%.

The conditional models shown in Table 9 were also applied to the same holdout sample. Of the two models , the first estimates the choice of modes in Wave 5 given that the choice in Wave 1 was "*sometimes or always shared ride*". In the second model the dependent variable is the same, but the choice in Wave 5 is conditioned on having chosen the "*always drive alone mode*" in Wave 1.

To estimate the joint probability of choice for the two periods a binary probit model for Wave 1 was estimated. The specification of the binary model is the same as that of the conditional model except that no difference explanatory variables and Wave 5 incentive variables are included. Parameter estimates and other pertinent information for this model are shown in Table 12, where the choice probability is specified for drive alone.

TABLE 12. Binary Marginal Probit Model Estimation Results

VARIABLE	ESTIMATION	
	Coefficient	(t-Score)
CONSTANT	1.69	(8.5)
DIST	-0.326)	(-4.56
RESERVPK	-0.275	(-2.27)
CSTSBSDY	-0.108	(-0.48)
GRNTDRID	-0.35	(-1.93)
-2L.L.(#of Parameters)	45.3 (4)	
% correctly predicted	74.5	

The conditional and the marginal probit models were combined to produce the joint probability of choice in Waves 1 and 5 for the holdout sample. The results of this prediction are shown in Table 13. The overall successful prediction rate for the conditional model is over 97% for aggregate shares prediction, and only about 63% for the disaggregate prediction.

TABLE 13. Conditional Model Validation Results - Holdout Sample

Modal Transitions	SR → SR	SR → DA	DA → SR	DA → DA	TOTAL
Observed Number	32	21	30	133	216
Predicted Number	3	0	0	133	136
% Correct	9.38	0.00	0.00	100.00	62.96
% Incorrect	90.62	100.00	100.00	0.00	**37.04**
Observed Shares (%)	14.81	9.73	13.89	61.57	
Predicted Shares (%)	15.25	8.95	13.21	62.59	
Predicted-Observed Shares	0.44	-0.78	-0.68	1.02	
$\frac{\textbf{Deviation in Shares}}{\textbf{Observed Shares}}$ (%)	2.97	-8.02	-4.90	1.66	
Absolute Sum of Deviations(%)					**2.92**

The prediction results for the dynamic beta-logistic model are shown in Table 14.

TABLE 14. Model Validation Results - Holdout Sample

Modal Transitions	SR → SR	SR → DA	DA → SR	DA → DA	TOTAL
Observed Number	32	21	30	133	216
Predicted Number	3	0	0	129	132
% Correct	9.38	0.00	0.00	96.99	61.40
% Incorrect	90.62	100.00	100.00	3.01	**38.89**
Observed Shares (%)	14.81	9.73	13.89	61.57	
Predicted Shares (%)	15.78	10.60	10.60	63.02	
Predicted-Observed Shares	0.97	0.87	-3.29	1.45	
$\frac{\text{Deviation in Shares}}{\text{Observed Shares}}(\%)$	6.55	8.94	-23.69	2.36	
Absolute Sum of Deviations(%)					**6.58**

As before, the prediction error for aggregate prediction is the difference between the predicted share and the observed share for each alternative. The total prediction error is the sum of the deviations for all four alternatives, which in this case is about 7%. The prediction success rate of the beta-logistic Model developed in this research is thus a little over 93%. In terms of disaggregate predictions, the success rate is only about 61%. One other noticeable feature of the beta-logistic model prediction is that none of the mode-switchers were correctly predicted. However, as in other cases, the propensity of choice was predicted much more closely to the actual shares observed.

Validation based on using the model to forecast travel behavior in the subsequent time period started with the 724 cases of Wave 8 data, from which a subset of 628 cases was created that included only respondents choosing "shared ride" or "drive alone" as their primary mode of commute. Some of the variables were recoded and some categories were regrouped to match

Wave 8 responses with those of Waves 1 and 5, since the Wave 8 questionnaire differed slightly from that used in the previous waves. The resulting data file was then merged with the data file containing Wave 5 responses. A file with 428 cases was then created with variables corresponding to the period spanning Waves 5 and 8. Predictions using these 428 cases were then calculated using the ordered probit and conditional models. The results for the ordered probit models are shown in Table 15. As can be seen from the table the ordered probit model predictions are grossly erroneous for the forecast sample.

TABLE 15. Ordered Probit Model Validation Results - Forecast Sample

Wave 8 Modal Choice	RS	SRS	DA	TOTAL
Observed Number	29	78	242	349
Predicted Number	27	2	39	68
% Correct	93.10	2.56	16.12	19.48
% Incorrect	6.9	97.44	83.88	**80.52**
Observed Shares(%)	8.31	22.35	69.34	
Predicted Shares(%)	69.87	15.64	14.49	
Predicted-Observed Shares	61.56	-6.71	-54.85	
$\frac{\text{Deviation in Shares}}{\text{Observed Shares}}$ (%)	740.79	-30.02	-79.10	
Absolute Sum of Deviations				**123.12**

As in the case of the hold-out sample, the conditional models used for predictions in this case are those shown in Table 9. The first of these two models is used to predict the choice in Wave 8 given that the choice in Wave 5 was "sometimes or always shared ride". The second model predicts the choice in Wave 8 given that the choice in Wave 5 was "always drive alone". To compute the joint probability of choice during the two-wave period, marginal probabilities of choice are needed for Wave 5. These Wave 5 marginal probabilities are computed using the coefficient values of a binary probit model estimated using Wave 1 data. The joint probabilities for the two-wave period is then obtained from the product of the marginal probability of choice for each of the two Wave 5 choice and conditional probabilities for each of the two Wave 8 choices conditioned on the corresponding Wave 5 choice. The prediction for the conditional model is shown in Table 16. Successful prediction for this case is only about 65%, for aggregate shares, and.about 61% for the disaggregate prediction.

TABLE 16. Conditional Model Validation Results - Forecast Sample

Modal Transitions	SR → SR	SR → DA	DA → SR	DA → DA	TOTAL
Observed Number	77	41	42	218	378
Predicted Number	38	0	5	186	229
% Correct	49.35	0.00	11.90	85.32	60.58
% Incorrect	50.65	100.00	88.10	14.68	**39.42**
Observed Shares (%)	20.37	10.85	11.11	57.67	
Predicted Shares (%)	27.51	0.99	21.78	49.72	
Predicted-Observed Shares	7.14	-9.86	10.67	-7.95	
$\frac{\textbf{Deviation in Shares}}{\textbf{Observed Shares}}$ (%)	35.05	-90.88	96.04	-13.79	
Absolute Sum of Deviations(%)					**35.62**

As seen from the above tables, although the ordered probit model predictions have relatively good correspondence with choices in the holdout sample, the prediction success drops significantly when applied to the Wave 8 forecasts. On the contrary, the conditional model is observed to exhibit relatively better performance.

Predictions using the new sample for the Beta logistic model were conducted next. The results are shown in Table 17.

TABLE 17. Model Validation Results - Wave 5 to Wave 8 Forecast

Modal Transitions	SR→SR	SR→DA	DA→SR	DA→DA	TOTAL
Observed Number	77	41	42	218	378
Predicted Number	33	0	0	192	225
% Correct	42.86	0.00	0.00	88.07	59.52
% Incorrect	57.14	100.00	100.00	11.93	**40.48**
Observed Shares (%)	20.37	10.85	11.11	57.67	
Predicted Shares (%)	23.09	10.00	10.00	56.91	
Predicted-Observed Shares	2.72	-0.85	-1.11	-0.76	
$\frac{\text{Deviation in Shares}}{\text{Observed Shares}}$ (%)	13.35	-7.83	-9.99	-1.32	
Absolute Sum of Deviations(%)					**5.44**

For the forecast sample the total prediction error for aggregate prediction is about 6%, which means the prediction success rate is 94%. In terms of disaggregate predictions the success rate is only about 60%. As in the hold-out sample, no mode-switcher was correctly classified by the model. However, as before, the prediction of propensities matched the observed aggregate shares very closely.

7. CONCLUSIONS

The research presented in this paper focuses on a major weakness of existing models of transportation behavior in adequately addressing time or state dependence of behavior. Dynamic models using panel data appear to offer a promising approach in the study of non-stationary behavioral patterns. Although beta distribution models have been used in the past for behavioral modeling, most are, however, static. Two previous cases of use of beta distribution models in dynamic modeling were observed from the literature research. The extension of static beta-logistic models to dynamic models in those previous studies was achieved only after making an unsupported assumption of logistically scaled posterior probabilities that were assumed to be logistic functions of a reference period probability. In this research the assumption of logistically scaled posterior probabilities was tested empirically and was found to be unjustified for the data set under study. It was found that the beta distribution could account for variation in choice probabilities over the two waves used in this research and no scaling of posterior probabilities was found to be needed.

The validation of the models was carried out with the help of a hold-out sample as well as with a new sample. Both with the hold-out sample, as well as with the new data set, the aggregate prediction success rate was high for the dynamic beta-logistic model. A comparative analysis using more conventional models was also undertaken. It was found that a two-period joint choice probit model performed slightly better than the beta-logistic model in disaggregae prediction. However, for aggregate shares prediction using a forecast sample, the beta-logistic model performed significantly better than the joint choice model. Both of these models outperformed the ordered probit model.

The dynamic model was not able to predict mode switching behavior at the disaggregate level. Since no variables measuring attributes of the alternatives in the choice set were used in the model, it can be speculated that the model performance might improve if such variables are included. Additionally, increasing the number of observations will allow the calibration of models with a better specification of variables that used in the dynamic model reported in this paper. Finally, modeling a choice process with a more evenly balanced split among the alternatives, might produce better prediction results than reported in this paper.

REFERENCES

Bates, G. and J. Neyman, 1951, "Contributions to the Theory of Accident Proneness II: True or False Contagion". *University of California Publications in Statistics,* **1,** 215-253.

Brownstone, D. and X. Chu, 1992, "Multiply Imputed Sampling Weights: A Simple but General Method for Consistent Inference with Panel Attrition". Paper prepared for presentation at the *First U.S. Conference on Panels for Transportation Planning,* Lake Arrowhead, CA.

Cervero, R. and B. Griensenbeck, 1988, "Factors Influencing Commuting Choices in Suburban Labor Markets: A Case Study of Pleasanton, California". *Transportation Research* **22A**, 151-162.

Clark, W.A.V. and J.O. Huff, 1977, "Some empirical tests of duration-of-stay effects in intraurban migration". *Environment and Planning A* **9**, 1357-1374.

Daganzo, C.F. and Y. Sheffi, 1982, "Multinomial probit with time-series data: unifying state dependence and serial correlation models". *Environment and Planning A* **14**, 1377-1388.

Davies, R.B., 1984, "A generalised beta-logistic model for longitudinal data with an application to residential mobility". *Environment and Planning A* **16**, 1375-86.

Davies, R.B. and R. Crouchley, 1984, "Calibrating longitudinal models of residential mobility and migration". *Regional Science and Urban Economics* **14**, 231-247.

Davies, R.B., R. Crouchley, and A.R. Pickles, 1982, "A family of hypothesis tests for a collection of short event series with an application to female employment participation", *Environment and Planning A* **14**, 603-614

Davies, R.B. and A.R. Pickles, 1985, "A panel study of life-cycle effects in residential mobility", *Geographical Analysis* **17 No 3**, 199-216.

Dunn, R., S. Reader, and N. Wrigley, 1987,"A nonparametric approach to the incorporation of heterogeneity into repeated polytomous choice models of urban shopping behaviour". *Transportation Research A* **21A No 4/5**, 327-343.

Dunn, R., and N. Wrigley, 1985, "Beta-logistic models of urban shopping center choice". *Geograhphical Analysis* **17 No 2** , 95-113.

Feeney, B.P., 1989, "A review of the impact of parking policy measures on travel demand". *Transportation Planning and Technology,* **13,** pp 229-244.

Ferguson, E., 1990, "The influence of employer ridesharing programs on employee mode choice". *Transportation,* **17**, 179-207.

Goodwin, P.B., 1989, "Family changes and public transport use 1984-1987 a dynamic analysis using panel data". *Transportation* **16**, 121-154.

Heckman, J.J., 1981a, "Statistical models for discrete panel data".*Structural Analysis of Discrete Data with Econometric Applications*, Eds C F Manski, D McFadden (MIT Press, Cambridge, MA), 114-178.

Heckman, J.J., 1981b, "Heterogeneity and state dependence". *Studies in Labor Markets,* Ed S Rosen (University of Chicago Press, Chicago, Il), 91-139.

Heckman, J.J. and R. Willis, 1977, "A beta-logistic model for the analysis of sequential labor force participation of married women", *Journal of Political Economy* **85**, 27-58.

Hensher, D.A. and V. Le Plastrier, 1985, "Towards a dynamic discrete-choice model of household automobile fleet size and composition". *Transportation Research B* **19B No 6** , 481-495.

Johnson, L. and D. Hensher, 1982, "Application of multinomial probit to a two-period panel data set", *Transportation Research A* **16A No 5-6** , 457-464.

Kitamura, R. and D.S. Bunch, 1990, "Heterogeneity and state dependence in household car ownership: A panel analysis using ordered-response probit models with error components" in *Transportation and Traffic Theory,* Ed M Koshi (Elsevier, New York, NY), 477-496.

Mahmassani, H.S., 1990, "Dynamic models of commuter behavior: experimental investigation and application to the analysis of planned traffic disruptions". *Transportation Research A* **24A No 6** , 465-484.

Mahmassani, H.S. and G.L. Chang, 1986,"Experiments with departure time choice dynamics of urban commuters". *Transportation Research B* **20B No 4**, 297-320.

Small, K.A., 1983, "Bus priority and Congestion Pricing on Urban Expressways",*Research in Transportation Economics*, **1**, 27-74.

Smith, N.C., D.A. Hensher, and N. Wrigley, 1986, "Modelling discrete choice outcome sequences with panel data: an application to automobile transactions". *Dimensions of Automobile Demand Project*, Working Paper **No 18**, (Macquarie University, North Ryde, Australia).

Teal, R.F., 1987, "Carpooling: Who, How, and Why", *Transportation Research,* **21A**, 203-214.

Uhlaner, C.J. and S. Kim, 1992, "Designing and Implementing a Panel Study of Commuter Behavior: Lessons for Future Research", Paper presented at the First U.S. Conference on Panels for Transportation Planning, Lake Arrowhead, CA, October 1992.