DYNAMIC TRAVEL DEMAND MODELS BASED ON LONGITUDINAL PERSON-TRIP DATA

Yoriyasu SUGIE
Professor
Graduate School for International Development and Cooperation
Hiroshima University
Kagamiyama 1-5-1
Higashi-Hiroshima 739-8529 Japan
Fax: +81-824-24-6919
E-mail: ysugie@hiroshima-u.ac.jp

Junyi ZHANG
Pacific Consultants Co., Ltd.
Nishi-Shinjuku 2-7-1, Shinjuku-ku,
Tokyo 163-0730, Japan
Fax: +81-3-3344-1549
E-mail: syunkitsu.chou@tk.pacific.co.jp

Akimasa FUJIWARA
Associate Professor
Graduate School for International Development and Cooperation
Hiroshima University
Kagamiyama 1-5-1
Higashi-Hiroshima 739-8529 Japan
Fax: +81-824-24-6921
E-mail: afujiwar@hiroshima-u.ac.jp

Abstract: This paper aims at improving the first three steps of the traditional four-step transportation model by using longitudinal person-trip data obtained at three different points in time in the Hiroshima metropolitan area. Important results can be obtained. Cross-sectional assumptions implicit in traditional travel demand models such as temporal stability, homogeneity and serial independence are all statistically rejected. Dynamic models with fixed-effects and random-effects are developed based on the statistical results of trip generation, attraction and distribution models. Through empirical analysis, newly developed dynamic models have proven to be superior to traditional ones in terms of prediction accuracy. Further, an aggregate logit model (linear form) is employed for modal split. Finally, a dynamic simultaneous-equation model with fixed-effects based on the seemingly unrelated regressions method is developed and its effectiveness is confirmed by empirical analysis.

Key Words: dynamic model, unobserved heterogeneity, serial correlation

1. INTRODUCTION

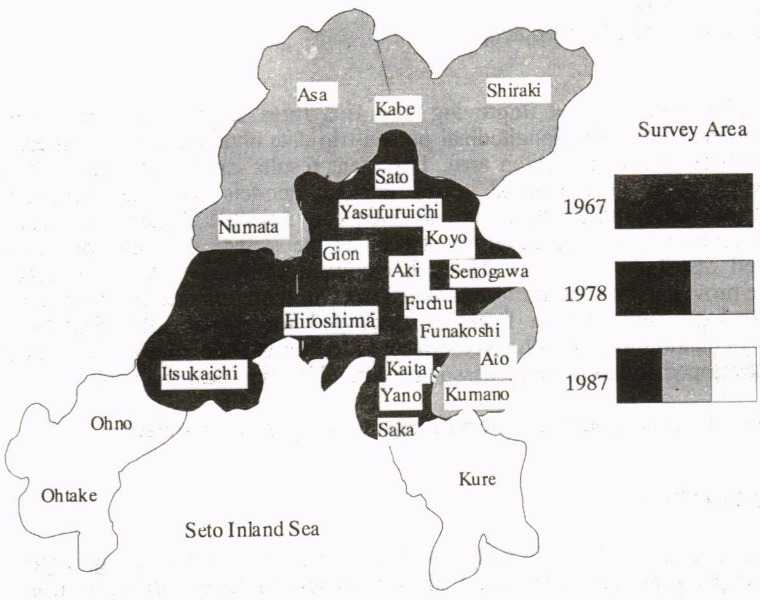
Cross-sectional data has been broadly used in travel demand modeling, especially urban transportation planning. However, there still remain several severe problems from a practical point of view. For example, models using cross-sectional data cannot provide travel information on temporal change, thereby reducing longer-term travel demand prediction accuracy.

To alleviate these infirmities, longitudinal data collected at multiple points in time has come to the fore of travel behavior research. According to whether or not samples surveyed are identical over time, longitudinal data can be generally classified as panel data or repeated cross-sectional data, respectively.

Although longitudinal data has its own specific problems such as expensive survey costs and attrition bias caused by repetitious surveying, its information provisional capabilities, especially regarding the temporal change of travel behavior, surpasses that of cross-sectional data. Unfortunately, most research thus far has been confined to individual behavioral analyses based on panel data in which the time-span between two surveys is very short, e.g. half a year or one year (e.g., Sugie *et al.*, 1999). Travel demand models at the zonal level using longitudinal data collected over multi-year intervals have not been satisfactorily developed, probably because a sufficient number of data sets could not be easily obtained for the area in question.

The goal of this paper is to improve the first three steps of the conventional four-step model using longitudinal travel data obtained in the Hiroshima metropolitan area during 1967, 1978 and 1987. Urbanization has expanded survey area size, hence the initial 32-zone area common to all survey years is used for this study (Sugie *et al*, 1982) (see Figure 1). The data originates from repeated cross-sectional data gathered at 10-year intervals. Though individuals sampled in the survey are different at each point in time, the analysis unit (i.e. zone) is fixed over the duration of the survey. Therefore, statistical methods developed for the analysis of individual panel data can be applied at the zonal level (Ito *et al.*, 1997).

In the field of travel behavior research, perhaps the most frequent reason that motivates a panel study is the evaluation of the impact of a change in the transportation system, or a specific transportation planning project (Kitamura, 1990). Accordingly, much research has been dedicated to disaggregate travel behavior using panel data (Special Issue: Longitudinal Data Methods, 1987; Special Issue: Panel Analysis of Travel Demand, 1989; Special Issue: Dynamic Travel Behavior Analysis, 1990), and useful results have been obtained.



Survey year	1967	1978	1987
Sampling rate (%)	5.0	1.5	7.5
Area (square kilometers)	413	850	1,151
Population (thousands)	770	1,060	1,580
Number of zones	110	40	196

Figure 1 Expansion of the Hiroshima Metropolitan Area

Most dynamic models in individual travel behavioral analysis have been developed using short-term panel data. However, if we consider forecasting transportation conditions 10 to 20 years down the road using dynamic models, assumptions at time t, which are functions of dependent variables at time t-1, are not completely plausible. Therefore, it seems critical to consider time series factors for studies conducted over long time intervals with few survey years. The objective of this study is to develop dynamic travel demand models incorporating unobserved heterogeneity and first-order serial correlation within the context of such a circumstance.

As for the format of this paper, section 2 is used to statistically test cross-sectional assumptions for trip generation, attraction and distribution models. Based on the test results, dynamic single-equation models integrating unobserved heterogeneity and first-order serial correlation are developed in section 3. Finally, section 4 is used to develop a new dynamic modal split model with simultaneous-equations, as it is not realistic to treat the error terms of different modes independently.

2. STATISTICAL TEST OF CROSS-SECTIONAL ASSUMPTIONS

2.1 Cross-sectional Assumptions

Traditional travel demand models using cross-sectional data can be expressed as follows:

$$y_{it} = \mu + \sum_{k=1}^{K} \beta_k x_{k,it} + v_{it} \tag{1}$$

where,

- i, t : zone (or zone pair) and time, respectively
- y_{it} : dependent variable (e.g. generated trips)
- $x_{k,it}$: k'th explanatory variable of y_{it}
- β_k : parameter of $x_{k,it}$
- μ : some constant
- v_{it} : error term having an identical and independent distribution (i.i.d.) for i and t
- K : total number of explanatory variables

The following is assumed for Eq. (1).

Assumption 1: temporal stability, i.e. β_k is independent of time.

Assumption 2: homogeneity, i.e. μ is constant across zones.

Assumption 3: serial independence of v_{it}.

Based on the above assumptions, Eq. (1) can be estimated using the ordinary least squares (OLS) method. However, if these assumptions do not hold, using the estimation results based on OLS will lead to erroneous conclusions.

2.2 Estimation of Trip Generation, Attraction and Distribution Models

In this section, traditional travel demand models are developed for statistical analysis. The indices related to population and employment in industry, business and commerce are used as explanatory variables for trip generation and attraction models expressed as Eq. (1).

For trip distribution, a gravity model (Eq. (2)) is employed to check the temporal stability of model parameters.

$$y_{ijt} = \mu(G_{it})^{\beta_G}(A_{jt})^{\beta_A} / (T_{ijt})^{\beta_T} \tag{2}$$

where,

y_{ijt} : interzonal trips between zones i and j at time t

G_{it} : trips generated in zone i

A_{jt} : trips attracted in zone j

T_{ijt} : a friction factor measuring average travel time between zones i and j

$\mu, \beta_G, \beta_A, \beta_T$: model parameters

This model is widely used in Japan. For computational convenience, Eq. (2) can be converted to its log-linear form as follows:

$$\ln(y_{ijt}) = \ln(\mu) + \beta_G \ln(G_{it}) + \beta_A \ln(A_{jt}) - \beta_T \ln(T_{ijt}) + v_{ijt} \tag{3}$$

This indicates that the above trip distribution model can be also expressed as Eq. (1). Accordingly, generation/attraction and distribution models are estimated using OLS (only the results with respect to total trip purpose are shown in Table 1 due to limited space). The sample size for trip distribution is smaller than expected (32x32=1,024), because intrazonal as well as zero-trip samples elicited during survey year 1978 (a mere 1.5% sampling rate) were excluded from the analysis. It can be seen that each model has an excellent goodness-of-fit (i.e. multiple correlation coefficient) and that population as well as business and commerce employment variables are significant in the generation/attraction models.

2.3 Test of Temporal Stability

To test whether or not the estimated parameters based on OLS are temporally stable, we use a covariance analysis method (Hsiao, 1986). First, we estimate Eq. (4) using OLS for each year.

$$y_{it} = \mu_t + \sum_{k=1}^{K} \beta_{k,t} x_{k,it} + v_{it} \tag{4}$$

Constant μ_t and parameter $\beta_{k,t}$ vary over time and their residual sum of squares can be calculated as S_1. Next, we estimate Eq. (1) using OLS and pooled data from 1967 and 1978, then calculate its residual sum of squares as S_2. The hypothesis of temporal stability for constantst and parameters can be thought of as Eq. (4) subject to (k+1)(T-1) linear restrictions:

$$H_0: \mu_1 = \mu_2 = \cdots = \mu_T \text{ and } \beta_{k,1} = \beta_{k,2} = \cdots = \beta_{k,T}$$

Based on S_1 and S_2, the following F-statistic can be employed to test temporal stability.

$$F = \frac{(S_2 - S_1)/[(T-1)(K+1)]}{S_1/[NT - T(K+1)]} \tag{5}$$

The test results of Eq. (5) are shown in Table 2. Test results for the school attraction model is not indicated because the number of students in 1967, an important explanatory variable in the model, cannot be obtained. From Table 2, it is obvious that temporal stability for all of the models is convincingly rejected.

Table 1 Estimation Results for Generation, Attraction and Distribution Models (Total Trip Purpose)

Explanatory variable	Trip generation 1967	Trip generation 1978	Trip generation 1987	Trip attraction 1967	Trip attraction 1978	Trip attraction 1987
Constant	-434 (0.13)	-3280 (1.26)	561 (0.17)	-502 (0.15)	-3335 (1.27)	597 (0.18)
Population	2.215 (17.1)**	1.750 (24.6)**	1.660 (23.4)**	2.210 (17.1)**	1.745 (24.4)**	1.655 (23.2)**
Employment in business and commerce	3.540 (29.3)**	2.190 (32.3)**	2.790 (25.7)**	3.561 (29.4)**	2.198 (32.1)**	2.804 (25.9)**
Sample size	32	32	32	32	32	32
Multiple correlation coefficient	0.990	0.990	0.995	0.990	0.990	0.988

Explanatory variable	Trip distribution 1967	Trip distribution 1978	Trip distribution 1987
Constant	-8.837 (9.11)**	-12.38 (8.80)**	-5.609 (6.70)**
Generated trips	0.876 (16.3)**	0.918 (11.5)**	0.938 (20.0)**
Attracted trips	1.106 (19.8)**	1.245 (14.0)**	0.827 (15.9)**
Average travel time	-1.792 (23.6)**	-1.235 (14.3)**	-1.962 (30.9)**
Sample size	458	458	458
Multiple correlation coefficient	0.891	0.739	0.881

(t scores in parentheses; *: significant at 5%, **: 1%)

Table 2 Temporal Stability Test Results

Trip purpose	Generation	Attraction	Distribution
Work	F(2, 60) = 8.09**	F(2, 60) = 3.94*	F(4, 802) = 5.62**
School	F(2, 60) = 5.29**		F(4, 434) = 5.48**
Home	F(3, 58) = 25.2**	F(2, 60) = 54.0**	F(4, 866) = 24.2**
Shopping	F(3, 58) = 174**	F(3, 58) = 46.7**	F(4, 208) = 21.9**
Personal	F(3, 58) = 126**	F(3, 58) = 127**	F(4, 540) = 35.1**
Business	F(3, 58) = 92.2**	F(3, 58) = 88.2**	F(4, 632) = 46.3**
Total	F(3, 58) = 123**	F(3, 58) = 123**	F(4, 908) = 32.3**

(Figures in F(,) are degree of freedom; *: significant at 5%; **: 1%)

2.4 Testing for Homogeneity

Consider the following equation with fixed-effects parameter δ_i.

$$y_{it} = \delta_i + \mu + \sum_{k=1}^{K} \beta_k x_{k,it} + u_{it} \tag{6}$$

The test of homogeneity determines whether or not the null hypothesis H_0: $\delta_i = 0$ holds. We estimate, first of all, the pooled model (Eq. (6)) in which $\delta_i = 0$ using OLS and obtain the estimated residual $\hat{u}_{it}$. Then the following Breusch-Pagan statistic λ can be used to test for homogeneity (Maddala, 1987; Meurs, 1990).

$$\lambda = \frac{NT}{2(T-1)}\left[\sum_{i=1}^{N}\left[\sum_{t=1}^{T}\hat{u}_{it}\right]^2 / \sum_{i=1}^{N}\sum_{t=1}^{T}\hat{u}_{it}^2 - 1\right]^2 \quad (7)$$

The λ statistic follows a χ^2 distribution with one degree of freedom when N is sufficiently larger than 1. The test results based on λ are shown in Table 3, and one can deduce that the existence of heterogeneity in most models is accepted.

Table 3 Homogeneity Test Results

Model	Work	School	Home	Shopping	Personal	Business	Total
Generation	0.013	0.737	2.533	21.1**	22.4**	14.5**	18.7**
Attraction	0.002		3.691	7.52**	20.4**	15.2**	18.7**
Distribution	49.2**	19.0**	143**	3.616	5.28*	2.514	42.8**

(*: significant at 5%; **: 1%)

2.5 Testing for Serial Independence

Here we test for the existence of serial correlation of error terms in the presence of heterogeneity. Therefore, we assume the following error structure:

$$u_{it} = \rho\, u_{it-1} + e_{it} \quad (8)$$

where ρ is a first-order serial correlation coefficient satisfying stationarity assumption $|\rho| < 1$.

By estimating Eqs. (6) and (8) with OLS when null hypothesis H_0: $\rho = 0$ holds, we can obtain the estimated residual $\hat{u}_{it}$ and establish the following generalized Durbin-Watson statistic (Bhargava *et al.*, 1982; Maddala, 1987).

$$DW = \sum_{i=1}^{N}\sum_{t=2}^{T}\left(\hat{u}_{it} - \hat{u}_{it-1}\right)^2 / \sum_{i=1}^{N}\sum_{t=1}^{T}\hat{u}_{it}^2 \quad (9)$$

The test results using Eq. (9) shown in Table 4 indicate the existence of first-order serial correlation in all of the models at significance levels of 5% (the critical value is approximately 2.00).

Table 4 Serial Independence Test Results

Model	Work	School	Home	Shopping	Personal	Business	Total
Generation	0.98*	1.15*	1.28*	1.81*	1.84*	1.67*	1.77*
Attraction	0.99*		1.34*	1.48*	1.80*	1.69*	1.76*
Distribution	0.65*	0.71*	0.43*	0.82*	0.86*	1.09*	0.69*

(*: significant at 5%)

3. DYNAMIC MODELS INCORPORATING UNOBSERVED HETEROGENEITY AND FIRST-ORDER SERIAL CORRELATION

The above test results would impel one to relax all three cross-sectional assumptions. However, because longitudinal data used here comprises only three time points, it is not possible to incorporate time-varying parameters into the models. For this reason, we develop dynamic models that incorporate heterogeneity and first-order serial correlation simultaneously for generation, attraction and distribution models. The general formulae can be represented as follows (Bhargava *et al.*, 1982; Hsiao, 1986):

$$y_{it} = \mu + \sum_{k=1}^{K} \beta_k x_{k,it} + v_{it} \tag{10}$$

$$v_{it} = \delta_i + u_{it} \tag{11}$$

$$u_{it} = \rho\, u_{it-1} + e_{it} \tag{12}$$

where v_{it}, u_{it}, e_{it} are error terms with e_{it} having an i.i.d.

The initial condition for Eqs. (10) ~ (12) is given as (Lillard *et al*, 1978):

$$u_{i1} = e_{i1} / \sqrt{1-\rho^2} \tag{13}$$

According to the assumptions about δ_i, we can obtain a model with fixed-effects (i.e. δ_i does not change stochastically) and a model with random-effects (i.e. δ_i is a random variable). Because error term u_{it} has a first-order serial correlation, the generalized least squares (GLS) method can be applied. The GLS estimator can be defined as (Amemiya, 1985):

$$\hat{\boldsymbol{\beta}} = \left[\mathbf{X}^{*\prime} \boldsymbol{\Omega}^{*-1} \mathbf{X}^{*} \right]^{-1} \mathbf{X}^{*\prime} \boldsymbol{\Omega}^{*-1} \mathbf{y} \tag{14}$$

where $\boldsymbol{\Omega}^* = \left(\mathbf{I}_N \otimes \boldsymbol{\Omega}\right)$ is a NT × NT matrix and $\boldsymbol{\Omega}$ is a T × T variance-covariance matrix of the stationary first-order auto regression, meaning $\boldsymbol{\Omega}$ has elements of the form (Bhargava *et al.*, 1982):

$$\omega_{ts} = \rho^{|t-s|} / \left(1-\rho^2\right) \tag{15}$$

Eq. (14) is expressed rather abstrusely, and practical estimation would be complex. In light of this, we set about transforming Eqs. (10) ~ (12) using a simple method.

3.1 Specification of Dynamic Models with Fixed-effects (DFIX)

Based on the above theoretical background, Eqs. (10) ~ (12) can be transformed as follows:

$$\sqrt{1-\rho^2}\,(y_{i1} - \bar{y}_i) = \sum_{k=1}^{K} \left[\sqrt{1-\rho^2}\, \beta_k (x_{k,i1} - \bar{x}_{k,i}) \right] + \varepsilon_{i1} \tag{16}$$

$$(y_{it} - \bar{y}_i) - \rho\,(y_{it-1} - \bar{y}_i) = \sum_{k=1}^{K} \beta_k \left[(x_{k,it} - \bar{x}_{k,i}) - \rho\,(x_{k,it-1} - \bar{x}_{k,i}) \right] + \varepsilon_{it} \tag{17}$$

where,

$$\varepsilon_{i1} = e_{i1} - \sqrt{1-\rho^2}\,\bar{u}_i\,, \quad \varepsilon_{it} = e_{it} - (1-\rho)\,\bar{u}_i$$

$$\bar{y}_i = \frac{1}{T}\sum_{t=1}^{T} y_{it}, \ \bar{x}_{k,i} = \frac{1}{T}\sum_{t=1}^{T} x_{k,it}$$

Because the error term ε_{it} ($t = 1, 2, \ldots, T$) is serially independent, OLS can be applied to Eqs. (16) and (17). However, for a study with a minimum number of survey years we propose $\bar{y} = 1/NT\sum_{i=1}^{N}\sum_{t=1}^{T} y_{it}$ and $\bar{x}_k = 1/NT\sum_{i=1}^{N}\sum_{t=1}^{T} x_{k,it}$ instead of $\bar{y}_i$ and $\bar{x}_{k,i}$ because the former can increase degrees of freedom for the estimation.

In order to estimate μ and δ_i separately, Hsiao (1986) assumes $\sum_{i=1}^{N} \delta_i = 0$. Using OLS to get the estimated value of β_k ($\hat{\beta}_k$) from Eqs. (16) and (17), along with $\bar{y}$ and $\bar{x}_k$, we can calculate the estimated values $\hat{\mu}$, $\hat{\delta}_i$ of μ, δ_i as follows:

$$\hat{\mu} = \bar{y} + \sum_{k=1}^{K} \hat{\beta}_k \bar{x}_k, \ \hat{\delta}_i = \bar{y}_i - \hat{\mu} - \sum_{k=1}^{K} \hat{\beta}_k \bar{x}_{k,i} \tag{18}$$

In fact, a consistent estimator of ρ must be pre-determined by the estimated parameter of y_{it-1} in Eq. (19), because it cannot be obtained directly from Eqs. (16) and (17) using OLS with an insufficient number of survey years.

$$y_{it} = \alpha + \rho\, y_{it-1} + \sum_{k=1}^{K} \left[\beta_k x_{k,it} + \gamma_k x_{k,it-1} \right] + e_{it} \tag{19}$$

Finally, the estimated value $\hat{y}_{it}$ of y_{it} can be expressed as a function of y_{it-1}, $x_{k,it-1}$ as well as $x_{k,it}$ as follows:

$$\hat{y}_{it} = \hat{\rho}\, y_{it-1} + (1 - \hat{\rho})(\hat{\mu} + \hat{\delta}_i) + \sum_{k=1}^{K} \left[\hat{\beta}_k (x_{k,it} - \hat{\rho}\, x_{k,it-1}) \right] \tag{20}$$

3.2 Specification of Dynamic Models with Random-effects (DRAN)

In contrast with DFIX, the variance-covariance matrix $\mathbf{\Omega}$ of error term v_{it} in DRAN is defined as follows (Lillard *et al.*, 1978):

$$\mathbf{\Omega} = \sigma_u^2 \begin{bmatrix} 1 & \rho & \rho^2 \cdots & \rho^{T-1} \\ \rho & 1 & \rho \cdots & \cdot \\ \rho^2 & \rho & 1 \cdots & \cdot \\ \vdots & \vdots & \vdots \ddots & \vdots \\ \rho^{T-1} & \cdot & \cdot \cdots & 1 \end{bmatrix} + \sigma_\delta^2\, \mathbf{ii}' \tag{21}$$

where σ_u^2, σ_δ^2 are variances of error terms u_{it} and δ_i, and $\mathbf{i}$ is a $T \times 1$ matrix in which all of the elements are 1.

Since substituting $\mathbf{\Omega}$ into Eq. (14) will cause the same problem as in DFIX, we propose another transformation method to specify DRAN.

$$y_{i1} = \mu + \sum_{k=1}^{K} \beta_k x_{k,i1} + \eta_{i1} \tag{22}$$

$$\frac{1}{1-\rho} y_{it} - \frac{\rho}{1-\rho} y_{it-1} = \mu + \sum_{k=1}^{K} \beta_k \left(\frac{1}{1-\rho} x_{k,it} - \frac{\rho}{1-\rho} x_{k,it-1}\right) + \eta_{it} \tag{23}$$

where,

$$\eta_{i1} = u_{i1} + \delta_i \,,\; \eta_{it} = e_{it} / (1-\rho) + \delta_i$$

The error term η_{it} (t = 1, 2, ..., T) has the following variance-covariance matrix.

$$\mathbf{\Psi} = \begin{bmatrix} \sigma_1^2 & \sigma_{cov} & \sigma_{cov} & \cdots & \sigma_{cov} \\ \sigma_{cov} & \sigma_2^2 & \sigma_{cov} & \cdots & \sigma_{cov} \\ \sigma_{cov} & \sigma_{cov} & \sigma_3^2 & \cdots & \sigma_{cov} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma_{cov} & \sigma_{cov} & \sigma_{cov} & \cdots & \sigma_T^2 \end{bmatrix} \tag{24}$$

where,

$$\sigma_2^2 = \sigma_3^2 = \cdots = \sigma_T^2 = \sigma_e^2 / (1-\rho)^2 + \sigma_\delta^2 \,,\; \sigma_{cov} = \sigma_\delta^2$$

σ_e^2 is the variance of error term e_{it}.

Eq. (24) is a special case of the GLS error structure. When T = 2, it follows the structure of the seemingly unrelated regressions (SUR) method (Zellner, 1962).

Similar to DFIX, the estimated value of y_{it} $(\hat{y}_{it})$ can also be calculated based on travel information at a previous time point.

$$\hat{y}_{it} = \hat{\rho}\, y_{it-1} + \hat{\mu}\,(1-\hat{\rho}) + \sum_{k=1}^{K} \hat{\beta}_k \,(x_{k,it} - \hat{\rho}\, x_{k,it-1}) \tag{25}$$

3.3 Estimation of DFIX and DRAN

In this section, we estimate DFIX and DRAN using data in 1967 and 1978 (only total trip purpose estimations are shown). It is obvious that most of the estimated parameters have the expected signs and are statistically significant.

To check the significance of DFIX and DRAN, we use the estimated parameters in Table 5 to predict travel demand in 1987 and then compare them with the predictions of models OLS-78, SUR-78 and FSUR-78 (see Table 6).

OLS-78 is a traditional prediction model, which extrapolates future conditions from present cross-sectional relationships, hence parameters of the base year (1978 in this study) are used for prediction. SUR-78 considers temporal variation of parameters, zonal variation of constants and arbitrary serial correlation. The difference between SUR-78 and FSUR-78 is that the latter does not assume parameters to vary over time. They use data in 1967 and 1978 for model estimation. However, since it is not clear how correlation between present and future error terms is considered, it is not addressed in this paper. Comparing these two models can make it clear whether or not time-varying parameters significantly influence prediction accuracy.

Table 5 Estimation Results for Generation, Attraction and Distribution Models (Total Trip Purpose)

Explanatory variable	Generation DFIX	Generation DRAN	Attraction DFIX	Attraction DRAN	Distribution DFIX	Distribution DRAN
Constant		-2099 (0.64)		-2157 (0.65)		-11.78 (12.2)**
Population	1.694 (8.10)**	1.762 (19.1)**	1.690 (8.03)**	1.754 (18.9)**		
Employment in business and commerce	2.476 (10.6)**	2.299 (26.5)**	2.484 (10.6)**	2.302 (26.3)**		
Trips generated					0.878 (17.3)**	1.057 (19.8)**
Trips attracted					0.932 (17.3)**	0.884 (14.8)**
Average travel time					-1.014 (15.8)**	-0.744 (11.3)**

(t scores in parentheses; *: significant at 5%; **: 1%)

Table 6 Prediction Models for 1987

Prediction model	Variation $\mu+\delta i$	Variation βt	Serial correlation	Estimation method
OLS-78	no	no	no	OLS
SUR-78	yes	yes	yes/no[a)]	SUR[b)]
FSUR-78	yes	no	yes/no[a)]	SUR[b)]
DFIX	yes	no	yes (1st)[c)]	OLS
DRAN	yes	no	yes (1st)[c)]	GLS

Prediction model	Data for model estimation	Parameters used for prediction
OLS-78	1978	1978
SUR-78	1967+1978	1978
FSUR-78	1967+1978	1967+1978[d)]
DFIX	1967+1978	1967+1978[d)]
DRAN	1967+1978	1967+1978[d)]

a) considered in model estimation, but not for prediction b) see section 3.2
c) first-order serial correlation d) commom parameters for 1967 and 78

Goodness-of-fit indices used to evaluate prediction accuracy regarding actual (Y_i) and estimated ($\hat{Y}_i$) trips in 1987 are correlation coefficient (R) and Theil's inequality coefficient (U_t: 0 ≤ its value ≤ 1) a (Theil, 1961). U_t can be expressed as Eq. (26).

$$U_t = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(Y_i - \hat{Y}_i)^2} \Big/ \left(\sqrt{\frac{1}{N}\sum_{i=1}^{N} Y_i^2} + \sqrt{\frac{1}{N}\sum_{i=1}^{N} \hat{Y}_i^2}\right) \tag{26}$$

A larger value of R and a smaller value of U_t means higher prediction accuracy. The prediction accuracy of each model defined in Table 6 is shown in Table 7.

We can see from Table 7 that FSUR-78 is superior to OLS-78 and SUR-78 in terms of model accuracy. This result means that zone-dependent constants (i.e. $\mu + \delta_i$) are more important than time-varying parameters (i.e. β_t), supporting the assumptions of DFIX and DRAN.

As heterogeneity parameters represent travel change due to unmeasurable zonal (or spatial) characteristics, they must be more effective than time-varying parameters. Moreover, incorporating first-order serial correlation into DFIX and DRAN makes it possible to consider travel information at previous time points explicitly. As a result, DFIX and DRAN are most accurate of all the models defined in Table 6. Besides, since the heterogeneity parameter can be explicitly incorporated in DFIX, it is more desirable to use this model to predict travel demand rather than DRAN.

Table 7 Prediction Accuracy of the Models Defined in Table 6
(Total Trip Purpose)

Model	Generation		Attraction		Distribution	
	R	Theil's	R	Theil's	R	Theil's
OLS-78	0.977	0.069	0.976	0.070	0.823	0.045
SUR-78	0.978	0.071	0.978	0.071	0.801	0.046
FSUR-78	0.983	0.058	0.983	0.058	0.840	0.044
DFIX	0.992	0.031	0.992	0.031	0.884	0.038
DRAN	0.982	0.058	0.982	0.059	0.879	0.035

4. DYNAMIC MODELS WITH SIMULTANEOUS-EQUATIONS

In this section, we extend the dynamic single-equation models of section 3 to the modal split phase. Although a number of modal split models have been employed in travel demand analysis, a logit model such as Eq. (27) is used here because it is more theoretically well-founded than other models.

$$P^m_{ij,t} = \exp\left(V^m_{ij,t}\right) / \sum\nolimits_{m'=1}^{M} \exp\left(V^{m'}_{ij,t}\right) \tag{27}$$

$$V^m_{ij,t} = \alpha^m + \sum\nolimits_{k=1}^{K_1} \beta_k x^m_{k,ij,t} + \sum\nolimits_{k=K_1+1}^{K} \beta^m_k x_{k,ij,t} \tag{28}$$

where,

- $P^m_{ij,t}$: trip share held by mode m between zones i and j at time t
- $V^m_{ij,t}$: linear utility function of mode m
- $x^m_{k,ij,t}$: k'th explanatory variable of mode m (e.g. average travel time)
- β_k : parameter common to all modes
- $x_{k,ij,t}$: explanatory variable common to all modes
- β^m_k : parameter for mode m
- α^m : constant for mode m

There exist two methods to estimate Eq. (27): one is the maximum likelihood (ML) method, and the another is GLS (or SUR). We adopt the latter (SUR) here because it incorporates time series information into the model better than the ML method. For travel modes car, bus and rail, Eq. (27) can be transformed as follows (Theil, 1969) :

$$\ln\left(P^{BUS}_{ij,t} / P^{CAR}_{ij,t}\right) = V^{BUS}_{ij,t} - V^{CAR}_{ij,t} + \omega_{ij,t} \tag{29}$$

$$\ln\left(P_{ij,t}^{RAIL} / P_{ij,t}^{CAR} \right) = V_{ij,t}^{RAIL} - V_{ij,t}^{CAR} + \eta_{ij,t} \tag{30}$$

The explanatory variable common to all modes is average travel time, and the following variables are used independently in the two equations:

1) accessibility Ψ_{it} (i.e. $\sum_{j=1}^{N} A_{jt}/T_{ijt}$) of origin zone i;

2) egressibility Ψ_{jt} (i.e. $\sum_{i=1}^{N} G_{it}/T_{ijt}$) of destination zone j;

3) car ownership in origin zone i;

4) business and commerce employment percentages at destination zone j (an indicator of parking difficulty);

where G_{it}, A_{jt} and T_{ijt} are defined as in Eq. (2).

The models developed in section 3 belong to the single-equation approach. For the modal split model, we must estimate Eqs. (29) and (30) simultaneously to determine the correlation between error terms $\omega_{ij,t}$ and $\eta_{ij,t}$.

4.1 Test of Cross-sectional Assumptions for the Modal Split Model

To carry out the tests we use total trip purpose data from 1967 and 1978. The covariance analysis method is handy to test temporal stability as was done in section 3. Applying the same method to Eqs. (29) and (30) would be too complicated, hence we first estimate Eqs. (29) and (30) using SUR for each year (see Table 8). The sample size decreases due to reasons shown in Table 1. Models obtained have relatively high multiple correlation coefficients, but their parameters seem to vary over time.

It was then tested whether or not the parameters' t-statistics each year were equal (see Table 9). It is clear that most of the parameters are significantly different for survey years 1967 and 1978. We used the same statistics employed in the previous section to test the homogeneity and serial independence assumptions, but the estimated residuals used here are from simultaneous SUR estimations of Eqs. (29) and (30), not from a separate OLS estimation. The test results of Table 10 indicate that all estimations are statistically rejected at significance levels of 5% or 1%, suggesting the existence of heterogeneity and first-order serial correlation.

Much like section 3, we can use the above test results to rewrite Eqs. (29) and (30) as follows:

$$\ln\left(P_{ij,t}^{BUS} / P_{ij,t}^{CAR} \right) = \delta_{ij}^{BC} + V_{ij,t}^{BUS} - V_{ij,t}^{CAR} + \omega_{ij,t} \tag{31}$$

$$\ln\left(P_{ij,t}^{RAIL} / P_{ij,t}^{CAR} \right) = \delta_{ij}^{RC} + V_{ij,t}^{RAIL} - V_{ij,t}^{CAR} + \eta_{ij,t} \tag{32}$$

$$\omega_{ij,t} = \rho^{BC}\, \omega_{ij,t-1} + \varepsilon_{ij,t}^{BC} \tag{33}$$

$$\eta_{ij,t} = \rho^{RC}\, \eta_{ij,t-1} + \varepsilon_{ij,t}^{RC} \tag{34}$$

where, δ_{ij}^{BC}, δ_{ij}^{RC} : heterogeneity parameters
ρ^{BC}, ρ^{RC} : first-order serial correlation coefficients

Table 8 SUR Estimation Results of Eqns (29) and (30) for Each Survey Year

Explanatory variable	1967	1978	1987
Average travel time (min.)	-1.61E-03 (0.36)	-2.08E-03 (1.25)	-3.10E-02 (4.59)**
Eqn (29)			
Constant	0.761 (1.56)	0.428 (0.76)	-1.58 (2.37)*
Accessibility of origin zone	7.54E-07 (0.81)	2.66E-06 (1.93)	4.50E-06 (4.13)**
Egressibility of destination zone	-9.36E-06 (2.09)*	-5.88E-06 (0.75)	-5.92E-06 (0.85)
Car ownership in origin zone	-2.35 (1.54)	-4.62 (2.73)**	-4.65 (2.81)**
Business and commerce employment percentages in destination zone	1.40 (4.28)**	1.49 (3.97)**	3.28 (6.72)**
Eqn (30)			
Constant	1.35 (1.76)	0.358 (0.46)	0.382 (0.44)
Accessibility of origin zone	-6.20E-06 (4.15)**	-7.42E-07 (0.39)	-1.37E-06 (0.97)
Egressibility of destination zone	-2.43E-05 (3.42)**	-2.53E-05 (2.30)*	-2.01E-05 (2.21)*
Car ownership in origin zone	1.07 (0.44)	-2.69 (1.15)	-0.452 (0.21)
Business and commerce employment percentages in destination zone	-0.219 (0.42)	1.11 (2.20)*	0.946 (1.48)
Sample size	126	126	126
Multiple correlation coefficient	0.672	0.763	0.859

(t scores in parentheses; *: significant at 5%; **: 1%)

Table 9 Test Results for Temporal Stability

Explanatory variable	1967 vs. 1978	
Average travel time	1.13	
	eqn (29)	eqn (30)
Constant	5.03**	10.1**
Accessibility of origin zone	12.8**	25.2**
Egressibility of destination zone	4.32**	0.810
Car ownership at origin zone	11.2**	12.5**
Business and commerce employment percentages in destination zone	2.14*	20.6**

(*: significant at 5%; **: 1%)

Table 10 Test Results for Heterogeneity and First-order Serial Correlation

Equation	Heterogeneity	Serial correlation
(29)	11.5**	0.699*
(30)	13.0**	0.827*

(*: significant at 5%; **: 1%)

4.2 Specification and Estimation of Dynamic Simultaneous-equation Modal Split Models with Fixed-effects (DSEFIX)

We can develop dynamic simultaneous-equations models for modal split using the same method demonstrated in section 3. However, it becomes very complicated to extend DRAN to simultaneous-equations due to error structure complexity. Therefore, we only discuss dynamic models with fixed-effects. Eqs. (31) ~ (34) can be transformed as follows:

$$y'^{BC}_{ij,t} = \sum_{k=1}^{K^{BC}} \beta_k^{BC} x'^{BC}_{k,ij,t} + \varepsilon^{BC}_{ij,t} \tag{35}$$

$$y'^{RC}_{ij,t} = \sum_{k=1}^{K^{RC}} \beta_k^{RC} x'^{RC}_{k,ij,t} + \varepsilon^{RC}_{ij,t} \tag{36}$$

where $y'^{BC}_{ij,t}$, $y'^{RC}_{ij,t}$, $x'^{BC}_{k,ij,t}$, $x'^{RC}_{k,ij,t}$ are transformed variables of $\ln(P^{BUS}_{ij,t} / P^{CAR}_{ij,t})$ in Eq. (31), $\ln(P^{RAIL}_{ij,t} / P^{CAR}_{ij,t})$ in Eq. (32) and their explanatory variables, respectively. These variables can be expressed similar to Eqs. (16) and (17) using the average values of i, j and t.

$$y'^{BC}_{ij,t} = \begin{cases} \sqrt{1-(\rho^{BC})^2}\,(y^{BC}_{ij,1} - \bar{y}^{BC}) & \text{if } t = 1 \\ (y^{BC}_{ij,t} - \bar{y}^{BC}) - \rho^{BC}(y^{BC}_{ij,t-1} - \bar{y}^{BC}) & \text{if } t > 1 \end{cases} \tag{37}$$

$$x'^{BC}_{k,ij,t} = \begin{cases} \sqrt{1-(\rho^{BC})^2}\,(x^{BC}_{k,ij,1} - \bar{x}^{BC}) & \text{if } t = 1 \\ (x^{BC}_{k,ij,t} - \bar{x}^{BC}) - \rho^{BC}(x^{BC}_{k,ij,t-1} - \bar{x}^{BC}) & \text{if } t > 1 \end{cases} \tag{38}$$

$$y'^{RC}_{ij,t} = \begin{cases} \sqrt{1-(\rho^{RC})^2}\,(y^{RC}_{ij,1} - \bar{y}^{RC}) & \text{if } t = 1 \\ (y^{RC}_{ij,t} - \bar{y}^{RC}) - \rho^{RC}(y^{RC}_{ij,t-1} - \bar{y}^{RC}) & \text{if } t > 1 \end{cases} \tag{39}$$

$$x'^{RC}_{k,ij,t} = \begin{cases} \sqrt{1-(\rho^{RC})^2}\,(x^{RC}_{k,ij,1} - \bar{x}^{RC}) & \text{if } t = 1 \\ (x^{RC}_{k,ij,t} - \bar{x}^{RC}) - \rho^{RC}(x^{RC}_{k,ij,t-1} - \bar{x}^{RC}) & \text{if } t > 1 \end{cases} \tag{40}$$

where,

$$y^{BC}_{ij,t} = \ln(P^{BUS}_{ij,t} / P^{CAR}_{ij,t}) \qquad y^{RC}_{ij,t} = \ln(P^{RAIL}_{ij,t} / P^{CAR}_{ij,t})$$

$$\bar{y}^{BC} = \frac{1}{NT}\sum_{ij=1}^{N}\sum_{t=1}^{T} y^{BC}_{ij,t} \qquad \bar{y}^{RC} = \frac{1}{NT}\sum_{ij=1}^{N}\sum_{t=1}^{T} y^{RC}_{ij,t}$$

$$\bar{x}_k^{BC} = \frac{1}{NT}\sum_{ij=1}^{N}\sum_{t=1}^{T} x^{BC}_{k,ij,t} \qquad \bar{x}_k^{RC} = \frac{1}{NT}\sum_{ij=1}^{N}\sum_{t=1}^{T} x^{RC}_{k,ij,t}$$

Note that N and T are the number of zone pairs and time points. The SUR method can be directly applied to Eqs. (35) and (36), and heterogeneity parameters and constants can be estimated as done in Eq. (18).

Using data for 1967 and 1978 we estimated DSFIX parameters and subsequently predicted the trips according to travel mode in 1987. Only final (1987) prediction accuracy is shown in Table 11. The traditional model listed in the Table is without heterogeneity and first-order serial correlation. Accordingly, DSFIX is relatively superior in terms of prediction accuracy, even though their goodness-of-fit indices are not satisfactory.

Table 11 Prediction Accuracy of Modal Split Models for 1987

Model	R	Theil's
Traditional model	0.534	0.281
DSFIX	0.596	0.268

5. CONCLUSIONS

Environments surrounding transportation experience more acute and more frequent changes now than ever before. For this reason, traditional travel demand models that extrapolate longitudinally from cross-sectional relationships have become impractical.

This paper have developed a new model system that incorporates unobserved heterogeneity and first-order serial correlation based on repeated cross-sectional data gathered at multi-year intervals. Notable results have been obtained.

The conventional cross-sectional assumptions of temporal stability, homogeneity and serial independence, accepted in traditional travel demand models are all statistically rejected. Further, the issue of temporal variation becomes difficult with only a minimal number of survey time points. Hence, in this paper we have proposed incorporating unobserved heterogeneity and first-order serial correlation of error terms into the model.

With respect to trip generation, attraction and distribution models, dynamic models with fixed-effects and random-effects are developed based on the above statistical results. Through empirical analysis, newly developed dynamic models have proven to be superior to traditional ones in terms of prediction accuracy. As heterogeneity parameters with fixed-effects can reflect different zonal characteristics directly, we have concluded that dynamic models with fixed-effects could be used for long-term predictions.

The log-linear form of an aggregate logit model is used for the modal split phase. Because modes chosen are not independent of one another, the correlation among error terms of different modes should be considered. However, since it was difficult to extend the single-equation model with random-effects to modal split, a dynamic simultaneous-equation model with fixed-effects based on SUR was developed and its effectiveness confirmed by empirical analysis.

These dynamic models are expected to improve prediction accuracy, but there still remain some problems. One is that we have used data only from the initial 1967 survey area. This area is common to all survey years, but since survey areas enlarge with the passage of time, it is necessary to incorporate this phenomenon into our models.

Gravity and logit models are fundamentally non-linear, so they must be transformed to log-linear form in order to apply the ideas introduced for trip generation/attraction models to them. Consequently, non-linear models need to be dealt with directly to further develop dynamic travel demand models.

Finally, we can say that the dynamic models proposed here would be also a useful tool for travel demand analysis and forecasting in developing countries. Because the longitudinal travel data will soon be available in these countries since the Person Trip Survey has been already done to make transportation plans in many Asian Metropolitan Areas and the second and third surveys are successively planning to be carried out to review them.

REFERENCES

Ameniya, T. (1985) **Advance Econometrics**, Cambridge, Harvard University Press.

Bhargava, A. et al. (1982) Serial correlation and the fixed effects model, **Review of Economic Studies, 49**, 533-549.

Hsiao, C. (1986), **The Analysis of Panel Data**, Cambridge University Press.

Ito, T. and Ishida, H. (1997) Regional and temporal analysis for car ownership in Japan, **Journal of the Eastern Asia Society for Transportation Studies, 2(3)**, 971-992.

Kitamura, R. (1990) Panel analysis in transportation planning: An overview, **Transportation Research, 24A(6)**, 401-415.

Lillard, l. A. and Willis, R. J. (1978) Dynamic aspects of earning mobility, **Econometrica, 46(5)**, 985-1012.

Maddala, G. S. (1987) Recent developments in the econometrics of panel data analysis, **Transportation Research, 21A(4/5)**, 303-326.

Meurs, H. J. (1990) Trip generation models with permanent unobserved effects, **Transportation Research, 24B(2)**, 145-158.

Special Issue (1987) Longitudinal data methods, **Transportation Research, 21A(4/5)**, 247-376.

Special Issue (1989) Panel analysis of travel demand, **Transportation, 16**, 97-194.

Special Issue (1990) Dynamic travel behavior analysis, **Transportation Research, 24A(6)**, 399-495.

Sugie, Y. and Kato, F. (1982) Transferability in time of conventional four step travel demand models for Hiroshima, **Memoirs of the Faculty of Eng., Hiroshima Univ., 8(1)**, 77-85.

Sugie, Y., Zhang, J. and Akimasa, F. (1999) Dynamic discrete choice models considering unobserved heterogeneity with mass point approach, **Journal of the Eastern Asia Society for Transportation Studies, 3(5)**, 245-260.

Theil, H. (1961) **Economic Forecast and Policy**, Amsterdam, North-Holland Publishing.

Theil, H. (1969) A multinomial extension of the linear logit model, **International Economic Review, 10**, 251-259.

Zellner, A. (1962) An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias, **Journal of the American Statistical Association, 57**, 348-368.