

Dynamic Travel Demand Models Incorporating Unobserved Heterogeneity and First-order Serial Correlation

Yoriyasu SUGIE

Professor

Graduate School for International Development and Cooperation,
Hiroshima University, Higashi-Hiroshima 739-8529, Japan

Junyi ZHANG

Pacific Consultants Co., Ltd.

7-1, Nishi-Shinjuku 2-Chome, Shinjuku-ku,
Tokyo 163-0730, Japan.

and

Akimasa FUJIWARA

Associate Professor

Graduate School for International Development and Cooperation,
Hiroshima University, Higashi-Hiroshima 739-8529, Japan

Abstract

Very little work using repeated cross-sectional data has been undertaken in transport research. This is especially true for travel data gathered at multiple points in time, especially data that is gathered every 5-10 years such as Urban Area Travel Survey Data and Road Traffic Census Data in Japan. Accordingly, travel demand modeling based on these types of data is not yet fully developed. This paper deals with methods for developing models which include time series factors for predicting travel demand using three time-points travel data gathered in Hiroshima. As a result, it was shown that model parameters based on cross-sectional data were not stable over time by using Covariance Analysis or T-Statistic. The existence of first-order serial correlation in residuals was confirmed by using Generalized Durbin-Watson Statistics, while unobserved heterogeneity was checked by using Breusch-Pagan Statistics. Fixed-effects models using these two factors were developed and it was shown that their predicting accuracy was improved in comparison to traditional cross-sectional models.

1 INTRODUCTION

Cross-sectional data has been broadly used in travel demand modeling, especially for 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, or longer-term travel demand prediction makes the accuracy at the target year worse.

To clear away these problems, longitudinal data which is collected at multiple points in time, becomes being utilized in the field of recent travel behavior research. According to whether samples surveyed are identical or not over time, longitudinal data can be generally classified into panel data and repeated cross-sectional data, respectively.

Although longitudinal data has also its own specific problems including expensive survey cost and attrition bias caused by repletion of surveys, it can give us more plentiful information, especially on temporal change of travel behavior than cross-sectional data. However, up to now, many researches have been done mainly within the context of individual behavioral analysis 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 the interval of long years have not been satisfactorily developed, probably because the enough number of longitudinal data sets cannot be easily obtained for the same area.

This paper aims at improving traditional four-step models, except assignment model, using longitudinal travel data obtained at three different points in time, i.e. 1967, 1978 and 1987 in Hiroshima Metropolitan area. Since the study area is different for the three surveys because of the development of urbanized area, the common area consisting of 32 zones which corresponds to the survey area in 1967 (Sugie *et al.*, 1982), is used for this study (see Figure 1). This data belongs to repeated cross-sectional data gathered at long years intervals (i.e. 10 years). Though individuals sampled in the survey are different at each point in time, the analysis unit, i.e. zone, is fixed over the twenty years from 1967 to 1987. Therefore, statistical methods developed for the analysis of individual panel data can be applied to the zonal aggregate level (Ito *et al.*, 1997).

In the field of travel behavior research, 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).

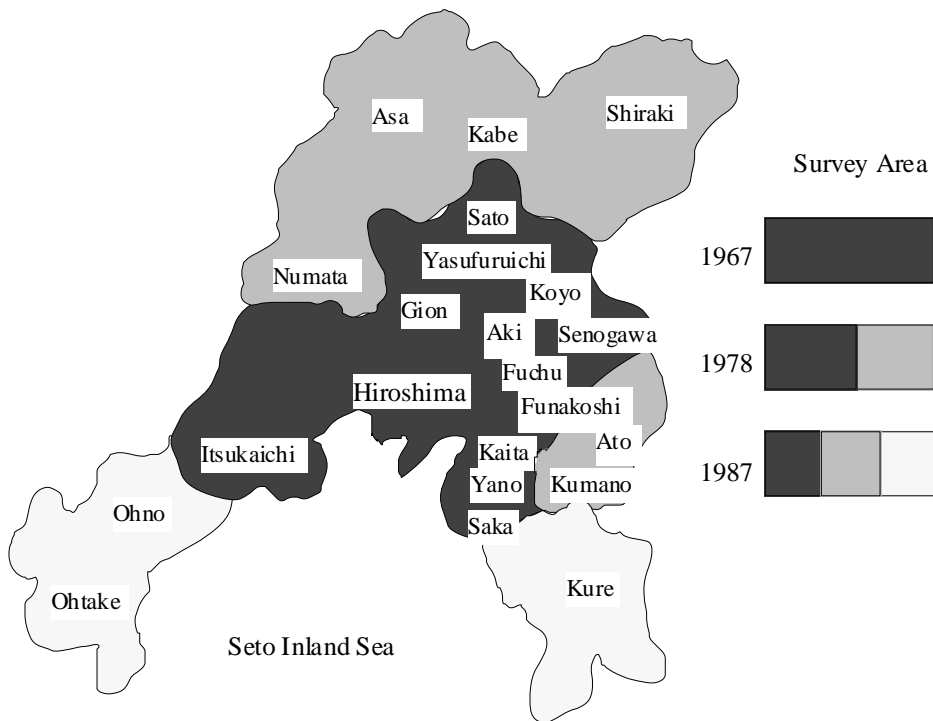


Figure 1 Development of survey area in Hiroshima and its surroundings

Accordingly, numerous researches on disaggregate travel behavior using panel data have been done (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.

Most dynamic models have been developed using short-term panel data in the field of individual travel behavioral analysis. However, when we consider transportation planning 10 to 20 years hence, the assumption of dynamic model at time t , which is a function of dependent variable at time $t-1$, is doubtful. Therefore, it seems essential to consider time series factors when the time intervals for such a survey are longer and the number of time points is small. 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.

Concerning the main structure of this paper, cross-sectional assumptions are statistically tested for trip generation, attraction and distribution models in section 2. Based on the test results, dynamic single-equation models considering unobserved heterogeneity and first-order serial correlation are developed in section 3. With respect to modal split model, because it is not realistic to treat the error terms of different modes independently, a new dynamic modal split model with simultaneous-equations is developed in section 4.

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:

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

where,

i, t : indicating zone (or zone pair) and time,

y_{it} : dependent variable, e.g. generated trips for trip generation,

$x_{k,it}$: k 'th explanatory variable of y_{it} ,

β_k : parameter of $x_{k,it}$,

μ : constant term,

v_{it} : error term following an identical and independent distribution (i.i.d.) for i and t ,

K : total number of explanatory variables.

The following assumptions are supposed in eqn (1).

Assumption 1: temporal stability, i.e. β_k is temporally invariant.

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

Assumption 3: serial independence of v_{it} .

Based on the above assumptions, eqn (1) can be estimated using 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 employees in industry, business and commerce are used as explanatory variables for trip generation and attraction models which can be expressed as eqn (1).

With respect to trip distribution model, a specific gravity model shown in eqn (2) is adopted in order to check the temporal stability of model parameters. Eqn (3) is its doubly constraint functions. Parameters β_G and β_A are often set to one for conventional gravity models. Constant parameter α is replaced by balancing factors in the calibration.

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

$$G_{it} = \sum_j y_{ijt} \quad \text{and} \quad A_{jt} = \sum_i y_{ijt} \quad (3)$$

where,

- y_{ijt} : interzonal trips between zone i and j at time t,
- G_{it} : generated trips at zone i,
- A_{jt} : attracted trips at zone j,
- T_{ijt} : average travel time between zone i and j,
- $\alpha, \beta_G, \beta_A, \beta_T$: parameters.

For the sake of practical use, transform eqn (2) as eqn (4).

$$\ln(y_{ijt}) = \ln(\alpha) + \beta_G \ln(G_{it}) + \beta_A \ln(A_{jt}) - \beta_T \ln(T_{ijt}) + v_{ijt} \quad (4)$$

This indicates that trip distribution model can be also treated as one of eqn (1). Accordingly, generation/attraction models and trip distribution model are estimated using OLS and only the results with respect to total trip purpose are shown in Table 1 because of limited space. The sample size for trip distribution is smaller than the expected one (i.e. $32 \times 32 = 1,024$), because intrazonal samples and some samples with zero trip which are caused by low sampling rate (i.e. 1.5%) in 1978, are excluded in the analysis. It is shown that each model has a high level of goodness-of-fit (i.e. multiple correlation coefficient) and that population and employees in business and commerce are significant in the trip generation/attraction models.

2.3 Test of temporal stability

To test whether the estimated parameters based on OLS are temporally stable or not, we use Covariance Analysis method (Hsiao, 1986). Firstly, estimate eqn (5) using OLS for each year.

$$y_{it} = \mu_t + \sum_{k=1}^K \beta_{k,t} x_{k,it} + v_{it} \quad (5)$$

In eqn (5), constant term μ_t and parameters $\beta_{k,t}$ vary over time and the residual sum of squares can be calculated as S_1 . Secondly, estimate eqn (1) based on OLS using the pooled data for 1967 and 1978,

then the residual sum of squares can be calculated as S_2 . The hypothesis of temporal stability for constant term and parameters can be viewed as eqn (5) subject to $(k+1)(T-1)$ linear restrictions:

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

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

$$F = \frac{(S_2 - S_1) / [(T - 1)(K + 1)]}{S_1 / [NT - T(K + 1)]} \tag{6}$$

The test results using eqn (6) are shown as Table 2. The reason why the test result for school attraction model is not indicated, is because the number of students in 1967, which is 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 significantly rejected.

Table 1 Estimation results of trip generation, attraction and distribution models with respect to total trip purpose

Explanatory variable	Trip generation			Trip attraction		
	1967	1978	1987	1967	1978	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	1978	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 Test results of temporal stability

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 Test of homogeneity

Consider the following eqn (7) with fixed-effects parameter δ_i .

$$y_{it} = \delta_i + \mu + \sum_{k=1}^K \beta_k x_{k,it} + u_{it} \tag{7}$$

The test of homogeneity means whether null hypothesis $H_0: \delta_i = 0$ holds or not. We estimate, first of all, the pooled model, i.e. eqn (7) 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 the homogeneity (Maddala, 1987; Meurs, 1990).

$$\lambda = \frac{NT}{2(T-1)} \left[\frac{\sum_{i=1}^N \left[\sum_{t=1}^T \hat{u}_{it} \right]^2}{\sum_{i=1}^N \sum_{t=1}^T \hat{u}_{it}^2} - 1 \right]^2 \tag{8}$$

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

Table 3 Test results of homogeneity

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 Test of serial independence

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

$$u_{it} = \rho u_{it-1} + e_{it} \tag{9}$$

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

By estimating eqns (7) and (9) based on OLS when null hypothesis $H_0: \rho_i = 0$ holds, we can obtain the estimated residual \hat{u}_{it} and establish the following generalized Durbin-Watson statistic (Bhargava *et al.*, 1982; Maddala, 1987).

$$DW = \sum_{i=1}^N \sum_{t=2}^T (\hat{u}_{it} - \hat{u}_{it-1})^2 / \sum_{i=1}^N \sum_{t=1}^T \hat{u}_{it}^2 \tag{10}$$

The test results using eqn (10) are shown as Table 4. This indicates that there exist first-order serial correlations in all of the models at the significant level 5% (i.e. the critical value is approximately 2.00).
 Table 4 Test results of serial independence

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 CONSIDERING UNOBSERVED HETEROGENEITY AND FIRST-ORDER SERIAL CORRELATION

From the above test results, it seems desirable to manage to relax all of these three cross-sectional assumptions. However, because the longitudinal data used here have only three time points, it is not possible to incorporate time-varying parameters into the models. For this reason, we develop dynamic models considering 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):

$$y_{it} = \mu + \sum_{k=1}^K \beta_k x_{k,it} + v_{it} \tag{11}$$

$$v_{it} = \delta_i + u_{it} \tag{12}$$

$$u_{it} = \rho u_{it-1} + e_{it} \tag{9'}$$

where, v_{it} , u_{it} , e_{it} are error terms and e_{it} is the one following an i.i.d.

The initial condition for eqns (11) ~ (9') is given as eqn (13) (Lillard *et al.*, 1978).

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

According to the assumptions on 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, generalized least squares (GLS) method could be applied. The GLS estimator can be defined as (Amemiya, 1985):

$$\beta = \left[\mathbf{X}^* \mathbf{\Omega}^{*-1} \mathbf{X}^* \right]^{-1} \mathbf{X}^* \mathbf{\Omega}^{*-1} \mathbf{y}$$

(14)

where, $\Omega^* = (I_N \otimes \Omega)$ is a $NT \times NT$ matrix, Ω being a $T \times T$ variance-covariance matrix of the stationary first-order auto regression, i.e. Ω has elements of the form (Bhargava *et al.*, 1982).

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

From a standpoint of practical estimation, eqn (14) is a form of very complicated expression. Here, we transform eqns (10) ~ (12) using a simple way.

3.1 Specification of dynamic model with fixed-effects (DFIX)

Based on the above theoretical background, eqns (11) ~ (12) can be transformed as follows:

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

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

where,

$$\begin{aligned} \varepsilon_{i1} &= e_{i1} - \sqrt{1 - \rho^2} \bar{u}_i, \quad \varepsilon_{it} = e_{it} - (1 - \rho) \bar{u}_i \\ \bar{y}_i &= \frac{1}{T} \sum_{t=1}^T y_{it}, \quad \bar{x}_{k,i} = \frac{1}{T} \sum_{t=1}^T x_{k,it} \end{aligned}$$

Because the error term ε_{it} ($t = 1, 2, \dots, T$) is serially independent, the OLS can be applied to eqns (16) and (17). However, when the number of time points for the survey is small, we propose to use which can increase degree of freedom for the estimation, instead of $\bar{y}_i, \bar{x}_{k,i}$.

$$\sum_{i=1}^N \delta_i = 0.$$

In order to estimate μ and δ_i separately, Hsiao (1986) assumes Using the estimated value $\hat{\beta}_k$ of β_k from eqns (16) and (17) with \bar{y}_i, \bar{x}_k by use of OLS, we can calculate the estimated values $\hat{\mu}, \hat{\delta}_i$ of μ, δ_i as follows:

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

In fact, the consistent estimator of ρ must be pre-determined by the estimated parameter of y_{it-1} in eqn (19), because it cannot be obtained directly from eqns (16) and (17) by using OLS when the enough number of time points for the survey cannot be obtained.

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

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 $x_{k,it}$ (eqn (20)).

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

3.2 Specification of dynamic model with random-effects (DRAN)

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

$$\Omega = \sigma_u^2 \begin{bmatrix} 1 & \rho & \rho^2 & \dots & \rho^{T-1} \\ \rho & 1 & \rho & \dots & \cdot \\ \rho^2 & \rho & 1 & \dots & \cdot \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{T-1} & \cdot & \cdot & \dots & 1 \end{bmatrix} + \sigma_\delta^2 \mathbf{ii}' \quad (21)$$

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

Since substituting Ω into eqn (14) will cause the same problem as in DFIX, we propose another transformation method to specify DRAN.

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

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

where,

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

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

$$\Psi = \begin{bmatrix} \sigma_1^2 & \sigma_{cov} & \sigma_{cov} & \dots & \sigma_{cov} \\ \sigma_{cov} & \sigma_2^2 & \sigma_{cov} & \dots & \sigma_{cov} \\ \sigma_{cov} & \sigma_{cov} & \sigma_3^2 & \dots & \sigma_{cov} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma_{cov} & \sigma_{cov} & \sigma_{cov} & \dots & \sigma_T^2 \end{bmatrix} \quad (24)$$

where,

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

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

Eqn (24) is a special case of the error structure of GLS. When $T = 2$, it turns out to be that of

Seemingly Unrelated Regressions (SUR) method (Zellner, 1962).

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

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

3.3 Estimation of DFIX and DRAN

In this section, we estimate DFIX and DRAN using the data in 1967 and 1978 and show only the estimation results with respect to total trip purpose in Table 5. 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 in order to predict the travel demand in 1987 and then to compare them with the predicting results by other models: OLS-78, SUR-78, FSUR-78, defined in Table 6.

OLS-78 is a traditional prediction model, which assumes that the present cross-sectional relationship will be extrapolated to the future, so parameters of the base year (here, i.e. 1978) are adopted for prediction. SUR-78 considers temporal variation of parameters, zonal variation of constant term and arbitrary serial correlation. The difference between SUR-78 and FSUR-78 is that the latter does not assume the parameters variable over time. They use the data in 1967 and 1978 for model estimation. However, since it is not clear how the correlation between error terms of the present and the future is considered, it is not included here for prediction. Comparing these two models can make it clear whether the time-

Table 5 Estimation results for generation, attraction and distribution models with respect to total trip purpose

Explanatory variable	Generation		Attraction		Distribution	
	DFIX	DRAN	DFIX	DRAN	DFIX	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)**		
Generated trips					0.878 (17.3)**	1.057 (19.8)**
Attracted trips					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 List of prediction models for 1987

Prediction model	Variation		Serial correlation	Estimation method
	$\mu + \delta_i$	β_t		
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) common parameters for 1967 and 78

varying parameters influence the prediction accuracy greatly or not.

The goodness-of-fit indices evaluating the prediction accuracy used are correlation coefficient (R) and Theil's inequality coefficient (U_t ; $0 \leq \text{its value} \leq 1$) between actual trips Y_i and estimated \hat{Y}_i in 1987 (Theil, 1961). U_t can be expressed as eqn (26).

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

Larger value of R and smaller value of U_t mean higher prediction accuracy. The prediction accuracy of each model defined in Table 6 is shown in Table 7.

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

Because heterogeneity parameter represents travel change due to the unmeasurable zonal (or spatial) characteristics, it must be more effective than time-varying parameters. Besides that, incorporating first-order serial correlation into DFIX and DRAN makes it possible to consider the travel information at previous time point explicitly. As a result, DFIX and DRAN have the best accuracy of all the models

Table 7 Prediction accuracy of the models defined in Table 6 with respect to 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

defined in Table 6. Besides, since the heterogeneity parameter can be explicitly incorporated in DFIX, it is more desirable to use DFIX to predict the travel demand rather than DRAN.

4 DYNAMIC MODELS WITH SIMULTANEOUS-EQUATIONS

In this section, we extend dynamic single-equation models in section 3 to modal split model. Even though a number of modal split models have been used in travel demand analysis, a logit-type model shown as eqn (27) is adopted because it has a more theoretical foundation than others.

$$P_{ij,t}^m = \exp(V_{ij,t}^m) / \sum_{m'=1}^M \exp(V_{ij,t}^{m'}) \quad (27)$$

$$V_{ij,t}^m = \alpha^m + \sum_{k=1}^{K_1} \beta_k x_{k,ij,t}^m + \sum_{k=K_1+1}^K \beta_k^m x_{k,ij,t}^m \quad (28)$$

where,

$P_{ij,t}^m$: share of mode m between zone i and j at time t ,

$V_{ij,t}^m$: linear utility function of mode m ,

$x_{k,ij,t}^m$: k 'th explanatory variable of mode m (e.g. average travel time),

β_k : common parameter of $x_{k,ij,t}^m$ across modes,

β_k^m : common explanatory variable across modes,

α^m : parameter of $x_{k,ij,t}^m$ for mode m ,

α^m : constant term of mode m .

There exist two methods to estimate eqn (27): one is Maximum Likelihood (ML) method, another is GLS (or SUR). We adopt the latter SUR here because it is easier to incorporate time series information into the model. In the case of three travel modes; CAR, BUS and RAIL, eqn (27) is transformed as follows (Theil, 1969) :

$$\ln \left(P_{ij,t}^{BUS} / P_{ij,t}^{CAR} \right) = V_{ij,t}^{BUS} - V_{ij,t}^{CAR} + \omega_{ij,t} \quad (29)$$

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

The common explanatory variable is average travel time for each mode 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 of origin zone i ;
- 4) percentage of employees in business and commerce at destination zone j which is an indicator to express parking difficulty.

Where, G_{it} , A_{jt} , T_{ijt} are defined as eqn (2).

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

4.1 Test of cross-sectional assumptions in modal split model

To carry out these tests, we use the data of total trip purpose in 1967 and 1978. It is easier to use Covariance Analysis method to test temporal stability like in section 3. Applying the same method to eqns (29) and (30) becomes so complicated that we estimate eqns(29) and (30) firstly by using SUR for each year as shown in Table 8. The sample size becomes smaller because of the same reason shown in Table 1. The models obtained have relatively high Multiple correlation coefficients, but parameters seem to be variable over time.

It is then tested whether the parameters in each year are equal or not by T-statistic (see Table 9). It is clear that most of the parameters are significantly different between 1967 and 1978. We use successively the same statistics employed in the previous section to test the assumptions of homogeneity and serial independence, but the estimated residuals used here are from simultaneous estimation of eqns (29) and (30) using SUR, not from the separate estimation using OLS. The test results shown in Table 10 indicate that all of them are statistically rejected at the significant level 5% or 1%. This suggests the existence of heterogeneity and first-order serial correlation.

Based on the above test results, similar to section 3, rewrite eqns (29) and (30) as follows:

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

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

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

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

where,

δ_{ij}^{BC} , δ_{ij}^{RC} : heterogeneity parameters,

ρ^{BC} , ρ^{RC} : first-order serial correlation coefficients.

Table 8 Estimation results of eqns (29) and (30) for each year using SUR method

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 at origin zone	-2.35 (1.54)	-4.62 (2.73)**	-4.65 (2.81)**
Rate of employment in business and commerce at 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 at origin zone	1.07 (0.44)	-2.69 (1.15)	-0.452 (0.21)
Rate of employment in business and commerce at 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 of temporal stability for modal split model

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**
Rate of employment in business and commerce at destination zone	2.14*	20.6**

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

Table 10 Test results of heterogeneity and first-order serial correlation for modal split model

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

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

4.2 Specification and estimation of dynamic simultaneous-equations modal split model with fixed-effects (DSEFIX)

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

$$y'_{ij,t}{}^{BC} = \sum_{k=1}^{K^{BC}} \beta_k^{BC} x'_{k,ij,t}{}^{BC} + \varepsilon_{ij,t}{}^{BC} \quad (35)$$

$$y'_{ij,t}{}^{RC} = \sum_{k=1}^{K^{RC}} \beta_k^{RC} x'_{k,ij,t}{}^{RC} + \varepsilon_{ij,t}{}^{RC} \quad (36)$$

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

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

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

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

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

Where,

$$y_{ij,t}^{BC} = \ln(P_{ij,t}^{BUS} / P_{ij,t}^{CAR})$$

$$y_{ij,t}^{RC} = \ln(P_{ij,t}^{RAIL} / P_{ij,t}^{CAR})$$

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

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

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

N and T are the number of zone pairs and time points. The SUR can be directly applied to eqns (35) and (36). Heterogeneity parameter and constant term are estimated in the same way as eqn (18).

Furthermore, we estimate DSFIX using the data in 1967 and 1978 and use the estimated parameters to predict the trips by travel mode in 1987. Only the final prediction accuracy for 1987 is shown in Table 11. Traditional one in the Table is the model without heterogeneity and first-order serial correlation. As a result, DSFIX is relatively superior to the traditional one in terms of prediction accuracy

Table 11 Prediction accuracy of modal split models

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

5 CONCLUSIONS

The environment surrounding transportation changes now largely more often than before. For this reason, traditional cross-sectional travel demand models assuming longitudinal extrapolation of cross-sectional relationships, becomes unrealistic for actual use.

This paper develops a new model system considering unobserved heterogeneity and first-order serial correlation based on repeated cross-sectional data gathered at long years intervals. As a result, some important conclusions can be obtained.

Cross-sectional assumptions: temporal stability, homogeneity and serial independence, supposed in traditional travel demand models are all statistically rejected. However, when available data has only a small number of time points, considering temporal variation becomes difficult. Therefore, it is proposed to incorporate 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 the empirical analysis, new developed dynamic models are proved to be superior to the traditional ones in terms of prediction accuracy. Considering heterogeneity parameters with fixed-effects can reflect different zonal characteristics directly, so it is concluded that the dynamic model with fixed-effects could be used for long-term prediction.

Associate with modal split model, an aggregate logit model transformed as a linear form is adopted for the study. Because choice of travel modes is not done independently each other, the correlation among error terms of different modes should be considered. However, it is difficult to extend the single-equation model with random-effects to modal split, so a dynamic simultaneous-equation model with fixed-effects based on SUR is developed and its effectiveness is confirmed by empirical analysis.

It is expected to improve the prediction accuracy by use of dynamic models which are presented in this study, but there still remain some problems.

Here, we use only the data from the common area for the three time-point surveys which corresponds to the 1967 survey area. It is often seen that the survey area is enlarged with the passage of time, so it is necessary to study more how to apply the models to the area newly included.

Gravity model and logit model are basically of non-linear type, so they must be log-transformed to linear ones in order to apply the ideas introduced for trip generation/attraction models to them. It is therefore necessary to treat with non-linear models directly for the further development of 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

- [1] Ameniya, T.(1985), *Advance Econometrics*, Cambridge, Harvard University Press.
- [2] Bhargava, A. et al. (1982), Serial correlation and the fixed effects model, *Review of Economic Studies*, 49, 533–549.
- [3] Hsiao, C. (1986), *The Analysis of Panel Data*, Cambridge University Press.
- [4] 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.
- [5] Kitamura, R. (1990), Panel analysis in transportation planning: An overview, *Transportation Research*, 24A(6), 401–415.
- [6] Lillard, I. A. and Willis, R. J. (1978), Dynamic aspects of earning mobility, *Econometrica*, 46(5), 985–1012.
- [7] Maddala, G. S. (1987), Recent developments in the econometrics of panel data analysis, *Transportation Research*, 21A(4/5), 303–326.
- [8] Meurs, H. J. (1990), Trip generation models with permanent unobserved effects, *Transportation Research*, 24B(2), 145–158.
- [9] Special Issue (1987), Longitudinal data methods, *Transportation Research*, 21A(4/5), 247–376.
- [10] Special Issue (1989), Panel analysis of travel demand, *Transportation*, 16, 97–194.
- [11] Special Issue (1990), Dynamic travel behavior analysis, *Transportation Research*, 24A(6), 399–495.
- [13] 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.
- [14] 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.
- [15] Theil, H. (1961), *Economic Forecast and Policy*, Amsterdam, North-Holland Publishing.
- [16] Theil, H. (1969), A multinomial extension of the linear logit model, *International Economic Review*, 10, 251–259.
- [17] 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.