Comparison of Alternative Methods for Updating Disaggregate Logit Mode Choice Models

DANIEL A. BADOE AND ERIC J. MILLER

An empirical assessment of alternative methods of updating disaggregate travel choice models so that their transferability from the estimation context within which they were originally developed to an application context (which differs from the original estimation context geographically or temporally, or both) is presented. The case study for the empirical tests performed is a long-term temporal transfer of work trip logit mode choice models estimated using 1964 data for the greater Toronto area (GTA) to represent 1986 work trip mode choice in the GTA. Three updating procedures that have been previously presented in the literature are examined (Bayesian updating, transfer scaling, and combined transfer estimation), plus a fourth new procedure, joint context estimation. All four procedures assume that a "small" data set of observed travel choices is available for the application context, which can be used in the updating procedure. The case study results indicate that the latter three procedures all possess merit as potential updating methods, with the choice among the three depending on such items as model specification and application context sample size. The results also indicate that if the application context sample size exceeds 400 to 500 observations, then updating may provide little or no improvement over simple estimation of an application context model, especially if "full" model specification is supported by the available data.

The spatial and temporal transferability of random utility models of travel demand is a matter of considerable practical interest. Although the empirical evidence concerning the transferability properties of random utility models is mixed (*1*), consensus exists that the potential for model transfer is greatly enhanced if local area (i.e., application context) data are used to update the model so that it better reflects application context conditions (*1–6*). At least two major reasons underlie this need to update transferred models (*4*):

1. Limitations in model specification, perhaps most notably as a result of omission of relevant variables; and
2. Differences in unmodeled "contextual factors" (geographical, historical, etc.) between the estimation and application context that affect the evolution over time of trip-makers' travel tastes and preferences.

Three major updating procedures have been presented in the literature to date:

1. Bayesian updating, in which parameter estimates from a small application context sample are combined with the estimation context parameter values using a classical Bayesian analysis to yield an updated set of parameters (*2*);
2. Transfer scaling, in which the application context utility function scales and alternative-specific constants are estimated from a small application context sample, assuming that the remainder of the utility function parameters are transferable from the estimation context (*3,4*); and
3. Combined transfer estimation, which can be viewed as a generalization of the Bayesian updating approach, which accounts for transfer scaling effects (*6*).

These approaches all assume that the estimation context model parameter values are known and that a small sample application context data set is available which permits the estimation of an application context model that is identical to the estimation context model being transferred. If, however, the estimation context data set used to estimate the original model parameters is also available (which in many instances may well be the case), a fourth approach is possible. This fourth approach, labeled joint *context estimation* involves estimating a new joint estimation/application context model, using both the estimation context and application context data sets.

This paper has two purposes. First, it provides a systematic comparison of the four updating techniques within a common empirical application. Second, this empirical application is unique in the literature because it involves assessing the relative effectiveness of the various updating procedures in achieving long-term temporal transferability of a disaggregate choice model within the same geographic area. Specifically the case study consists of updating 1964 morning peak-period work trip mode choice models developed for the greater Toronto area (GTA), Canada, over a 22-year period to reflect 1986 conditions.

The next section of this paper briefly reviews the four updating procedures. The paper's third section briefly describes the data sets used in the study, and the fourth section describes the test procedure employed. The fifth section presents and discusses the results obtained. The final section of the paper then summarizes the findings of the study and their implications for the state of practice in model updating and transfer.

MODEL UPDATING METHODS

It is assumed that a disaggregate multinomial logit choice model is to be transferred from an original (estimation) context to a new (application) context; that the estimation context parameter estimates are known; and that a small sample data set drawn from the application context that is suitable for estimating a model specified identically to the estimation context model is available.

Notation used throughout this discussion of methods includes the following:

D. A. Badoe, University of Toronto, Joint Program in Transportation, 42 St. George Street, Toronto, Ontario M5S 1A4, Canada. E. J. Miller, Department of Civil Engineering, University of Toronto, 35 St. George Street, Toronto, Ontario M5S 1A4, Canada.

$\boldsymbol{\Theta}$ = $[K \times 1]$ vector of utility function parameters, where $K = N + M - 1 = [\begin{smallmatrix}\alpha \\ \beta\end{smallmatrix}]$;

α = $[(N - 1) \times 1]$ vector of alternative-specific constants, where N is maximum number of alternatives available in choice set;

β = $[M \times 1]$ vector of parameters consisting of utility function weights for M explanatory variables in model;

$\mathbf{X}_{it}$ = $[M \times 1]$ vector of explanatory variables for alternative i for individual t;

V_{it} = systematic utility for alternative i for individual t $= \beta'\mathbf{X}_{it} + \alpha_i$ (where $\alpha_N = 0$ by definition); (1)

P_{it} = probability that individual t will choose alternative i from choice set C_t $= \exp(V_{it})/\Sigma_{j \in C_t}\exp(V_{jt})$; (2)

$\boldsymbol{\Theta}_1, \boldsymbol{\Theta}_2$ = estimates of $\boldsymbol{\Theta}$ derived from estimation context and application context data sets, respectively;

Σ_i = estimated parameter covariance matrix for context i ($i = 1,2$); and

$\boldsymbol{\Theta}_{\text{update}}$ = final estimates of $\boldsymbol{\Theta}$ to be used in application context, as generated by updating procedure.

Bayesian Updating

On the basis of the seminal work of Atherton and Ben-Akiva (*2*), it is well known in the literature that the Bayes theorem can be applied to the updating problem to yield asymptotically normal updated parameters with the following mean:

$$\boldsymbol{\Theta}_{\text{update}} = \left(\sum_1^{-1} + \sum_2^{-1}\right)^{-1}\left(\sum_1^{-1} \boldsymbol{\Theta}_1 + \sum_2^{-1} \boldsymbol{\Theta}_2\right) \tag{3}$$

and covariance matrix:

$$\sum\nolimits_{\text{update}} = \left(\sum_1^{-1} + \sum_2^{-1}\right)^{-1} \tag{4}$$

Thus, to use this updating procedure, one must estimate an application context model using the available application context small sample using standard maximum likelihood methods to compute $\boldsymbol{\Theta}_2$ and Σ_2. These can then be combined with the known values of $\boldsymbol{\Theta}_1$ and Σ_1 from the estimation context using Equation 3 to yield the updated model parameters. Atherton and Ben-Akiva (*2*) used this procedure with considerable success to update a 1968 Washington, D.C., work trip mode choice model to reflect 1963 New Bedford, Massachusetts, and 1967 Los Angeles applications.

Transfer Scaling

It is well recognized that alternative-specific constants are likely not to be transferable between applications, given the extent to which systematic but unmodeled "contextual factors" are captured within these terms. It is equally true that the overall scale of the model's utility functions (which are not statistically identifiable within a standard cross-sectional model) are also likely to vary from one application to another, again because of unmodeled contextual factors.

A not unreasonable hypothesis on which to construct an updating procedure, therefore, is to assume that the utility function parameters computed in the estimation context, excluding the alternative-specific constants, are transferable to the application context, up to scale (note that, among other implications, this results in values of time—as defined by the ratios of time-to-cost parameters within the utility functions—being equal in the two contexts). Indeed, as shown elsewhere (*1,3,4*), much of the transfer bias can, in fact, be eliminated by adjusting model constants and scales. The updating problem then becomes one of determining changes in alternative-specific utility function constants and scales relative to estimation context values. For example, given a set of estimation context parameters (excluding alternative-specific constants), β_1, one can assume that the application context systematic utilities, $V_{it,2}$, take the following form:

$$V_{it,2} = \mu_{i,2}\beta_1'\mathbf{X}_{it,2} + \alpha_{i,2} \tag{5}$$

where $\mu_{i,2}$ is the ratio of the application context utility function scale to the (unidentified) estimation context utility function scale for alternative i, and all other terms are as previously defined, with the addition of the subscript 2 to indicate the application context.

Alternatively, Gunn et al. (*3*) apply scale factors to various groupings of parameters, where these groupings are defined on the basis of variable type rather than alternative. Equation 5 is thus a special case of the Gunn et al. formulation, which also includes as special cases complete reestimation of the model parameters on the basis of the application context data set (i.e., 1 μ for every parameter) and "naive" transfer of the estimation context model ($\mu = 1$).

The updated application context alternative-specific constants (α) and scale adjustments (μ) are readily estimated given an application context small sample using standard maximum likelihood methods, with the constructed variable $W_{it} = \beta_1'\mathbf{X}_{it,2}$ being the single explanatory variable in the utility function for each alternative.

Given that W_{it} is constructed using the estimated values β_1, the standard errors reported by typical logit model estimation packages will be biased downwards. If it is critical to the evaluation of the updating results to eliminate this bias, then appropriate corrections can be computed. More typically, the estimation results obtained will be sufficiently robust to allow the modeler to use the unadjusted standard errors, with the recognition that they somewhat overestimate the precision of the parameter estimates.

Successful applications of transfer scaling techniques include the following:

1. Gunn et al. (*3*), in which alternative transfer scaling schemes were applied to four different models: joint mode and destination choice models for personal business trips and shopping trips and trip frequency choice models for the same two trip purposes; with the transfer occurring between two regions in the Netherlands (Rotterdam/The Hague and Utrecht), and with the data sets for the two urbanized regions being collected 5 years apart and at different times of the year; and
2. Koppleman et al. (*4*), in which both intraregional transferability within the Washington, D.C., area and interregional transferability among the metropolitan areas of Washington, D.C.; Baltimore; and Minneapolis-St. Paul were investigated for the case of work trip mode choice models.

Combined Transfer Estimation

Implicit in the Bayesian updating approach is the assumption that $\boldsymbol{\Theta}_1 = \boldsymbol{\Theta}_2 = \boldsymbol{\Theta}$; that is, that the estimation and application contexts share the same underlying set of parameters. The transfer scaling method, on the other hand, explicitly assumes that a "transfer bias," $\boldsymbol{\Delta}$, exists, where

$$\mathbf{\Delta} = \mathbf{\Theta}_2 - \mathbf{\Theta}_1 \quad (6)$$

Ben-Akiva and Bolduc (*6*) present a generalization of the Bayesian approach, which accounts for a nonzero $\mathbf{\Delta}$, and which yields the minimum mean square error estimate of $\mathbf{\Theta}_{\text{update}}$ achievable from a linear combination of the estimation and application context parameter estimates. As shown in Equation 6, this minimum mean square error estimate is provided by

$$\mathbf{\Theta}_{\text{update}} = [(\Sigma_1 + \mathbf{\Delta\Delta}')^{-1} + \Sigma_2^{-1}]^{-1}[(\Sigma_1 + \mathbf{\Delta\Delta}')^{-1}\mathbf{\Theta}_1 + \Sigma_2^{-1}\mathbf{\Theta}_2] \quad (7)$$

Comparison of Equation 7 with Equation 3 indicates that the combined transfer estimator reduces to the Bayesian estimator in the case of $\mathbf{\Delta} = 0$. In practice, the unknown transfer bias $\mathbf{\Delta}$ is approximated by the estimated bias $d = \mathbf{\Theta}_2 - \mathbf{\Theta}_1$. Ben-Akiva and Bolduc also demonstrate theoretically that the combined transfer estimator is superior to simply using the application context parameter estimates $\mathbf{\Theta}_2$, providing the transfer bias, $\mathbf{\Delta}$, is small. If the transfer bias $\mathbf{\Delta}$ is large, then the term $(\Sigma_1 + \mathbf{\Delta\Delta}')^{-1}$ in Equation 7 becomes negligible and hence $\mathbf{\Theta}_{\text{update}} \approx \mathbf{\Theta}_2$.

Joint Context Estimation

The transfer scaling procedure described above for updating model constants and scales makes the following assumptions concerning the other model parameters (i.e., β):

1. $\beta_1 = \beta_2, = \beta$;
2. $\beta_1 - \beta$ is small (i.e., the sample error in the estimates of β obtained from the estimation context are small); and
3. These parameter estimates are obtained solely from the estimation context data, independent of and before consideration of application context data (which are allowed only to affect the application context constants and scales).

A much more general model that is fully consistent with the behavioral assumptions mentioned earlier is one in which β is jointly estimated using both the estimation and application context data sets, simultaneously with the estimation of the alternative-specific constants for both contexts and the scales of one context relative to the other.

The following notation is used in developing the joint context estimation procedure:

p = 1 for estimation context; = 2 for application context;
$\mathbf{s}_{it}^p$ = vector of explanatory variables for alternative i common to Periods 1 and 2 (i.e., associated with the constant parameter vector γ), but with values given for person t in context p;
α^p = vector of utility function parameters assumed to be specific to context p (at a minimum, this includes alternative-specific constants for context p);
$\mathbf{r}_{it}^p$ = vector of context-specific explanatory variables for alternative i for individual t within context p;
μ_i = utility function scale for alternative i in context 2 (context superscript has been suppressed to simplify notation; context 1 scales are assumed to be "imbedded" within α^p and γ; given this, μ_i is actually the ratio of context 2 scale to constant 1 scale for alternative i, with absolute values of either of these scales not being identifiable);
$\mathbf{\Theta}$ = combined vector of all parameters to be estimated within joint context model, excluding utility function scales

$$= \begin{bmatrix} \alpha^1 \\ \alpha^2 \\ \gamma \end{bmatrix}$$

$\mathbf{x}_{it}^p$ = combined vector of all explanatory variables in joint context model, for alternative i for person t in context p

$$= \begin{bmatrix} \mathbf{r}_{it}^1 \\ \mathbf{0} \\ \mathbf{S}_{it}^1 \end{bmatrix} \quad \text{for } p = 1$$

and

$$\begin{bmatrix} \mathbf{0} \\ \mathbf{r}_{it}^2 \\ \mathbf{S}_{it}^2 \end{bmatrix} \quad \text{for } p = 2$$

Given these definitions, the systematic utility components for the two contexts are

$$V_{it}^1 = \alpha^{1T}\mathbf{r}_{it}^1 + \gamma^T\mathbf{S}_{it}^1 = \mathbf{\Theta}^T\mathbf{X}_{it}^1 \quad (8)$$

$$V_{it}^2 = \mu_i(\alpha^{2T}\mathbf{r}_{it}^2 + \gamma^T\mathbf{S}_{it}^2) = \mu_i\,\mathbf{\Theta}^T\mathbf{x}_{it}^2 \quad (9)$$

Given the explicit accounting for changes in scales and constants between the two contexts, the usual IID Gumbel Type I distribution is assumed for the random utility terms in each context, leading to conventional multinomial logit choice models:

$$P^p{}_{it} = \frac{\exp(V_{it}^p)}{\sum_{j \in C_t} \exp(V_{jt}^p)} \qquad p = 1,2 \quad (10)$$

If n_p is the number of observations in the context p data set and y_{it}^p is the observed choice indicator for person t in context p (equals 1 if alternative i is chosen; equals 0 otherwise), then the joint log-likelihood function for the joint context model is simply

$$L = \ln L^* = \sum_{p=1}^{2} \sum_{t=1}^{np} \sum_{i \in C_t} y_{it}^p \ln P_{it}^p \quad (11)$$

Substituting Equations 8 through 10 into Equation 11 yields, on rearrangement;

$$L = \ln L^* = \sum_{t=1}^{n1} \sum_{i \in C_t} y_{it}^1 \left[\mathbf{\theta}^T\mathbf{x}_{it}^1 - \ln\left(\sum_{j \in C_t} \exp(\mathbf{\theta}^T\,\mathbf{x}_{jt}^1)\right)\right] + \sum_{t=1}^{n2} \sum_{i \in C_t} y_{it}^2 \left[\mu_i\mathbf{\theta}^T\mathbf{x}_{it}^2 - \ln\left(\sum_{j \in C_t} \exp(\mu_j\mathbf{\theta}_T\mathbf{x}_{jt}^2)\right)\right] \quad (12)$$

With straightforward changes in notation, this model is identical to that developed by Ben-Akiva and Morikawa for combining revealed and stated preference data sets within the same choice model (*7,8*). As noted by Morikawa et al. (*8*), "nested logit" full information likelihood estimation procedures can be applied to this model. Such a procedure was programmed in Fortran by the authors and used in computing the joint context model parameter estimates presented in this paper.

Joint context estimation can clearly be used as a model updating technique, providing that the estimation context data are available for combination with the small sample application context data set.

Although this will not always be the case, access to estimation context data is surely sufficiently feasible in many instances to warrant the testing of this approach compared with the previous three approaches discussed. In particular, potential advantages of joint context estimation relative to conventional transfer scaling techniques include the following:

1. It eliminates biases within the updated application context parameters caused by estimation context sampling errors [a problem discussed in detail elsewhere (6)]; and
2. It provides an operational full-information maximum likelihood procedure for parameter estimation when multiple cross-sectional data bases are available, as opposed to current methods, which are all limited information estimation procedures and hence inefficient in their use of data.

DATA

The 1964 estimation context data set is obtained from the 1964 Metropolitan Toronto and Region Transport Study (MTARTS) survey, which was a home interview survey of 3.3 percent of the households in Metropolitan Toronto and the surrounding regions, consisting of 24,000 households in total. This survey is documented elsewhere (*9*). The 1986 application context data set is obtained from the 1986 Transportation Tomorrow Survey (TTS), a telephone interview survey of 4 percent of the households in the GTA, or 67,000 households in total. This survey is documented elsewhere (*10,11*).

Both surveys were one-day travel surveys that collected generally comparable information, with the single biggest difference being that the 1986 TTS did not collect information on worker occupations and household income. Although coded to different zone systems, these zone systems are roughly similar in definition. Similarly, the study areas for the two surveys vary slightly but not significantly.

All level-of-service data required, with the exception of parking costs and transit fares (which were assembled from other sources), were generated using EMME/2 network assignment procedures applied to 1964 and 1986 road and transit networks. All costs were scaled to 1986 Canadian dollars on the basis of consumer price indexes for transportation.

RESEARCH METHOD

Test Procedure

In this study, the morning peak-period work trips contained in the 1964 MTARTS data base define the estimation context data set. Two multinomial logit work trip mode choice models are estimated using the 1964 data set: one that contains only level-of-service variables (i.e., modal travel times and costs), and one that in addition to these level-of-service variables includes as full a set of socioeconomic variables as is supported by the available data (referred to as the "fully specified" model).

The morning peak-period work trips contained in the 1986 TTS data base then define the application context data set. All four of the updating procedures discussed assume the existence of a "small" sample of trips drawn from the application context to be used in the updating calculations.

To simulate this small sample, random subsets of trip records are drawn from the full TTS data base (which consists in total of 32,328 usable records for this application). To explore the impact of sample size on updating performance, samples of 400, 800, 1,600, 3,200, and 6,400 are used (with each larger sample containing all the records included in the smaller samples). Both the level-of-service and fully specified models are then updated using each of the four updating procedures for each sample size.

In addition, level-of-service and fully specified models are estimated using each of the 1986 small samples. These small sample models are then used to compare the impact that the information contained in the transferred models contributes to predictive performance in the application context with respect to simply using the available application context data.

The performances of the four updated models and the 1986 small model are evaluated for each sample size–model specification combination in terms of how well they replicate the full 32,328 record 1986 TTS set of observed trips. The primary test statistic used is the log-likelihood value generated by the given model when it is applied to the entire 1986 TTS data set.

In addition, however, various aggregate prediction test statistics were constructed, all of which compare in various ways the aggregate number of predicted trips by mode m for a given aggregate group $g, \hat{N}_{mg}$, with the observed number of trips by this mode for this group, N_{mg}. In this paper, only one of these test statistics is discussed, the mean absolute error (MAE) defined as

$$\mathrm{MAE} = \left\{\sum_{m} \sum_{g} |\hat{N}_{mg} - N_{mg}|\right\} \Big/ \left\{\sum_{m} \sum_{g} N_{mg}\right\} \quad (14)$$

Two aggregations are examined: seven major destination groups and worker gender.

Model Specification and Estimation Context Parameters

Three modes are potentially included in the choice set in this study: automobile drive allway, transit allway, and walk. Although automobile passenger, automobile access to transit, and (in 1986) commuter rail modes in principle were also available, these modes were excluded from this analysis to reduce the modeling complexity with respect to specification, decision structure, and introduction of new modes (the commuter rail service did not exist in 1964). Table 1 defines the explanatory variables used in the two models, whereas Table 2 presents the estimation results obtained through standard maximum likelihood estimation of the models using the 1964 MTARTS data set.

RESULTS

Table 3 contains the 1986 full-sample log likelihood values computed for each model specification–sample size combination for the four updated models as well as the estimated 1986 small-sample models. Figures 1 and 2 display these log likelihoods for the level-of-service and fully specified models, respectively. Points to note from these figures and Table 3 include the following.

First, in view of the significant transfer bias, the combined transfer procedure as expected yields results that are virtually indistinguishable from the 1986 small-sample results. At very small samples (e.g., 400 observations), the "prior" information provided by the estimation context parameters contributes a very marginal amount of additional information (resulting in a 0.2 percent improvement in the full-sample log-likelihood value for the level-

TABLE 1 Definition of Variables

dauto	= 1 for auto-drive mode; = 0 otherwise
dwalk	= 1 for walk mode; = 0 otherwise
aivtt	= auto in-vehicle travel time (min.) for auto mode; = 0 otherwise
tivtt	= transit in-vehicle travel time (min.) for transit mode; = 0 otherwise
twait	= transit wait time (min.) for transit mode; = 0 otherwise
twalk	= transit access + egress time (min.) for transit mode; = 0 otherwise
ivtc	= auto in-vehicle travel costs ($) for auto mode; = 0 for walk mode; = transit fare ($) for transit mode
pkcst	= auto daily parking cost ($) for auto mode; = 0 otherwise
wdist	= walk distance (km.) for walk mode; = 0 otherwise
avplic	= number of vehicles per licensed person in household for auto mode; = 0 otherwise
wcbd	= 1 if worker works in PD1 (Planning District 1) for walk mode; = 0 otherwise
amal	= 1 for male worker for auto mode; = 0 otherwise
tcbd	= 1 if worker destination is PD1 for transit mode; = 0 otherwise
tgend	= 1 if worker is female for transit mode; = 0 otherwise

of-service model and a 0.1 percent improvement for the fully specified model relative to the 1986 small-sample model results). Beyond this point, however, it is clear that the transfer scaling component of the procedure completely dominates the calculations. Because in this case the transfer scaling adjusts every parameter in the model, this is equivalent to simply reestimating the model on the basis of the small-sample application context data and using the reestimated parameters directly. To the extent that this result is verified in other empirical settings, it implies that combined transfer updating contributes little relative to simply reestimating the model on the basis of the application context small sample (a theme that is discussed more generally later), except perhaps in the case of extremely small samples.

Of the remaining procedures, the joint estimation procedure always performs the best, regardless of model specification or sample size used. This is not surprising given that the joint procedure is the only full-information procedure of the three. The improvement in performance achieved with the joint procedure increases with model specification: at the application sample size of 1,600, for example, the joint procedure reduces the log-likelihood value relative to the Bayesian procedure by only about 1 percent for the level-of-service model, whereas it generates about a 4 percent improvement for the fully specified model.

Conversely, the joint estimation procedure performs marginally better than the combined transfer procedure for small sample sizes for the simpler level-of-service model, whereas the combined transfer procedure performs slightly better than joint estimation at all sample sizes for the fully specified model. In comparing these two procedures, however, it should be noted that the combined transfer procedure effectively requires the estimation of $2(N + M - 1)$ parameters, where N is the number of alternatives and M is the number of utility function parameters (excluding alternative-specific constants); that is, $(N + M - 1)$ parameters from each of the estimation and application contexts. The joint context estimation procedure, on the other hand, requires $2(N - 1)$ alternative-specific constants, N scales, and M other utility function parameters to be estimated, for a total of $M + 3N - 2$ parameters, $M - N$ fewer than the combined transfer procedure.

Given $N = 3$ in this case, for the level-of-service model ($M = 7$) 18 parameters are estimated in the combined transfer model, whereas 14 parameters are estimated in the corresponding joint context model. This is a 29 percent increase in model parameters yielding no improvement in model performance below the 800 sample level and at most a 0.5 percent improvement in full-sample log likelihood over the entire range investigated.

Similarly, for the fully specified model ($M = 12$), 28 parameters are required by the combined transfer procedure versus 19 for the joint context procedure, a 47 percent increase in parameters, which yields at most a 3.3 percent improvement in full-sample log likelihood. Thus, joint context estimation would appear to be the more parsimonious of the two updating procedures, and, hence, all else being equal, preferred.

The constant/scale updating procedure performs surprisingly well at small sample sizes. For the level-of-service model, it performs virtually as well the joint procedure up to the 1,600 observation level and it clearly outperforms the computationally more complex Bayesian procedure up to at least the 6,400 observation level. The procedure's performance relative to the others decreases with improved model specification, but it is still comparable to the joint procedure at the 400 observation level and with the Bayesian procedure up to the 1,600 observation level for the fully specified model. This sensitivity to model specification is a sensible one, given that the relative role of constants (in particular) within the model should decline as model specification improves.

Given that small sample updating generally utilizes sample sizes in the order of 1,000 or less, these results imply that updating model scales and constants—a simpler and less onerous task than Bayesian updating—may well outperform the Bayesian procedure. The

TABLE 2 1964 (Estimation) Context Model Parameter Estimates

Parameter	Level of Service (I) Model		Fully Specified Model	
	Estimate	t-value	Estimate	t-value
dauto	0.090	0.452	-1.266	-4.362
dwalk	0.924	3.853	1.592	6.029
aivtt	-0.031	-10.731	-0.009	-2.217
tivtt	-0.043	-9.584	-0.029	-6.205
twait	-0.205	-12.444	-0.202	-11.653
twalk	-0.046	-3.205	-0.026	-1.729
wdist	-1.961	-21.892	-1.884	-20.935
ivtc	-0.389	-3.697	-0.388	-3.488
pkcst	-0.333	-11.583	-0.282	-9.134
avplic			1.874	12.185
amal			0.740	4.842
tcbd			1.224	9.142
tgend			0.759	5.299
wcbd			0.943	4.951
Number of observations	8066		8066	
Log-likelihood at Zero	-5929.6		-5929.6	
Log-likelihood at Convergence	-2839.4		-2590.5	
Adjusted rho-square	0.5204		0.5625	

TABLE 3 Full-Sample 1986 TTS Log-Likelihood Values for Alternative Models and Updating Sample Sizes

Model Type	Sample Size	Log-Likelihood Values				
		Bayesian Updating	Transfer Scaling	Joint Context Estimation	Combined Transfer Estimator	Small Sample
Level of Service Model	400	-11201	-11081	-11076	-11081	-11105
	800	-11156	-11076	-11075	-11077	-11086
	1600	-11133	-11052	-11045	-11012	-11014
	3200	-11072	-11012	-10991	-10937	-10937
	6400	-11013	-11000	-10963	-10930	-10930
	32328	-10942	-10996	-10931	-10920	-10920
Fully Specified Model	400	-10277	-10128	-10100	-9762	-9774
	800	-10138	-10025	-9903	-9621	-9626
	1600	-9999	-10013	-9805	-9555	-9555
	3200	-9865	-9952	-9622	-9471	-9470
	6400	-9712	-9945	-9545	-9470	-9470
	32328	-9501	-9941	-9475	-9453	-9453

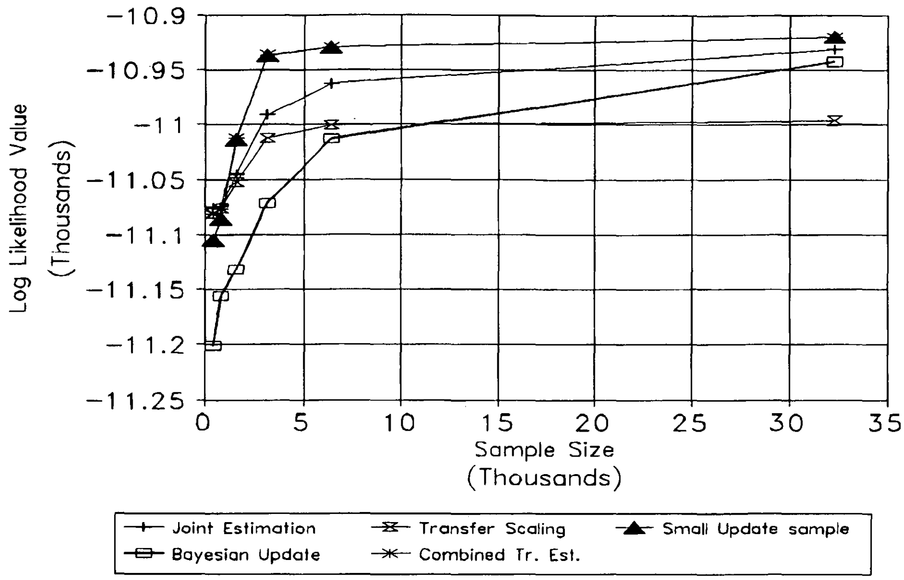

FIGURE 1 Full-sample 1986 TTS log likelihood values, level-of-service models.

results also imply that the joint estimation procedure may add little additional information to the updated model, relative to simply updating constants and scales, at least for sample sizes of 400 to 500 or less.

Comparison of the performance of the updated models with that of the 1986 small-sample models (i.e., the models simply estimated using the 1986 small samples) raises some question concerning the utility of updating a transferred model at all given the availability of an application context small sample. That is, the small-sample models outperform most of the updated models at most sample sizes. Thus, if one has a small sample of at least 400 to 500 observations, these results imply that one would be at least as well off to simply estimate an application context model, rather than to update a model developed elsewhere, especially if a relatively good specification is supported by the application data set.

Indeed, Table 3 and Figures 1 and 2 reinforce the importance of model specification in the determination of model performance by showing that the differences between the level-of-service model log-likelihoods and the corresponding fully specified model log likelihoods are far greater than the total differences between updating procedures or across sample sizes within either of the model specifications. In particular, note that the log likelihood for the 400-sample 1986 fully specified model of −9773.83 is larger than any of the full 32,328 sample level-of-service models.

Figures 3 and 4 present the aggregate MAE statistics for the four updated models and the 1986 small-sample models as a function of sample size for the level-of-service and fully specified models, respectively. The results here are less clear cut, reflecting the fact that different aggregations result in different combinations of com-

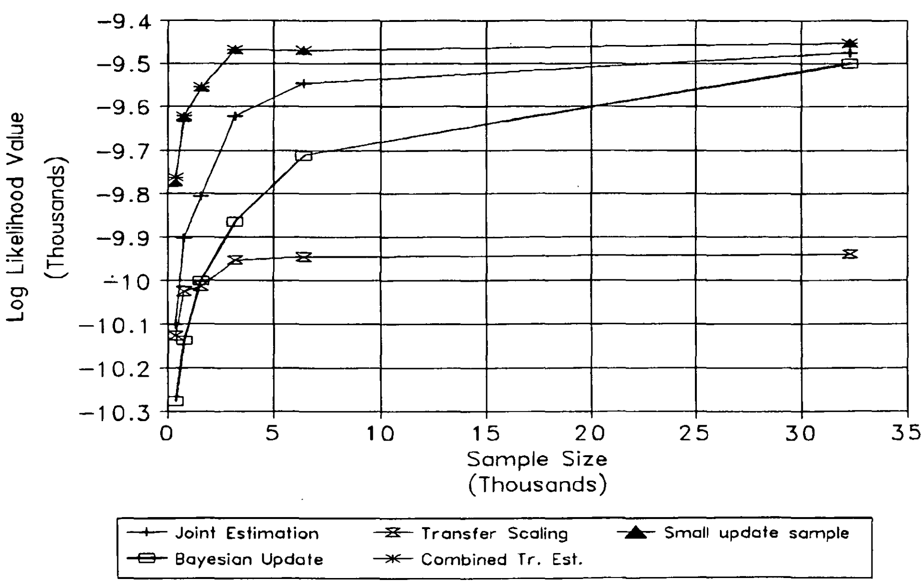

FIGURE 2 Full-sample 1986 TTS log-likelihood values, fully specified models.

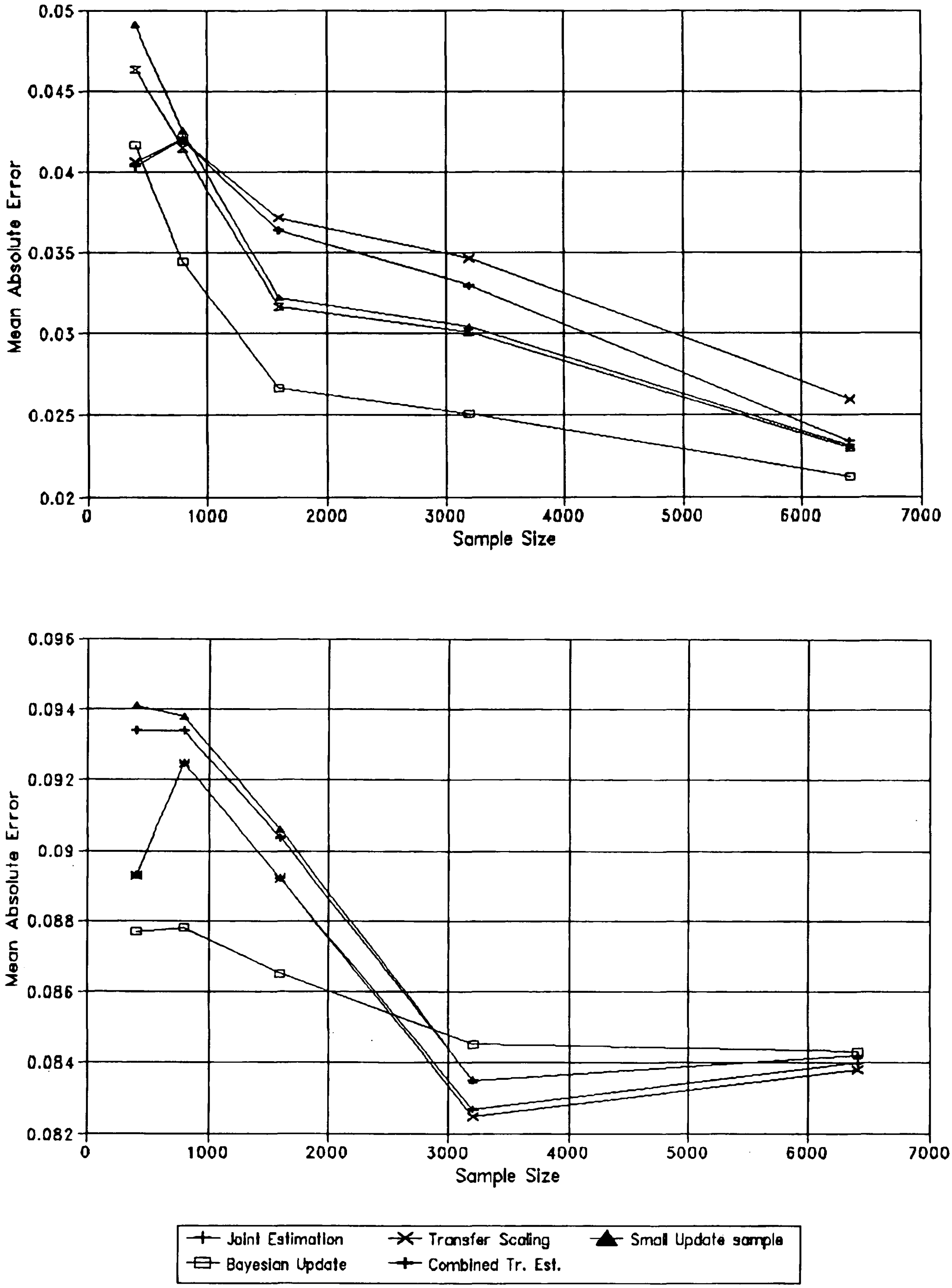

FIGURE 3 Aggregate mean absolute prediction errors, level-of-service models: *top,* aggregation by destination region; *bottom,* aggregation by worker gender.

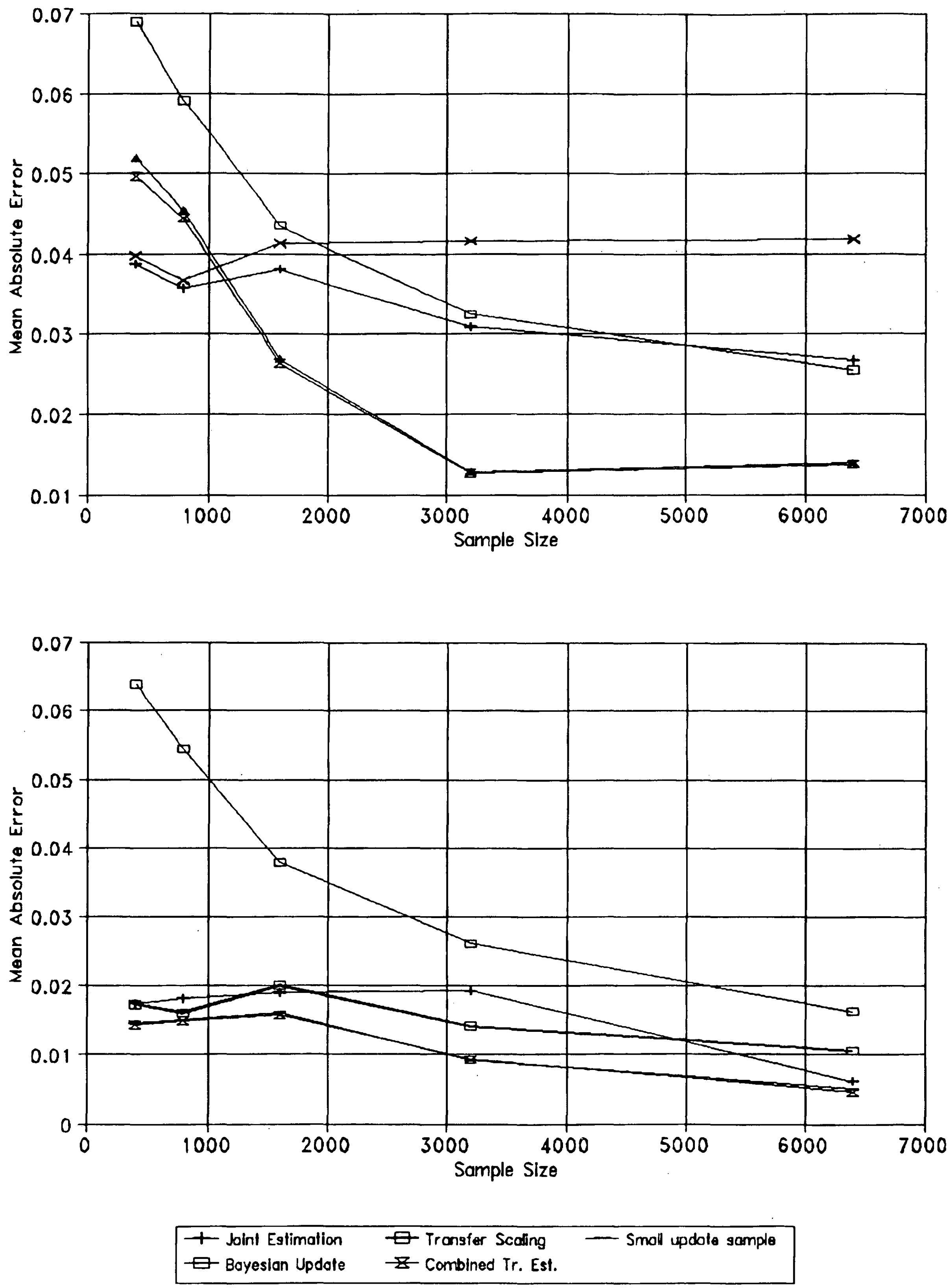

FIGURE 4 Aggregate mean absolute prediction errors, fully specified models: ***top,*** **aggregation by destination region;** ***bottom,*** **aggregation by worker gender.**

pensating errors. Nevertheless, some general trends are evident in these figures.

In the case of the level-of-service models and small sample sizes (e.g., under 1,000), the Bayesian procedure consistently yields the best aggregate predictions, the joint estimation and constants/scales updating generally yield results similar to one another that are slightly poorer than the Bayesian results, and the combine transfer and 1986 small-sample models generally yield the poorest aggregate predictions. The results for the fully specified model are more mixed but, in general, are different from the level-of-service results in that the combined transfer and 1986 small sample model results are, overall, the best, whereas overall the Bayesian procedure performs the most poorly (especially at sample sizes under 1,000). The other two updating procedures are again fairly comparable at small sample sizes and again generally lie between the best and the worst, although in this case their performance is generally close to the best.

Figures 3 and 4 again reinforce the importance of model specification in that the aggregate prediction errors are generally smaller for the fully specified model and the sensitivity to sample size is generally larger for the fully specified model as well.

SUMMARY AND CONCLUSIONS

This paper has provided an empirical comparison of four disaggregate choice model updating procedures using two data sets from the GTA representing travel behavior at two points in time 22 years apart (1964 and 1986). All the results obtained are based on this one case study, implying the need for additional tests employing other estimation/application contexts to be able to generalize any conclusions that arise from this study. On the basis of this study's results, however, the following findings are noteworthy.

1. The combined transfer estimation procedure consistently yields the best predictive performance in the 1986 application context, on the basis of the disaggregate full-sample log-likelihood measure used. This, however, is largely the result of the dominance of the transfer scaling component of the procedure, which effectively results in the procedure corresponding to a simple re-estimation of the model using the application context data set.
2. The joint context estimation yields results generally comparable to the combined transfer procedure, but with a significantly more parsimonious parameter structure. Hence, if the estimation context data set is available to support joint context estimation, generally it should be preferred relative to combined transfer estimation.
3. The computationally simpler transfer scaling procedure yields results that are similar to those of joint context estimation for small sample sizes. Hence, if the software required for joint context estimation or the estimation context data set, or both, are not available, then transfer scaling may well provide a useful and credible model update.
4. The Bayesian updating procedure is generally dominated by the other updating procedures examined, all of which explicitly deal with transfer biases in various ways. Thus, on the basis of this case study, Bayesian updating cannot be recommended as an updating procedure, especially given alternative techniques, such as transfer scaling and combined transfer estimation, which are not any more burdensome computationally and yet yield superior results.
5. Once the application context small sample reaches the 400 to 500 observation level, simply reestimating the model for the application context may yield results that are comparable or superior to any updated model transferred from an estimation context—providing that the application context data set supports development of a "fully specified" model.
6. Model specification is important in the updating/transfer process. In this case study, improving the model specification yielded far greater improvements in model performance than either "optimizing" the updating procedure or increasing the application context sample size.

In conclusion, this study indicates, in keeping with other studies cited in the paper, that updating a model estimated in another context through use of a small sample drawn from a new context significantly improves the model's transferability to this new context. In comparing the performance of a range of updating methods suggested in the literature, this study indicates that three procedures that all explicitly address the issue of transfer bias (transfer scaling, combined transfer estimation, and joint context estimation) all perform well at small sample sizes and possess merit as possible updating procedures for practical application. The choice among these methods depends on model specification, application context sample size, and availability of the estimation context data set. All else being equal, however, the joint context estimation procedure may be preferred given that it is a parsimonious, full-information approach to the problem.

ACKNOWLEDGMENTS

The work reported in this paper was funded by an operating research grant from the Natural Science and Engineering Research Council (Canada). Access to the travel behavior data bases and EMME/2 network modeling system used in the study was provided by the University of Toronto Joint Program in Transportation.

REFERENCES

1. Badoe, D. A., and E. J. Miller. Analysis of the Temporal Transferability of Disaggregate Work Trip Mode Choice Models. Presented at 74th Annual Meeting of the Transportation Research Board, Washington, D.C., Jan. 1995.
2. Atherton, T. J., and M. E. Ben-Akiva. Transferability and Updating of Disaggregate Travel Demand Models. In *Transportation Research Record 610*, TRB, National Research Council, Washington, D.C., 1976, pp. 12–18.
3. Gunn, H. F., M. E. Ben-Akiva, and M. A. Bradley. Tests of the Scaling Approach to Transferring Disaggregate Travel Demand Models. In *Transportation Research Record 1037*, TRB, National Research Council, Washington, D.C., 1985, pp. 21–30.
4. Koppelman, F. S., G. K. Kuah, and C. G. Wilmot. Transfer Model Updating with Disaggregate Data. In *Transportation Research Record 1037*, TRB, National Research Council, Washington, D.C., 1985, pp. 102–107.
5. McCoomb, L. A. Analysis of the Transferability of Disaggregate Demand Models Among Ten Canadian Cities. *Transportation Forum*, Vol. 3-1, 1986, pp. 19–28.
6. Ben-Akiva, M. E., and D. Bolduc. Approaches to Model Transferability and Updating: The Combined Transfer Estimator. In *Transportation*

Research Record 1139, TRB, National Research Council, Washington, D.C., 1987, pp. 1–7.
7. Ben-Akiva, M. E., and T. Morikawa. Estimation of Travel Demand Models from Multiple Data Sources. *Proc., 11th International Symposium on Transportation and Traffic Theory,* Yokohama, Japan, July 1990.
8. Morikawa, T., M. E. Ben-Akiva, and K. Yamada. Forecasting Intercity Rail Ridership Using Revealed Preference and Stated Preference Data. In *Transportation Research Record 1328,* TRB, National Research Council, Washington, D.C., 1991, pp. 30–35.
9. Traffic Research Corporation. *An Analysis of Some Travel Trends Between 1956 and 1964 Based on Home Interview Surveys.* Metropolitan Toronto Planning Board, Ontario, Canada, 1965.
10. *The Transportation Tomorrow Survey, Design and Conduct of the Survey.* Ministry of Transportation, Toronto, Ontario, Canada, 1987.
11. Data Management Group. *TTS Version 3: Data Guide.* University of Toronto Joint Program in Transportation, Ontario, Canada, 1990.

Publication of this paper sponsored by Committee on Passenger Travel Demand Forecasting.