Summary of Final Report, BD432
October 2005

ACCURACY STANDARDS AND DATA COVERAGE REQUIREMENTS FOR MODEL VALIDATION IN FSUTMS

PROBLEM STATEMENT

Model validation is an important step in travel demand modeling. It ensures that the calibrated model produces outputs that are consistent with the observed data. Various consistency checks have been suggested to validate the distribution and assignment models for highway networks in the Florida Standard Urban Transportation Model Structure (FSUTMS). They include (1) comparison of gravity model trip length outputs and the trip survey data outputs used to calibrate the model; (2) comparison of predicted link volumes and ground counts at screenlines, cordon lines, and/or cutlines; and (3) comparison of the derived statistics from the subject model and those from case studies of other areas with similar characteristics. Major criteria used for validation include systemwide vehicle miles traveled (VMT), vehicle hours traveled (VHT), link volume by facility type, area type, facility size (in terms of number of lanes), and volume group Root Mean Square Error (RMSE). In the transit modeling arena, accuracy standards for validating transit assignments have been extremely limited.

In FSUTMS, specific accuracy standards are prescribed in *Urban Transportation Planning Model Update-Phase II, Task C, Develop Standard Distribution and Assignment Models* (COMSIS, 1981) for each validation criterion to determine if further model diagnoses are needed. These accuracy standards are established as the critical benchmarks for gauging the model prediction capability and establishing the minimal levels for highway demand forecast model acceptance, primarily for determining the number of lanes for new roads. While these standards have been used for more than two decades, they are inadequate when the models are used for more than determining the number of lanes. Consequently, the current standards need to be amended.

OBJECTIVES

The main goal of this project is to set up and/or modify the accuracy standards of the validation task for the FSUTMS model. A set of procedures for validating this model that reflects the current household travel survey and transportation data available will be used. Specific objectives include the following:

- Develop a refined statistical methodology for consistency checking to support future model validation tasks.
- Review and enhance current validation accuracy standards using the latest count data collected from different types of facilities in different areas.
- Devise count sample size requirements at the desired level of statistical significance for the validation standards to facilitate consistency checking.
- Develop validation standards for the transit assignments.

FINDINGS AND CONCLUSIONS

The following are among the major findings and conclusions:

- Researchers provided information that will be useful to support future model validation tasks (pp. 51-59).

- In highway networks, sampling and model misspecification errors are considered in deriving accuracy standards. For sampling errors, except for several facility types with a limited sample size (less than 50), almost all error limits are below 10% (average is 5%) at the 95% level of confidence. Errors for one-way and toll facilities were slightly higher than others, which indicates that the variability of traffic counts observed on these facilities was higher generally.

- Allowable error limits to account for model misspecification errors are derived based on the perception error associated with travelers in choosing routes. The overall standards are derived by adding 5% sampling error on the model misspecification error, assuming the sampling error is independent of the model misspecification error. As a comparison, FHWA standards are about 10% higher than the proposed ones for ADTs 15,000 vehicles per day (vpd) or higher, and are substantially lower for ADTs below 15,000 vpd. The proposed accuracy standards meet Michigan DOT standards quite closely, except when the average assigned volume falls between 2,500 vpd and 15,000 vpd. The current practices by both FHWA and Michigan DOT seem to underestimate the assignment error by 10-15 percent within the abovementioned ADT range, as compared to the proposed standards.

- The current areawide RMSE standard of 35-50% seems arbitrary. Based on actual (or estimated) distribution of the roadway AADT groups, a worksheet is proposed for planners in different jurisdictional areas to develop their own areawide RMSE standards. Further, facility-type and size-of-facility specific accuracy standards should be eliminated, since they have created some contradictions to the standards by volume-group. While this recommendation may seem to contradict to the initial intent of this study (i.e., to accommodate the two-digit codes for facility type), elimination of these categories would simplify and standardize the structure of the current accuracy standards for highway networks.

- Based on the current accuracy standards, the required sample sizes are computed for each functional class area-wide at the 68%, 85%, and 95% levels of confidence. The results show that the required sample size increases significantly as the level of confidence increases. However, the required sample size remains at an affordable range between 10 and 60. Except for the one-way facility, which may require a larger sample (e.g., 50), the use of 30 count stations is a rule of thumb for general sampling purposes.

- For transit networks, the accuracy standards were derived to account for statistical variability, aggregation errors, and prediction errors. Errors in modal share subjected to aggregation and prediction variability are derived for both multinomial logit and nested logit models based on a second-order Taylor expansion theory. Variability of the modal share is also derived to establish confidence intervals at a user-specified level of significance as an alternative form of accuracy standards. Due to the limitation of data sources in both quantity and quality, the accuracy standards that account for aggregation and prediction errors could not be developed in this study. However, errors due to statistical variability are significantly lower than the current accuracy standards, in terms of daily ridership, which is to be expected since no other sources of error are considered.

BENEFITS

This study explicitly takes into account sampling and model specification errors for deriving the accuracy standards for model validation in FSUTMS. The derived accuracy standards reflect the notion that allowable error limits are more sensitive to link volume groups instead of facility types and sizes (number of lanes), which suggests that there should be more specific accuracy standards by volume groups and that facility-type and size-of-facility specific accuracy standards should be eliminated. These recommendations may seem to contradict to the initial intent of this study to accommodate the two-digit codes for facility type; however, they would greatly simplify and standardize the structure of current accuracy standards on highway networks.

Improved accuracy standards will result in increased confidence in the traffic forecasts produced by the subject models. With increased confidence in forecasts, engineers are more likely to produce optimal designs for new or enhanced (e.g., widening projects) facilities. The alternatives are overdesigned facilities and underdesigned facilities; the former are more costly than necessary, while the latter fail to meet the desired level of service.

This research project was conducted by Chang-Jen Lan, Ph.D., of the University of Miami. For more information, contact Shi-Chiang Li, Project Manager, at (954) 777-4655, shi-chiang.li@dot.state.fl.us