

**Proposed Validation and Sensitivity Testing of
Denver Region Activity-Based Models**

Prepared for
Innovations in Travel Modeling 2006 Conference

Prepared by
David L. Kurth
Senior Technical Consultant
Parsons

Suzanne Childress
Transportation Planner
Parsons

Erik Sabina
Denver Regional Council of Governments

Thomas Rossi
Principal
Cambridge Systematics, Inc.

May 21-23, 2006
Austin, Texas

Background

Traditional, four-step travel modeling procedures have evolved over the last half of the 20th century. Many of the innovations and improvements made to the four-step travel model process were made incrementally and subjected to either formal or informal validation and sensitivity tests. Formal validation tests were normally applied at the end of the model calibration process and, quite frequently, focused on the “super test” – the concept that reproduction of observed traffic volumes and transit boardings at some reasonable level of aggregation somehow showed that the models were, in fact, valid. Informal validation and sensitivity tests, unfortunately, too often consisted of discovering modeling problems after illogical travel forecasts were produced.

In the late 1990s, the Federal Highway Administration (FHWA) commissioned the development of a *Model Validation and Reasonableness Checking Manual* through the Travel Model Improvement Program. <1> This manual summarized validation standards used in various parts of the country and recommended a validation process that focused on the validation of the individual four-step model components as well as the traditional overall model system validation focused on traffic volumes and transit boardings.

Sensitivity testing has, by and large, been somewhat less formal. It has frequently focused on the sensitivity of individual model components using measures such as elasticity. Sensitivity testing of modeling systems by validating model results over time has not been as common since it requires observed travel data from more than one point in time for the same region. In addition, even if data for multiple points in time existed, many regions did not have measurable changes in their transportation system such as the significant addition of new roadway capacity or the opening of a new transit line. Without major changes to the underlying transportation system, it would be difficult to test the veracity of the underlying models (unless the test proved that the models were, in fact, poorly calibrated).

In 1997, the Denver Regional Council of Governments (DRCOG) initiated the collection of travel survey data to be used for the refreshing of their traditional four-step travel model and for the longer-term development of state-of-the-art modeling techniques. The “refresh” of the traditional model using these data took place from 2002 to 2004. The model validation included model component validation, validation to the 1997 base year and validation to travel conditions in 2001. While the time period between 1997 and 2001 was short, Denver’s light rail system was more than doubled in length during the period with the opening of the eight mile long Southwest LRT line, enhancing the effectiveness of transit component calibration/validation.

As with its current trip-based model, DRCOG is committed to rigorous validation and sensitivity testing of their activity-based modeling (ABM) system that will be developed over the next 18 months. The ABM is being designed to make use of many of the most recent developments in ABM theory in order to provide better representation of the travel decision making process and provide reasonable sensitivity to a wider range of future travel options and constraints. Nevertheless, DRCOG realizes that the increased detail of the ABM approach carries the price tag of an increased number of places where the models can fail. DRCOG has addressed this concern by committing approximately the same budget to the validation and sensitivity testing of the ABM that they committed to the entire refresh of the existing traditional four-step travel model. Perhaps a statement made by Chandra Bhat and Frank Koppelman in a recent web-based focus group discussion best summarizes the philosophy that DRCOG is using regarding their ABM development:

“...researchers and practitioners have not thought carefully enough about the criteria for validation of models. Researchers have the habit of asking practitioners to believe that activity-based methods will produce better impact assessment and forecasts because such models more appropriately represent the actual decision process (we plead guilty to this charge). There is a good basis for this line of thought, but researchers need to go beyond this argument. They need to develop clear validation criteria and demonstrate the value of activity-based methods in ways that are easily understood.” <2>

Since the ABM development process for the Denver region has just been initiated, this paper, by necessity, focuses on the initial plans for the validation and sensitivity testing of the models.

DRCOG Activity-Based Model Approach

The approaches to be used in ABM for the Denver region will, by and large, be based on approaches used in ABMs developed in other parts of the country, especially the ABM developed for the San Francisco Transportation Authority. Most of the components will be nested or multinomial logit models, sensitive to person and household demographic variables and transportation level of service variables. While detailed plans for the ABM model currently are being finalized, anticipated model components include:

- synthetic population generator;
- regular workplace location choice model for each worker;
- regular school location choice model for each student;
- household auto ownership choice model;
- daily activity pattern choice model for each person-day;
- number of tours choice model for each person day;
- work-based sub-tour generation;
- tour level destination choice;
- tour level mode choice;
- tour level time-of-day choice;
- trip level destination choice;
- trip level mode choice (conditional on tour mode choice); and
- trip level time of day choice (conditional on time windows remaining after all previous choices).

Several model components will be transferred or adapted from the existing four-step model for the region. Examples of these models include the area type model, the parking cost model, and traffic and transit assignment procedures.

DRCOG Validation and Sensitivity Testing Approach

A validation plan <3> describing all of the validation tests to be conducted for the ABM components and the overall model system has been developed concomitantly with the specification of the ABM for the region. The validation plan includes the standards by which the tests will be evaluated. It is anticipated that the validation tests for the model components will include the following:

- Checks to ensure that the model component is producing the correct results (i.e. verification of computations);

- Comparisons of model parameters to comparable parameters in similar models in other areas;
- Disaggregate validation of all model components estimated using disaggregate methods through the application of the model component using the estimation data set and comparison the model outputs to the estimation data;
- Testing of each model's sensitivity to variables through controlled modification of those input variables;
- Comparisons of the model component outputs to the results from the survey data set; and
- Comparisons, where data are available, of the base year outputs from each model component to independent observed data (e.g. comparisons of mode choice model outputs to linked trips estimated from transit boarding counts).

The above tests are typical of model validation tests that have been recommended in documents such as the *Model Validation and Reasonableness Checking Manual*. Such tests should be performed for all model development efforts.

Since there are more components in the proposed ABM than in a conventional trip based model, there will be significantly more component testing. It will be important to design validation tests that are appropriate for each component. While some of the tests will be analogous to those performed for components of trip based models, others will be completely different. Examples of tests similar to those used for trip-based models include comparisons of modeled trip length frequencies to those from the household survey (although tour lengths must also be compared) and comparisons of modeled and observed mode shares. Examples of tests to be performed without comparable trip based tests include the number of trips per tour by purpose, amount of time spent in activities versus traveling on tours by purpose, and the number of activities performed by each person. The DRCOG model validation plan <3> provides a complete list of all tests to be performed.

One difficulty in performing the "new" validation tests is the lack of experience to determine standards for validation. For example, how close should the modeled number of activities per person be to the observed number? In some cases, established standards for trip based models may be used to inform the choice of standards for the ABM. In other cases, the acceptable error ranges will be determined by estimating the variation expected in aggregate model statistics (such as vehicle miles traveled) resulting from the deviation on the particular model component, in effect tying new validation tests that as yet have no standards to existing tests that do have such standards.

Sensitivity testing will also play an important role in the validation of the ABM components. The amount of change in the results of applying a model component relative to changes in input variable values will be evaluated. Again, there are no established standards for reasonable elasticities for some of the newer model components, and so this will be more of a reasonableness test than anything else. In such cases, it may not be possible to tie these tests to existing tests with standards already developed. In such cases, it may be possible to compare Denver model test outcomes to observed outcomes in other cities with conditions similar to those being evaluated in the sensitivity tests.

In addition to the model component calibration and validation tests outlined above, significant effort will be placed on the validation of the overall model system. Again, this validation will be comparable to previously recommended validation procedures and will include the following:

- Reasonableness and logic checks of demographic and network data/skim data input to the models;
- Traditional validations for the model estimation year (1997) and for 2005 against independent observed data. Depending on data availability, these traditional checks will include:
 - Root-mean-squared-error of modeled to observed traffic volumes by appropriate segmentation variables (such as facility type, traffic volume level, etc.).
 - Matching regional observed VMT within approximately 1% error.
 - Matching observed VMT by facility type.
 - Matching VMT by area type.
 - Matching total transit boardings.
 - Matching transit boardings by sub-mode.
 - Rapid transit boardings by corridor, sub-mode and station.
 - Park and ride lot usage.
 - Matching a series of at least 10 highway and transit screenline volumes.
 - Highway volumes on individual freeways.
 - Toll road usage.
 - Acceptable matching of peak and off-peak speeds.
 - Roadway speeds by several time-of-day periods (AM vs. PM vs. off peak, etc.)
- Tests of the sensitivity of the overall model system to input variables (similar to the procedures used for the model component sensitivity testing).

The ABM will be subjected to the same validation standards that were used for the recently refreshed four-step model. Validation results for the ABM are expected to be as good as or better than those produced using the four-step model. While this may seem to be a rather lenient standard, it must be remembered that the four-step model was, in fact, calibrated to produce reasonable validation results for 1997 and 2001. One would expect that since the ABM can consider more aspects of personal activity performance and travel behavior, the amount of “adjustment factoring” not tied to specific measurable behavior should be less in the ABM.

Another validation activity under consideration is “back-casting” to a prior year (besides the model estimation year of 1997). This will be done if the necessary data are available and the resources to perform the back-cast are available.

In addition to specifying traditional model validation standards, input and coordination with Federal agencies will also be sought in the validation and sensitivity testing of the ABM for the Denver region. This coordination will be particularly important in the development of the ABM since operational experience with them is limited; Federal agencies may be expected to evaluate them closely for validity and for consistency with the calibration outcomes of the numerous trip-based models in existence. DRCOG intends to involve federal agencies early in the model development process, through oversight panels or other means, and will include their requirements in the calibration/validation plan at the earliest possible point.

Temporal and Policy Sensitivity Testing

The normal validation testing outlined above includes some temporal validation in that the model will be validated against observed travel data for 1997 and 2005. Such testing is crucial for model validation but does not address the hypothesized true value of ABMs – the production of better impact assessment information and travel forecasts that will result from the more appropriate representation of the actual decision process.

Two primary approaches will be used to test the sensitivity of the ABM. The first will be the application of the ABM for an existing forecast year and scenario and comparison of those results to results produced by the calibrated four-step model. While the “true” results for such a forecast year cannot be known, the results from the existing four-step model provide one outcome that has been deemed reasonable. Several questions will be asked:

- How similar are the results? The traditional validation measures outlined in the previous section regarding traffic volumes and transit boardings can be used to measure the similarity.
- Which model produces “more believable” results? Two outcomes from the comparison of the results are possible: the forecasts from the two models are not substantially different or the forecasts are substantially different. In either case, an assessment will need to be made whether or not the outcome is acceptable since either outcome, ultimately, will need to be presented to and supported to local decision makers.

The second approach for testing the ABM will be directed at assessing the desire to develop a model that is more sensitive to policy variables. The policy-oriented tests will include:

- Evaluation of outcomes in designated transit-oriented development areas.
- Evaluation of effects of different regional development densities (e.g., single family housing versus multi-family, etc.).
- Evaluation of development in known industrial areas.
- Evaluation of development of specific “greenfield” areas, to see how well the model can predict the spread of the urban area.
- Evaluation of outcomes in redevelopment areas.

The policy oriented tests will be even more subjective than the comparison of forecasts from the ABM and the traditional four-step model. To improve the usefulness of the tests, it will be important to reach a reasonable consensus regarding the expected outcome of the tests. If the outcome from the model does not match the expected outcome, the model results will need to be assessed to determine whether they are illogical or providing valuable information that would modify the expected outcomes developed prior to running the model. As discussed above, the definition of reasonableness may be derived from observed conditions in other cities.

References

1. *Model Validation and Reasonableness Checking Manual*, Travel Model Improvement Program, Federal Highway Administration, February 1997.
2. Bhat, C. and F. Koppelman, Travel Model Improvement Program E-mail dated December 16, 2005.
3. Parsons Corporation and Cambridge Systematics, Inc. “Validation Plan for the DRCOG Activity Based Model,” March 2006.