Estimating Risks and Model Credibility

Session 1:
Model Credibility as a Criteria for Design and Implementation

Date:
December 12, 2013

Speaker:
Rick Donnelly
Disclaimer

The views and opinions expressed during this webinar are those of the presenters and do not represent the official policy or position of FHWA and do not constitute an endorsement, recommendation or specification by FHWA. The webinar is based solely on the opinions and experience of the presenters and is made available for knowledge and experience sharing purposes only.
• Do high-level decision-makers view model validity different than we do?
• If so, what can we learn from them?
• Is there any difference between the views held by public and private executives?
The model is **designed** to meet specific analytical needs. The **specification** is the roadmap for building and testing it.

Formal statistical methods are used to **estimate** some parts of the model, while other parts are **synthesized** from models built elsewhere, reported in the literature, or asserted by the analyst in absence of statistically reliable data.

The estimated or synthesized pieces of the model are tested as a whole against observed flows and patterns, a process called **calibration**. An existing model might be compared to newer or different data than was used to estimate or calibrate the model, a process called **validation**.

A process collectively called **model development**
How We View Model Validity

• Entirely a statistical exercise
• Heavy focus on matched observed traffic counts
• Start when development is over
• Matching pre-conceived ideas about accuracy
  – Comparison with competing models
  – DOT guidelines
  – Accepted norms
Cooking the Numbers

OLS: adj r² = 0.67, n = 1114

OLS: adj r² = 0.97, n = 1175
Data Points

- Investment bankers (3)
- State legislators (3)
- State DOT directors (2)
- HSR board member
- HSR executive director
- Governor
- State Transportation Commissioner (2)
- Congressional legal counsel
- CEO of infrastructure consulting firm
Data Collection

• Two questions
  – “How do you assess the validity of a travel model or forecast derived from one?”
  – “What could forecasters do to increase your confidence in their work?”

• Administration
  – Relaxed social setting
  – After interacting with forecasts
  – No advance warning

• Ad hoc analysis
## Assessing Validity (Q1)

### Accomplished modelers (n=15)

<table>
<thead>
<tr>
<th>Criteria</th>
<th># mentions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Replicates traffic counts</td>
<td>9</td>
</tr>
<tr>
<td>Theoretically plausible</td>
<td>5</td>
</tr>
<tr>
<td>Established practice</td>
<td>5</td>
</tr>
<tr>
<td>Matches calibration targets</td>
<td>5</td>
</tr>
<tr>
<td>Software verification</td>
<td>4</td>
</tr>
<tr>
<td>Sensitivity testing</td>
<td>2</td>
</tr>
<tr>
<td>Market understanding</td>
<td>2</td>
</tr>
<tr>
<td>Parallel studies</td>
<td>2</td>
</tr>
<tr>
<td>Peer review panel</td>
<td>2</td>
</tr>
</tbody>
</table>
## Assessing Validity (Q1)

### Accomplished modelers (n=15)

<table>
<thead>
<tr>
<th>Criteria</th>
<th># mentions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Replicates traffic counts</td>
<td>9</td>
</tr>
<tr>
<td>Theoretically plausible</td>
<td>5</td>
</tr>
<tr>
<td>Established practice</td>
<td>5</td>
</tr>
<tr>
<td>Matches calibration targets</td>
<td>5</td>
</tr>
<tr>
<td>Software verification</td>
<td>4</td>
</tr>
<tr>
<td>Sensitivity testing</td>
<td>2</td>
</tr>
<tr>
<td>Market understanding</td>
<td>2</td>
</tr>
<tr>
<td>Parallel studies</td>
<td>2</td>
</tr>
<tr>
<td>Peer review panel</td>
<td>2</td>
</tr>
</tbody>
</table>

### Decision-makers (n=15)

<table>
<thead>
<tr>
<th>Criteria</th>
<th># mentions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Independent review</td>
<td>11</td>
</tr>
<tr>
<td>Confidence in analyst</td>
<td>11</td>
</tr>
<tr>
<td>Comparable forecasts</td>
<td>10</td>
</tr>
<tr>
<td>Squares with intuition</td>
<td>8</td>
</tr>
<tr>
<td>Free from obvious flaws</td>
<td>8</td>
</tr>
<tr>
<td>Agency/investor buy-in</td>
<td>6</td>
</tr>
<tr>
<td>Effective presentation</td>
<td>6</td>
</tr>
<tr>
<td>Established practice</td>
<td>3</td>
</tr>
<tr>
<td>Aligns with theory</td>
<td>2</td>
</tr>
</tbody>
</table>
But Do They Really Believe Us?

Possible

Plausible

Probable

Source: Adapted from Voros (2003) and Hancock & Bezold (1994)
Solutions (Q2)

• Lose “replication tunnel vision”
• Make risk and uncertainty explicit
• Standards
• Transparency
  – Delphi panel
  – Open models (data + assumptions + tools)
  – “Stakeholder access”
• Reference class forecasting (outside view)
• Independent review
• Crash testing
• Model alignment
Purposeful Model Design

Current and expected information needs informed by travel analysis methods

- Analytical requirements
- Institutional constraints

Informed by best practices and validation standards

- Model acceptance criteria
- Performance measures

Includes usability, interface, and connectivity considerations as well as desired outputs

- Required functionality

Evidence-based

Includes both existing and emerging approaches

- Candidate methods and techniques

An ideal solution generates performance measures at the desired levels of resolution and fidelity

Appropriate tools and approaches
Explicitly Modeling Risk and Uncertainty

Actors:  
- **DM**: Decision-makers and executives  
- **EP**: Expert panels, peer review  
- **DP**: Delphi panels

Not shown: modelers, planners, and the public
AND MY REVENUE FORECAST SAYS... DID YOU MAKE ANY ASSUMPTIONS?

I MADE A LOT OF THEM.

THEN WE DON'T BELIEVE YOUR FORECAST.

CAN I TELL YOU ABOUT IT ANYWAY?

DO WHATEVER MAKES YOU FEEL LESS ABSURD.

Rick Donnelly | PB | donnellyr@pbworld.com | 505-881-5357
Estimating Risks and Model Credibility

Session 2: Estimating Risks Associated with Travel Demand Forecasts

Date: December 11, 2013

Speaker: Tom Adler (RSG)
The Data
The Data– Transit New Starts Forecasts

- Travel demand model forecasts are not always accurate

Source: Federal Transit Administration, TRB Session 371, January 2008
• Travel demand model forecasts are not always accurate

Three Major Sources of Inaccuracy

• Model structure and data
  – Travel forecasting models are not perfect representations of actual travel behavior
  – Data used in models are not perfect representations of transportation system

• Analysis bias
  – General tendency to be overly optimistic
  – May be influenced be desired outcome

• Inherent uncertainties about future
  – Large uncertainties in many important future conditions
  – Uncertainties interact in ways not identified by simple sensitivity analyses
Ways of Addressing the Problem
There is broad recognition of issues with forecasting models and of the continuing need to improve their reliability.

- Improve the models that are used for forecasting
  - Progress has been made along a number of fronts
    - Model structure
    - Model inputs

- Improve the review and evaluation process
  - Lesson from transit ridership forecasting

- Recognize and quantify inherent uncertainties and simplifications
  - Quantified probability analysis

**Comparisons of FTA New Start predicted vs. actual ridership by year**

<table>
<thead>
<tr>
<th>Study</th>
<th>Average</th>
<th>50th Percentile</th>
</tr>
</thead>
<tbody>
<tr>
<td>“Pickrell” Report 1990</td>
<td>42.4%</td>
<td>41.4%</td>
</tr>
<tr>
<td>FTA 2003</td>
<td>68.9%</td>
<td>70.5%</td>
</tr>
<tr>
<td>FTA 2007</td>
<td>74.5%</td>
<td>63.8%</td>
</tr>
</tbody>
</table>

Source: Federal Transit Administration, TRB Session 371, January 2008
The Objective

- Produce estimates of the likelihood of obtaining different ranges of model outputs
One Approach: Quantified Probability Analysis

• Quantified probability analysis involves two steps:
  1. Estimating the probability distribution associated with each model uncertainty
  2. Estimating the resulting probability distribution of the model outputs

• Step 1 can use a combination of quantitative and qualitative assessment – still better than just ignoring

• Step 2 can in some cases be calculated analytically, and in most other cases requires Monte Carlo simulation which means calculating outcomes for many, many different combinations of inputs (for each project construction scenario and each year)

• HOWEVER – most travel demand models take hours or days to run and so running them many times as required by Monte Carlo simulation becomes an intractable problem
  – Newer dynamic/agent-based/activity-based models are especially computation time-intensive
Ways to Quantify Probabilities of Different Outcomes

• Approaches for “Step 1” – Estimating ranges for inputs
  – Historical data
  – Expert judgment
  – Delphi method

• Approaches for “Step 2” – Quantifying probabilities of outcomes
  – Use simple variable-by-variable sensitivity analysis to estimate effects
    • Ignores interactions among inputs, can be computationally inefficient
  – Use Bayesian melding
    • Requires data on past outcomes
  – Create new “structure and logic” model
    • May produce results that are inconsistent with T&R model
  – Streamline T&R model for faster computation
    • May be difficult to accomplish
  – Create synthesized model using statistical response surface methods
    • Requires structured forecasting model runs, statistical analysis
Uses additional steps to create information that supports a more complete enumeration of all reasonable future conditions.

Key insight is that corridor traffic can be modeled as a multivariate closed-form function of model inputs.

Source: Technical Memorandum Planning-Level Traffic and Revenue Study Interstate 4 Tolled Managed Lanes, URS Corporation, October 2012
Case Study
Case Study: Orlando I-4 Traffic and Revenue Study

• Proposed 21-mile managed lane project
  – Existing highway has 4 lanes in each direction
  – Project will add 2 dynamically-priced lanes in each direction
  – Toll revenue will be important piece of project finance
  – Needed 75th percentile traffic and revenue estimates

• Regional travel demand model used to estimate demand
  – Travelers’ values of time and regional growth rates are key drivers of demand
  – The actual future growth rate is highly uncertain
  – Model system takes 4-5 hours for a single run
Probability Distributions of Inputs

- **Value of time (VOT)** distribution based on survey sampling error.
- **Land use growth** distribution based on historical accuracy of Florida’s official land use forecasts (BEBR).
### Forecasting Runs for Response Surface Modeling

- Forecasting model runs: 9 orthogonal fractional factorial experiments
- Different from sensitivity analysis where inputs are varied individually

#### I-4 Managed Lanes Experimental Design

<table>
<thead>
<tr>
<th>Alternative</th>
<th>VOT/hour</th>
<th>Economy</th>
<th>Network</th>
<th>Toll Rate/Mile</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>$10.67</td>
<td>BEBR Med-Low</td>
<td>E+C</td>
<td>$0.05</td>
</tr>
<tr>
<td>P2</td>
<td>$10.67</td>
<td>BEBR Med</td>
<td>150</td>
<td>$0.15</td>
</tr>
<tr>
<td>P3</td>
<td>$10.67</td>
<td>BEBR Med-High</td>
<td>125</td>
<td>$0.10</td>
</tr>
<tr>
<td>P4</td>
<td>$18.08</td>
<td>BEBR Med-Low</td>
<td>125</td>
<td>$0.10</td>
</tr>
<tr>
<td>P5</td>
<td>$18.08</td>
<td>BEBR Med</td>
<td>150</td>
<td>$0.05</td>
</tr>
<tr>
<td>P6</td>
<td>$18.08</td>
<td>BEBR Med-High</td>
<td>E+C</td>
<td>$0.15</td>
</tr>
<tr>
<td>P7</td>
<td>$25.49</td>
<td>BEBR Med-Low</td>
<td>125</td>
<td>$0.15</td>
</tr>
<tr>
<td>P8</td>
<td>$25.49</td>
<td>BEBR Med</td>
<td>E+C</td>
<td>$0.10</td>
</tr>
<tr>
<td>P9</td>
<td>$25.49</td>
<td>BEBR Med-High</td>
<td>150</td>
<td>$0.05</td>
</tr>
</tbody>
</table>

Source: Technical Memorandum Planning-Level Traffic and Revenue Study Interstate 4 Tolled Managed Lanes, URS Corporation, October 2012

E+C is existing plus committed highway projects; 150 and 125 represent additional improvements necessary to keep traffic on all links within 150% and 125% of capacity, respectively.
The Synthesized I-4 Models

Traffic = 128686 + 48486 * growth – 91138245 * (tollRate / VOT) + 9432 * ln(rampUp) – 11099 * roadEC – 15820 * road150 + yearCon

Revenue = –6130 + 7.9 * (Traffic * tollRate) + 3532 * roadEC + 2088 * road150 + yearCon

Where: Traffic is the number of daily one-way trips that use the I-4 Express Lanes Revenue is the estimated gross revenue from the trips on the I-4 Express Lanes growth is the ratio of dwelling units in the given year to dwelling units in 2010 minus one tollRate is the average toll rate charged on I-4 in 2010 $ rampUp is the number of years the project has been operating in the given year roadEC represents the road improvements included only in the E+C conditions road150 represents an improvement program that maintains all roads below V/C of 1.5 yearCon is a vector of constants representing the years for which the forecasts are being made

Model R² = 0.98

Source: Technical Memorandum Planning-Level Traffic and Revenue Study Interstate 4 Tolled Managed Lanes, URS Corporation, October 2012
Monte Carlo Model

• Many easy-to-use software packages
Illustrative Distributions of Outputs*

*Note: The data in these graphs are illustrative only and do not reflect the final I-4 project forecasts.

Outputs reflect 1 million Monte Carlo draws from input distributions; runs took ~30 min. each.
Other Relevant Work
Another Approach: Empirical “Evidence-based”

- Robert Bain’s current work
- Compiled information about performance of toll road model estimates vs. actual toll traffic
- Distilled information into simple model with dependent variable as the width of the confidence interval

Width of 90% confidence interval $\propto \sqrt{n}$, where $n$ is the forecast horizon in years
Conclusions
Conclusions

• Uncertainties and, thus, risks are inherent in travel demand forecasts

• For some types of projects, such as those financed by toll revenues, quantification of risks is especially important

• Travel demand models are too cumbersome to be used to directly simulate the probability distributions of their outputs

• Simple sensitivity analyses that are commonly conducted do not provide robust information about these distributions because they do not consider interactions among variables

• Response surface methods can be used to develop closed-form models that very effectively estimate the effects of key model inputs on corridor traffic and revenue

• Response surface models in turn can be used to effectively simulate risks associated with much more complex travel demand forecasting models

• Where data are available “evidence-based” approaches can also be used to estimate uncertainties in forecasts

• Whatever approach is used, it is important to formally recognize inherent uncertainties in travel demand forecasts
Questions?

For More Information

Contact: Tom Adler
RSG
tadler@rsginc.com