An Adaptation of the Incremental Logit Model for Forecasting Work Zone Diversion

John Miller, Hillary Goldstein, and Emily Seay
Virginia Center for Transportation Innovation and Research

April 29, 2014
Problem: Forecast Diversion Due to Work Zones Quickly
Availability of Historical Data

- Can get travel times for crossing paths
- Can get volumes at crossing nodes
- Often cannot get precise duration of previous closures

Figure courtesy of University of Maryland’s Regional Integrated Transportation Information System (www.ritis.org)

Figure from Virginia Department of Transportation Traffic Engineering Division TMS Count Database
Goal of the Tool

- Adapt the incremental logit model to forecast potential traffic volume changes at the water crossings based on work zone delay.

- Scope: limited to forecasting diverted volumes

- Rationale: A credible methodology for forecasting delay based on work zones already exists
We Benefitted from 3 Ideas

- **Idea 1. Traffic Shift Methodology for Corridors**

- **Utility of route** $i = \theta$ (**Travel time for route** $j$)

\[
P_i = \frac{e^{U_{routei}}}{e^{U_{routei}} + e^{U_{routej}}} = \frac{e^{\theta t_j}}{e^{\theta t_i} + e^{\theta t_j}}
\]

Utility is based on $\theta$(time)
Idea 2. Calibrate $\theta$ From Multiple Observations

If difference in travel times has little impact on volume, then $\theta$ will be closer to zero.
Idea 3. Incremental Logit Model

\[
P_i' = \frac{P_i e^{\Delta U_{\text{route}i}}}{P_i e^{\Delta U_{\text{route}i}} + P_j e^{\Delta U_{\text{route}j}}}
\]

\[
P_i' = \frac{P_i e^{\theta \Delta t_i}}{P_i e^{\theta \Delta t_i} + P_j e^{\theta \Delta t_j}}
\]

\[
\Delta t_i = \text{delay attributable to the work zone at route } i \text{ rather than the travel time on route } j
\]
How do we know what value of $\theta$ to use?

- Consider 3 alternative crossings with average volume percentages for a certain time of day (say weekday mornings)

  \[ P_{i}^{\text{mean}} = 50\% \quad P_{j}^{\text{mean}} = 25\% \quad P_{k}^{\text{mean}} = 25\% \]

- Suppose for a given period (say this Tuesday morning)

  \[ P_{i}' = 40\% \quad P_{j}' = 28\% \quad P_{k}' = 32\% \]

  \[ t_{i}' = 15 \text{ min} \quad t_{j}' = 7 \text{ min} \quad t_{k}' = -1 \text{ min} \]

  (slower than usual) (faster than usual)

- Maximize \((P_{i}'^{40\%})(P_{j}'^{28\%})(P_{k}'^{32\%})\) where

  \[
P_{i}' = \left[ P_{i}^{\text{mean}} e^{\theta t_{i}'} + P_{j}^{\text{mean}} e^{\theta t_{j}'} + P_{k}^{\text{mean}} e^{\theta t_{k}'} \right] e^{\theta t_{i}'}
  \]

Maximize \(40\% \ln(P_{i}') + 28\% \ln(P_{j}') + 32\% \ln(P_{k}')\)
Data Collection: Travel Times and Volumes (March 2012-May 2013)

- Obtained travel times for paths by hour
- Excluded toll crossing
- Obtained crossing volumes by hour

©2013 Google
Humbled by Initial Calibrations with Real Data

- $\theta_k$ is positive!!! More delay makes this route more attractive.

- Initial assumptions
  - A complex model is better than a simpler one.
  - We should automatically use all of the data.

- Solution
  - Let $\theta_i = \theta_j = \theta_k = \theta$.
  - After computing a mean delay and volume, select periods in which delay exceeded mean delay by 2 standard deviations.
Are $\theta$ Statistically Significant?

- Likelihood Ratio Test

- Friday eastbound

$$
\sum_{z=1}^{77} O_i^z \ln(P_i^{z}) + O_j^z \ln(P_j^{z}) + O_k^z \ln(P_k^{z})
$$

<table>
<thead>
<tr>
<th>Quantity</th>
<th>Log likelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta = 0$</td>
<td>-77.660</td>
</tr>
<tr>
<td>$\theta = -0.0894$</td>
<td>-76.656</td>
</tr>
<tr>
<td>2 $</td>
<td>\text{Difference}</td>
</tr>
<tr>
<td>$\rho$-value</td>
<td>0.16</td>
</tr>
</tbody>
</table>
## Example of Results

<table>
<thead>
<tr>
<th></th>
<th>Eastbound</th>
<th>Westbound</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Summer</strong></td>
<td>-0.1023</td>
<td>-0.0932</td>
</tr>
<tr>
<td>(Memorial Day-Labor Day)</td>
<td>(p = 0.02)</td>
<td>(p = 0.01)</td>
</tr>
<tr>
<td><strong>Winter</strong></td>
<td>-0.0880</td>
<td>-0.0872</td>
</tr>
<tr>
<td>(December 1-February 28)</td>
<td>(p = 0.12)</td>
<td>(p &lt; 0.01)</td>
</tr>
<tr>
<td><strong>Monday-Tuesday</strong></td>
<td><strong>-0.0872</strong></td>
<td><strong>-0.0867</strong></td>
</tr>
<tr>
<td></td>
<td>(p &lt; 0.01)</td>
<td>(p &lt; 0.01)</td>
</tr>
<tr>
<td><strong>Friday</strong></td>
<td>-0.0894</td>
<td>-0.0855</td>
</tr>
<tr>
<td></td>
<td>(p = 0.16)</td>
<td>(p &lt; 0.01)</td>
</tr>
</tbody>
</table>
To Finish:

- How are these parameters used?
- How do these $\theta$ values compare with those from other studies?
- To try this elsewhere, what should operators consider?
- What weaknesses can be resolved?
How Are These Parameters Used?

<table>
<thead>
<tr>
<th>Crossing</th>
<th>Normal volume</th>
<th>Normal percentage</th>
<th>Work zone delay</th>
<th>Diversion parameter</th>
</tr>
</thead>
<tbody>
<tr>
<td>US-58E</td>
<td>1,034</td>
<td>14%</td>
<td>25.5</td>
<td>-0.0335</td>
</tr>
<tr>
<td>I-264E</td>
<td>2,420</td>
<td>33%</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Route 337</td>
<td>163</td>
<td>2%</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>US-13N</td>
<td>524</td>
<td>7%</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

θ = -0.034 (SHRP 2)
θ = -0.050 (NCHRP 365 if like OVTT)
θ = -0.087 (calibrated)
<table>
<thead>
<tr>
<th>Crossing</th>
<th>US-58E</th>
<th>I-264E</th>
<th>Route 337</th>
<th>US-13N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elizabeth River</td>
<td>1,034</td>
<td>2,420</td>
<td>163</td>
<td>524</td>
</tr>
<tr>
<td>River Bridge</td>
<td>14%</td>
<td>33%</td>
<td>2%</td>
<td>7%</td>
</tr>
<tr>
<td>Midtown Tunnel</td>
<td>25.5</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Tunnel</td>
<td>-0.0872</td>
<td>1.0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Diversion parameter</td>
<td>2%</td>
<td>37%</td>
<td>2.51%</td>
<td>8%</td>
</tr>
<tr>
<td>New Percentages</td>
<td>128</td>
<td>2,766</td>
<td>186</td>
<td>599</td>
</tr>
<tr>
<td>New Volumes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Areas for Exploration in Practice

- Estimation of volumes at locations without counts
- Estimation of toll parameters where applicable
- Selection of diversion parameters
- Manner in which delay is calculated
- Use of an hour as the unit of analysis
Areas for Exploration of Theory

- **Positive coefficients**
  - Is there heterogeneity from incident publicity, a need to nest alternatives, or some other factor?

- **Highly sensitive weekend values**
  - A reviewer suggested we may need more specific origins and destinations

- **Significance testing**
  - The impact of the perfect model not having a likelihood of zero is not known.

- **Selection of paths relative to volume locations**
Conclusions

1. Calibrating the incremental logit model, based on the sensitivity of changes in traffic volume to changes in delay, is an attractive option.

2. The sign of the diversion parameter is logical and is expected provided a single parameter and appropriate data subset are used.

3. The diversion parameter shows relatively little variation when the data are stratified by season or weekday.

4. Diversion parameters range from -0.07 to -0.10 and are nominally higher than those parameters reported elsewhere.
Next Steps and Acknowledgments

- True test: compare predicted to actual forecasts at a given work zone

- Account for tolls that are being added to existing crossings.

- Insights: Mike Fontaine, Stephany Hanshaw, Tim Haynam, Rose Lawhorne, Amy O’Leary, and Eric Stringfield.

- Authors are responsible for errors!