1-24-2014

Modeling Injury Outcomes of Crashes involving Heavy Vehicles on Texas Highways

Salvador Hernandez  
Oregon State University

Let us know how access to this document benefits you.

Follow this and additional works at: http://pdxscholar.library.pdx.edu/trec_seminar

Part of the Transportation Commons, and the Urban Studies and Planning Commons

Recommended Citation

http://pdxscholar.library.pdx.edu/trec_seminar/63

This Book is brought to you for free and open access. It has been accepted for inclusion in TREC Friday Seminar Series by an authorized administrator of PDXScholar. For more information, please contact pdxscholar@pdx.edu.
Background Intro

- As the national economy continues to recover, the volume of large-truck (GVWR>10,000 lbs.) will also experience a similar recovery.
- The increase in truck traffic poses an increased hazard to passenger vehicle traffic.
- Large trucks were responsible for more fatalities than passenger vehicles in the U.S. (FHWA, 2010; NHTSA, 2009).
Some Statistics

- Large trucks accounted for roughly
  - 4% of registered vehicles and
  - 8% of VMT in 2008,
  - 11% of motor vehicle involved crash deaths in 2008. (FHWA, 2010)

- Crash costs related to large truck collisions

<table>
<thead>
<tr>
<th>Injury Categories</th>
<th>Cost (2005 USD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average cost per fatal crash</td>
<td>US$3,604,518</td>
</tr>
<tr>
<td>Average cost per non-fatal crash</td>
<td>US$195,258</td>
</tr>
<tr>
<td>Average cost per non-injury (PDO)</td>
<td>US$15,114</td>
</tr>
</tbody>
</table>

Source: Zaloshnja and Miller, 2006
Societal Impacts

• Associated costs are remarkably high
  • expenses related to loss of lives,
  • medical attention, and insurance, and
  • short term and long term physical and emotional effects
• Economic impacts
Motivation

• State of Texas faced the greatest economic losses of any state at $22.9 billion dollars in 2008
  • US from 2006 to 2010 → $107.4 billion dollars
• Recent data indicated that in 2010 the State of Texas experienced:
  • 3,023 deaths
  • 59,660 serious injury crashes
  • 82,685 people sustaining serious injuries
• Very few studies have investigated injury severities associated with large truck involved crashes, especially utilizing state specific crash databases
Research Objectives – Purely Exploratory

• The primary objective of this study is:
  • to analyze injury severity and understand and identify the factors of injury severity of large truck involved crashes

• A secondary objective is:
  • to account for any “unobserved heterogeneity” (i.e., unobserved factors that may influence an injury outcome) related to human, vehicle, and road-environment

WHY? Thorough understanding of the factors that affect the likelihood of a crash occurring or those which have occurred allows for effective countermeasure development
Research Area
Research Steps

1. Data Collection
2. Define Variables
3. Process/Clean Data
4. Injury Severity Modeling and Estimation
5. Variables and Scenario Explanation
6. Safety/Policy Implications
Data – Texas Crash Data

• The Crash Record Information System (CRIS) crash database of Texas was used to estimate injury severity models, where the KABCO severity scale was used to help define the injury outcomes as follows:

• fatal, incapacitating, non-incapacitating, possible, and no injury (Property-Damage-Only),

<table>
<thead>
<tr>
<th>Injury Levels</th>
<th>Police Reporting Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fatal</td>
<td>K</td>
</tr>
<tr>
<td>Incapacitating</td>
<td>A</td>
</tr>
<tr>
<td>Non-incapacitating</td>
<td>B</td>
</tr>
<tr>
<td>Possible</td>
<td>C</td>
</tr>
<tr>
<td>No injury (PDO)</td>
<td>O</td>
</tr>
</tbody>
</table>

January 27, 2014
Data Processing – Data Fusion

2006 to 2010

- Interstate Highways
- Large Trucks

- Vehicle and Person IDs

- Crash IDs

- Maximum Severity

SAS Data Processing/ Cleaning

Vehicle

Person

Crash

Sample for Models

Figure: Data Processing Steps in SAS
Methodological Approach

• Have detailed accident information, however the State of Texas is extremely large and diverse that:
  • Unobserved heterogeneity is likely to exist among the population of large-truck crash-involved road users
    • Due to differences in risk-taking behavior, physiological factors, driver training (especially for international drivers form neighboring Mexico), etc...

• Mixed logit (random parameters model) may be appropriate
Methodological Approach (Cont’d)

• Standard **Multinomial Logit (MNL)** estimation assumes that the estimated parameters are the same (fixed) for all observations.

• Violations of this assumption can occur if there is some compelling reason (to believe) that parameters vary across observations in a way that cannot be accounted for in the model

• If not accounted for, may result in potential bias and erroneous statistical inferences
Define:
\[ S_{in} = \beta_i X_{in} + \varepsilon_{in} \]

where

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>( S_{in} )</td>
<td>is a severity function determining the injury-severity category ( i ) for large truck involved crashes ( n );</td>
</tr>
<tr>
<td>( X_{in} )</td>
<td>is a vector of explanatory variables (e.g., driver, vehicle, road, and environment variables);</td>
</tr>
<tr>
<td>( \beta_i )</td>
<td>is a vector of estimable parameters; and</td>
</tr>
<tr>
<td>( \varepsilon_{in} )</td>
<td>is error term.</td>
</tr>
</tbody>
</table>
Methodological Approach (Cont’d)

If $\varepsilon_{in}$’s are assumed to be generalized extreme value distributed,

$$P_n(i) = \frac{\text{EXP}[\mathbf{\beta}_i X_{in}]}{\sum_l \text{EXP}[\mathbf{\beta}_l X_{ln}]}$$

where

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P_n(i)$</td>
<td>is probability of large truck involved incident $n$ having severity outcome with $I$ denoting all possible injury severity outcomes as hereunto presented</td>
</tr>
</tbody>
</table>
The Mixed Logit is:

\[ P_{in} = \int \frac{\text{EXP}[\beta_i X_{in}]}{\sum_l \text{EXP}[\beta_l X_{in}]} f(\beta | \phi) \, d\beta \]

where

\( f(\beta | \phi) \) is the density function of \( \beta \) with \( \phi \) referring to a vector of parameters of the density function (mean and variance).

- With this, \( \beta \) can now account for segment-specific variations of the effect of \( X \) on large truck involved crash probabilities, with the density function \( f(\beta | \phi) \) used to determine \( \beta \).
Mixed Logit

• Relaxes possible (independence of irrelevant alternatives) IIA problems with a more general error-term structure.

• Can test a variety of distribution options for $\beta$.
  - Normal, log-normal, triangular, etc...

• Estimated with simulation based maximum likelihood.

• For more on the Mixed logit I refer you to:

Empirical Setting

• Seek to model the maximum injury level sustained by the drivers involved in a crash

• Injury severities:

<table>
<thead>
<tr>
<th>Injury Severity</th>
<th>Observations</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Property damage only (PDO)</td>
<td>18,223</td>
<td>88.96%</td>
</tr>
<tr>
<td>Possible Injury (POSS)</td>
<td>1,120</td>
<td>5.46%</td>
</tr>
<tr>
<td>Non-incapacitating injury (Non-INCAP)</td>
<td>601</td>
<td>2.93%</td>
</tr>
<tr>
<td>Incapacitating injury (INCAP)</td>
<td>463</td>
<td>2.26%</td>
</tr>
<tr>
<td>Fatality (FATAL)</td>
<td>78</td>
<td>0.38%</td>
</tr>
</tbody>
</table>

• Total observations (N=20,495)
### Selected Descriptive Statistics

#### Meaning of Variables in the Model

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Severity outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Terrain of roadway (1 if level roadway surface, 0 otherwise)</td>
<td>0.764</td>
<td>0.425</td>
<td>Fatal</td>
</tr>
<tr>
<td>Time of day (1 if between 3 pm to 7 pm, 0 otherwise)</td>
<td>0.203</td>
<td>0.402</td>
<td></td>
</tr>
<tr>
<td>Weather condition (1 if clear weather, 0 otherwise)</td>
<td>0.682</td>
<td>0.466</td>
<td></td>
</tr>
<tr>
<td>Land-use pattern at crash location (1 if rural area, 0 otherwise)</td>
<td>0.268</td>
<td>0.443</td>
<td></td>
</tr>
<tr>
<td>Traffic flow at the time of crash (ADT per lane in each direction – veh/day/ln)</td>
<td>15397.3</td>
<td>8845.23</td>
<td></td>
</tr>
<tr>
<td>Month of the year (1 if summer (June to August), 0 otherwise)</td>
<td>0.257</td>
<td>0.437</td>
<td>Incapacitating Injury</td>
</tr>
<tr>
<td>Weather condition (1 if clear weather, 0 otherwise)</td>
<td>0.682</td>
<td>0.466</td>
<td></td>
</tr>
<tr>
<td>Shoulder width (right shoulder width (feet))</td>
<td>19.727</td>
<td>3.555</td>
<td></td>
</tr>
<tr>
<td>Time of the day (1 if between 3 pm to 7 pm, 0 otherwise)</td>
<td>0.203</td>
<td>0.402</td>
<td></td>
</tr>
<tr>
<td>Land-use pattern at crash location (1 if urban area, 0 otherwise)</td>
<td>0.564</td>
<td>0.496</td>
<td></td>
</tr>
<tr>
<td>Number of lanes on highways (1 if 4 lanes in one direction, otherwise)</td>
<td>0.428</td>
<td>0.495</td>
<td>Non-Incapacitating Injury</td>
</tr>
<tr>
<td>Time of the day (1 if between 12 am to 6 am, 0 otherwise)</td>
<td>0.159</td>
<td>0.366</td>
<td></td>
</tr>
<tr>
<td>Months of year (1 if fall (September to December), 0 otherwise)</td>
<td>0.318</td>
<td>0.428</td>
<td></td>
</tr>
<tr>
<td>Land-use pattern at crash location (1 if rural area, 0 otherwise)</td>
<td>0.268</td>
<td>0.443</td>
<td></td>
</tr>
<tr>
<td>Gender of the occupants (1 if male, 0 otherwise)</td>
<td>0.944</td>
<td>0.229</td>
<td>Possible Injury</td>
</tr>
<tr>
<td>Time of the day (1 if between 3 pm to 7 pm, 0 otherwise)</td>
<td>0.203</td>
<td>0.402</td>
<td></td>
</tr>
<tr>
<td>Traffic flow at the time of crash (ADT per lane in each direction – veh/day/ln) (1 if veh/day/ln &gt;2000, 0 otherwise)</td>
<td>0.985</td>
<td>0.122</td>
<td></td>
</tr>
<tr>
<td>Surface condition at the time of crash (1 if dry surface, 0 otherwise)</td>
<td>0.810</td>
<td>0.392</td>
<td></td>
</tr>
<tr>
<td>Age group (1 if age between 25 to 35, 0 otherwise)</td>
<td>0.178</td>
<td>0.382</td>
<td></td>
</tr>
<tr>
<td>Weather condition at the time of crash (1 if clear weather condition, 0 otherwise)</td>
<td>0.682</td>
<td>0.466</td>
<td>No-injury (PDO)</td>
</tr>
<tr>
<td>Light condition of street (1 if the surrounding area is dark but outside is lighted, 0 otherwise)</td>
<td>0.119</td>
<td>0.324</td>
<td></td>
</tr>
<tr>
<td>Land-use pattern at crash location (1 if urban area, 0 otherwise)</td>
<td>0.564</td>
<td>0.496</td>
<td></td>
</tr>
<tr>
<td>Age group (1 if age between 45 to 55, 0 otherwise)</td>
<td>0.239</td>
<td>0.427</td>
<td></td>
</tr>
<tr>
<td>Surface condition at the time of crash (1 if dry surface, 0 otherwise)</td>
<td>0.810</td>
<td>0.392</td>
<td></td>
</tr>
<tr>
<td>Traffic flow at the time of crash (ADT per lane in each direction – veh/day/ln)</td>
<td>15397.3</td>
<td>8845.23</td>
<td></td>
</tr>
</tbody>
</table>
Five Utility functions used in the Mixed Logit Approach

\[ S_{FATAL} = \sum \beta_i X_{in} + \varepsilon_{in} \]
\[ S_{INCAP} = \sum \beta_i X_{in} + \varepsilon_{in} \]
\[ S_{NON-INCAP} = \sum \beta_i X_{in} + \varepsilon_{in} \]
\[ S_{POSS} = \sum \beta_i X_{in} + \varepsilon_{in} \]
\[ S_{PDO} = \sum \beta_i X_{in} + \varepsilon_{in} \]
Human

Vehicle

Safety

Roadway

Environment
### Texas Case Study Findings

**Drivers Characteristics**
- Male [POSS]  
- 25-35 yrs [POSS]  
- 45-55 yrs [PDO]  

**Temporal Characteristics**
- 3-7 PM [INCAP, POSS]  
- 12-6 AM [NON-INCAP]  
- Jun-Aug [INCAP]  
- Sept-Dec [NON-INCAP]  

**Land use Characteristics**
- Rural [F, NON-INCAP]  
- Urban [INCAP, PDO]  

**Traffic Characteristics**
- ADT/LN/DAY >2K [POSS]  
- ADT/LN/DAY [F] [PDO]  

**Lighting Characteristics**
- Dark but lighted [PDO]  

**Weather Characteristics**
- Clear [F, PDO] [INCAP]  

**Road geometry Characteristics**
- 4-lane [NON-INCAP]  
- Right shoulder width [INCAP]  
- Level surface [F]  
- Dry surface [POSS, PDO]  

---

Increase likelihood  
Decrease likelihood
Random Parameter Findings

• **POSS:** The male indicator
  • Parameter normally distributed mean=-3.936, S.D.=1.622
  • This suggests that about 0.7% of large truck involved crashes involving male drivers on average were more likely to result in possible injuries

• **PDO:** Age group indicator between 45 to 55 years of age
  • Parameter normally distributed mean=1.097, S.D.=1.866
  • This implies that about 72.2% of large truck involved crashes with this age group (45 to 55 years’ old drivers) on average were less likely to result in non-injury crashes
Research Hypothesis Testing

Null hypothesis – $H_0$: Statistically equivalent models FIX=RAN

Alternate hypothesis – $H_A$: RAN is superior

$$\chi^2 = -2 \left[ LL_{\text{FIX}} \left( \beta^{\text{FIX}} \right) - LL_{\text{RAN}} \left( \beta^{\text{RAN}} \right) \right]$$

The $\chi^2$ statistic for the likelihood ratio test with six degrees of freedom gave a value greater than the 99.99% ($\chi^2 = 62.944$) confidence interval, indicating that the mixed logit model (i.e., random parameter model) is statistically superior to the corresponding multinomial model (i.e., fixed parameter model).
Summary (Key Findings)

• Higher traffic flow (i.e., vehicles per day per lane) reduces the likelihood of fatalities, but increases that of PDO crashes.

• Crashes occurring in rural settings result in a higher likelihood of fatalities and non-incapacitating injuries. In contrast, crashes occurring in urban settings result in a lower likelihood of incapacitating and PDO crashes.
Summary (Key Findings)

• Time of day such as 3 pm to 7 pm results in lower likelihood of fatalities, incapacitating and possible injuries. Additionally, time of day such as 12 am to 6 am results in higher likelihood of non-incapacitating injuries.

• Summer is likely to increase likelihood of incapacitating injuries, whereas fall is likely to decrease likelihood of non-incapacitating injuries.
Future Work

• Rural vs. Urban
• Time of Day
• Transferability of Models, e.g., Texas to Oregon etc...
Thank You!

Sources:

