Auto Loss Costs: Dynamic Linear Models with Changepoints

October 2020
Industry-wide auto insurance losses can be difficult to model but are very important for companies to understand. A new dynamic linear model is developed with seasonality, regression on congestion, and a linear trend with a changepoint. The changepoint allows the slope and intercept to change, enabling modeling of structural shifts in the industry, regardless of why they occur (e.g., regulatory, economic, or social changes). The changepoint improves the model fit and will likely lead to improved predictions of future losses; urban congestion best describes the loss process; frequency has generally decreased; and severity has generally increased. Loss cost has mainly increased, but a large number decreased at the beginning of the time window (Q1 1999 through Q3 2019). This model is likely better able to model the uncertainty in the industry going forward.

Introduction

The personal auto insurance industry is constantly changing. There are small and steady changes over time (e.g., inflation affecting the cost of repairs or general progress in the safety of vehicles) and sharp spikes (e.g., disasters and legislation). That change has never been more apparent with the Covid-19 pandemic, civil unrest, and driverless cars all presently having a significant impact on personal auto insurance.

Personal auto insurance frequency (number of claims per covered car year) had been consistently falling for many years prior to the financial crisis (2008-2010). This is largely attributed to improvements in collision avoidance technology, increased safety awareness, and changes in enforcement. Since the financial crisis, countrywide frequency has remained relatively constant. Both collision (Coll) and property damage (PD) frequency seem to increase from 2010-2016 before falling to the present. Countrywide severity (total loss per claim) has increased exponentially rather consistently since 1999, though there might have been some slowing around the financial crisis especially in PD and Coll. Countrywide loss cost (total loss per covered car year) is a combination of frequency and severity. Bodily injury (BI), PD, and Coll all were pretty constant (with falling frequency and increasing severity cancelling each other out) until the financial crisis and have been significantly increasing since then. Personal injury protection (PIP) has increased pretty steadily and comprehensive (Comp) has stayed constant. Figure 1 plots all three metrics on a countrywide level.

In addition to the overall trends, there are obvious seasonal patterns for frequency, severity, and loss cost. A typical model choice for time series data such as this would be an AR, ARMA, or ARIMA model with a seasonal component. However, the shifts or changepoints in the data, as is apparent when looking at the plots of frequency, severity, and loss cost, suggest that a structural change should be incorporated directly into a model.
Figure 1: Countrywide Trends
The main modeling tool used in this paper is a Dynamic Linear Model (DLM) (West and Harrison, 2006; Prado and West, 2010). A dynamic linear model is a very powerful tool as it allows for a realistic separation of the underlying process from the observed noisy data. Another simple yet different approach to changepoint modeling in DLMs is used that essentially treats time as regression variable and allows the coefficient to change at some point in the process.

Practicing actuaries can use this model and the results in this paper to better understand when structural changes and historical trends exist in auto insurance. Also, directors of state operations can use these results to better forecast losses and claims in their states. Additionally, interested researchers can apply these models to their own datasets to understand the shifts and changes in their data.

Data

The loss data is gathered from the ISS Fast Track Plus database.¹ Quarterly loss amounts, claims, and earned car years for each state and coverage (BI, PD, Comp, Coll, PIP, and property protection (PPI)) were obtained from Q1 1999 through Q3 2019. Because of the date range, any impacts of the current public health crisis or social unrest are unable to be inferred. Frequency (claims/earned car years), severity (loss amount/claims), and loss cost (loss amount/earned car years) are calculated. For the remainder of this report, the focus is on those three metrics.

In addition to the loss data, congestion data from the Federal Highway Administration was gathered. Congestion is defined as vehicle miles traveled/total road miles. This data is adjusted to include only urban roads, only rural roads, and all roads. One shortcoming of this information is that it is only at the annual level, all quarters are assumed to be the same within the year. That is obviously incorrect, but the seasonal effect of congestion will largely be incorporated in the seasonal effect in the model.

Results

Fitting a single model to a wide range of data is a difficult task. This is a rather large and varied dataset with

- 229 state and coverage combinations
  - 52 states each, all 50 states + DC + countrywide (CW) for BI, PD, Comp, and Coll
  - 20 PIP states
  - 1 PPI state (MI)
- Three different metrics: frequency, severity, and loss cost
- Four different congestion possibilities: total, urban, rural, and none (except DC does not have rural areas, so no rural congestion values)
- 52 different possible changepoints and one model without a changepoint

for a total of \((229 \times 3 \times 4 - 15) \times 53 = 144,849\) possible models to compare.

To parse through these results, the overall trends in the results are presented first and then are solidified with a few examples. For those combinations where the model did poorly, potential next steps are outlined to overcome the issues. Results from all the combinations are available in the appendix.

Energy scores (Gneiting and Raftery, 2007) are used to compare each model fit on the individual time series. Energy scores are similar to AIC or BIC in that they penalize for model complexity, so a more complicated model (like adding a changepoint) must fit significantly better to have a lower (better) energy score. The model that has the lowest energy score is denoted as the optimal model. In the time series plots, not only is the optimal changepoint designated, but also the second and third optimal changepoints. This helps give a broader understanding of possible periods where the time series has a structural shift.

¹ For more information please see https://www.iss-statistical.net/ or http://viewer.zmags.com/publication/9427a07c#9427a07c/1
Overall Trends

The discussion of the overall trends is divided into several subsections. First, the changepoint results are discussed, followed by the slopes of the linear trends, and finally the congestion measures.

Changepoints. Almost all the state/coverage/metric combinations chose a model with a changepoint as the optimal model. The only two which did not include a changepoint are Ohio comp severity and Utah comp loss cost. Plots of both those values are included in Figure 2. With Ohio comp severity the pattern is rather consistent throughout the period with an outlier in 2007 Q1. The magnitude of the seasonal trend does appear to increase after the outlier. For simplicity, the overall slope of the model can only change before and after the changepoint. If the seasonality was included in the changepoint, the model likely would have found a changepoint around 2007 Q1. For Utah comp loss cost the trend is interesting. There appears to be no steadily increasing trend, but rather three constant mean periods. The first is between 1999 and about 2004 with an annual mean around 70. Then, from 2004 until 2015, the mean drops to around 55. After 2015, the mean bounces back up to around 63. The model likely would have caught this pattern if it allowed for two separate changepoints. That will be a common theme as the models are discussed further. Another changepoint or two would likely change the results dramatically. Note that the changepoints in the plots are the best locations for changepoints, but no changepoint is preferred.

![Figure 2: Ohio Comprehensive Severity and Utah Comprehensive Loss Cost](image)

It would make sense that many of the changepoints would be chosen around the financial crisis. Table 1 shows that only slightly more optimal changepoints occurred during the financial crisis. The total proportion of models with changepoints selected in the financial crisis was only slightly more than would have been selected through random chance. Originally this result was disappointing. Upon closer visual investigation, however, it seemed that there were many instances of structural shifts in periods other than the financial crisis. For time series that shifted around the financial crisis, like countrywide collision loss cost or California PD frequency, the model chose those time periods to include a changepoint (Figure 3). However, most of the datasets did not include a significant change around the financial crisis or had a more significant changepoint elsewhere. This result does not suggest a faulty model, but rather suggests that major structural shifts over the 20-year period of study were not limited to the financial crisis.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Best CP</td>
<td>0.00</td>
<td>0.13</td>
<td>0.18</td>
<td>0.24</td>
<td>0.24</td>
<td>0.20</td>
</tr>
<tr>
<td>2nd Best CP</td>
<td>0.00</td>
<td>0.15</td>
<td>0.19</td>
<td>0.24</td>
<td>0.23</td>
<td>0.19</td>
</tr>
<tr>
<td>3rd Best CP</td>
<td>0.01</td>
<td>0.11</td>
<td>0.20</td>
<td>0.26</td>
<td>0.25</td>
<td>0.17</td>
</tr>
</tbody>
</table>

Table 1: Optimal Changepoints
Slopes. The chosen slopes of the model are described by dividing them into four categories (Table 2).

<table>
<thead>
<tr>
<th>Name</th>
<th>Slope before changepoint</th>
<th>Slope after changepoint</th>
</tr>
</thead>
<tbody>
<tr>
<td>Up</td>
<td>Positive</td>
<td>Positive</td>
</tr>
<tr>
<td>Down</td>
<td>Negative</td>
<td>Negative</td>
</tr>
<tr>
<td>Mountain</td>
<td>Positive</td>
<td>Negative</td>
</tr>
<tr>
<td>Valley</td>
<td>Negative</td>
<td>Positive</td>
</tr>
</tbody>
</table>

Table 2: Slope name definitions

As examples of each type of slope, Figure 4 shows Alaska PD severity (Up), Minnesota BI Frequency (Down), Delaware PIP Severity (Mountain), and Tennessee Collision Loss Cost (Valley). Notice that the classification of the slope could change if the second- or third-best changepoint were chosen instead of the best.

The slopes were largely driven by the metric. Most of the frequency models are Down and most of the severity models are Up. The combination of the two (Loss Cost) is mainly Up with a large number showing a Valley (Tables 3-5).

<table>
<thead>
<tr>
<th></th>
<th>Down</th>
<th>Valley</th>
<th>Mountain</th>
<th>Up</th>
</tr>
</thead>
<tbody>
<tr>
<td>BI</td>
<td>0.56</td>
<td>0.23</td>
<td>0.17</td>
<td>0.04</td>
</tr>
<tr>
<td>Coll</td>
<td>0.42</td>
<td>0.50</td>
<td>0.08</td>
<td>0.00</td>
</tr>
<tr>
<td>Comp</td>
<td>0.42</td>
<td>0.35</td>
<td>0.21</td>
<td>0.02</td>
</tr>
<tr>
<td>PD</td>
<td>0.81</td>
<td>0.15</td>
<td>0.04</td>
<td>0.00</td>
</tr>
<tr>
<td>PIP</td>
<td>0.45</td>
<td>0.30</td>
<td>0.25</td>
<td>0.00</td>
</tr>
<tr>
<td>PPI</td>
<td>1.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Table 3: Frequency Slopes

<table>
<thead>
<tr>
<th></th>
<th>Down</th>
<th>Valley</th>
<th>Mountain</th>
<th>Up</th>
</tr>
</thead>
<tbody>
<tr>
<td>BI</td>
<td>0.00</td>
<td>0.06</td>
<td>0.06</td>
<td>0.88</td>
</tr>
<tr>
<td>Coll</td>
<td>0.00</td>
<td>0.04</td>
<td>0.02</td>
<td>0.94</td>
</tr>
<tr>
<td>Comp</td>
<td>0.00</td>
<td>0.06</td>
<td>0.13</td>
<td>0.81</td>
</tr>
<tr>
<td>PD</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>PIP</td>
<td>0.00</td>
<td>0.10</td>
<td>0.25</td>
<td>0.65</td>
</tr>
<tr>
<td>PPI</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Table 4: Severity Slopes

<table>
<thead>
<tr>
<th></th>
<th>Down</th>
<th>Valley</th>
<th>Mountain</th>
<th>Up</th>
</tr>
</thead>
<tbody>
<tr>
<td>BI</td>
<td>0.02</td>
<td>0.33</td>
<td>0.23</td>
<td>0.42</td>
</tr>
<tr>
<td>Coll</td>
<td>0.00</td>
<td>0.56</td>
<td>0.06</td>
<td>0.38</td>
</tr>
<tr>
<td>Comp</td>
<td>0.06</td>
<td>0.29</td>
<td>0.21</td>
<td>0.02</td>
</tr>
<tr>
<td>PD</td>
<td>0.81</td>
<td>0.15</td>
<td>0.04</td>
<td>0.00</td>
</tr>
<tr>
<td>PIP</td>
<td>0.45</td>
<td>0.30</td>
<td>0.25</td>
<td>0.00</td>
</tr>
<tr>
<td>PPI</td>
<td>1.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Table 5: Loss Cost Slopes
Congestion. In previous studies, congestion consistently rated as one of the most important variables when trying to predict statewide losses. This exercise proved to be no different. The best model had no congestion variable in 10% of the combinations. Not counting PPI, urban congestion was consistently the most often chosen. The proportion of models by coverage type that had each congestion as most significant is shown in Table 6.

<table>
<thead>
<tr>
<th></th>
<th>None</th>
<th>Rural Congestion</th>
<th>Urban Congestion</th>
<th>Total Congestion</th>
</tr>
</thead>
<tbody>
<tr>
<td>BI</td>
<td>0.10</td>
<td>0.24</td>
<td>0.51</td>
<td>0.15</td>
</tr>
<tr>
<td>Coll</td>
<td>0.06</td>
<td>0.35</td>
<td>0.37</td>
<td>0.22</td>
</tr>
<tr>
<td>Comp</td>
<td>0.13</td>
<td>0.23</td>
<td>0.26</td>
<td>0.38</td>
</tr>
<tr>
<td>PD</td>
<td>0.07</td>
<td>0.33</td>
<td>0.33</td>
<td>0.28</td>
</tr>
<tr>
<td>PIP</td>
<td>0.15</td>
<td>0.28</td>
<td>0.32</td>
<td>0.25</td>
</tr>
<tr>
<td>PPI</td>
<td>0.33</td>
<td>0.67</td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Table 6: Congestion type in the best model

Conclusion

In this paper a new model is developed to better understand the complicated dynamics in personal auto insurance. A dynamic linear model with seasonality, regression on congestion, and a linear trend with a changepoint is designed. The changepoint enables modeling of structural shifts in the industry, regardless of why they occur (e.g., regulatory, economic, or social changes).

The changepoint improves the model fit and will likely lead to improved predictions of future losses; urban congestion best describes the loss process; frequency has generally decreased; and severity has generally increased. Loss cost has mainly increased, but a large number decreased at the beginning of the time window.
For future work, it will be interesting to see how the model deals with the Covid-19 pandemic. Another changepoint or two would make the model more flexible. Finally, more covariates and their impacts on losses would be interesting to study.

Acknowledgments
The authors are grateful to the Casualty Actuarial Society, American Property Casualty Insurance Association, and the Society of Actuaries for supporting this project. They are especially grateful for the thoughtful feedback and support of the members of the project oversight group (Joan Barrett, Kevin Brazee, Dave Clark, Dave Core, David DeNicola, Peter Drogan, Brian Fannin, Russell Fox, Rick Gorvett, Dale Hall, Chris Harris, Linda Jacob, Ben Kimmons, Tyler Lantman, Scott Lennox, James Lynch, Kim MacDonald, Lawrence Marcus, Rob Montgomery, Thomas Myers, Norman Niami, Bob Passmore, Dave Prario, Jacob Robertson, Michelle Rockafellow, Erika Schulty, Janet Wesner, and Ken Williams).

References


Appendix: Results for all states and coverages
First Change Point  Second Change Point  Third Change Point

AR Comp Frequency

Year

AR Comp Severity

Year

AR Comp Loss Cost

Year
**MS Comp Frequency**

- First Change Point
- Second Change Point
- Third Change Point

**Year**

- 2000
- 2005
- 2010
- 2015
- 2020

**Frequency**

- 0.05
- 0.10
- 0.15
- 0.20

**MS Comp Severity**

- First Change Point
- Second Change Point
- Third Change Point

**Year**

- 2000
- 2005
- 2010
- 2015
- 2020

**Severity**

- 1000
- 2000
- 3000
- 4000

**MS Comp Loss Cost**

- First Change Point
- Second Change Point
- Third Change Point

**Year**

- 2000
- 2005
- 2010
- 2015
- 2020

**Loss Cost**

- 0
- 250
- 500
- 750
- 1000
First Change Point
Second Change Point
Third Change Point

MS PD Frequency
0.026 0.028 0.030 0.032

Year
Frequency

Change Point
ES

RuralCong
TotalCong
UrbanCong

MS PD Severity
2500 3000 3500 4000

Year
Severity

Change Point
ES

RuralCong
TotalCong
UrbanCong

MS PD Loss Cost
60 80 100

Year
Loss Cost

Change Point
ES

RuralCong
TotalCong
UrbanCong
### MT Coll Frequency

<table>
<thead>
<tr>
<th>Year</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000</td>
<td>0.035</td>
</tr>
<tr>
<td>2005</td>
<td>0.040</td>
</tr>
<tr>
<td>2010</td>
<td>0.045</td>
</tr>
<tr>
<td>2015</td>
<td>0.050</td>
</tr>
<tr>
<td>2020</td>
<td>0.055</td>
</tr>
</tbody>
</table>

### MT Coll Severity

<table>
<thead>
<tr>
<th>Year</th>
<th>Severity</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000</td>
<td>26</td>
</tr>
<tr>
<td>2005</td>
<td>27</td>
</tr>
<tr>
<td>2010</td>
<td>28</td>
</tr>
<tr>
<td>2015</td>
<td>30</td>
</tr>
<tr>
<td>2020</td>
<td>32</td>
</tr>
</tbody>
</table>

### MT Coll Loss Cost

<table>
<thead>
<tr>
<th>Year</th>
<th>Loss Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000</td>
<td>100</td>
</tr>
<tr>
<td>2005</td>
<td>125</td>
</tr>
<tr>
<td>2010</td>
<td>150</td>
</tr>
<tr>
<td>2015</td>
<td>175</td>
</tr>
<tr>
<td>2020</td>
<td>200</td>
</tr>
</tbody>
</table>
ND PD Frequency

First Change Point  Second Change Point  Third Change Point

Year

ND PD Frequency

ES

None  RuralCong  TotalCong  UrbanCong

Change Point

ND PD Severity

First Change Point  Second Change Point  Third Change Point

Year

ND PD Severity

ES

None  RuralCong  TotalCong  UrbanCong

Change Point

ND PD Loss Cost

First Change Point  Second Change Point  Third Change Point

Year

ND PD Loss Cost

ES

None  RuralCong  TotalCong  UrbanCong

Change Point
First Change Point
Second Change Point
Third Change Point

NE PD Frequency

Year
Frequency

NE PD Severity

Year
Severity

NE PD Loss Cost

Year
Loss Cost

Change Point
ES
None
RuralCong
TotalCong
UrbanCong
First Change Point
Second Change Point
Third Change Point

UT BI Frequency

Year

Frequency


0.007 0.008 0.009 0.010

Change Point

ES

None RuralCong TotalCong UrbanCong


6.5 7.0 7.5 8.0

UT BI Severity

Year

Severity


1000 0 0 0 0

Change Point

ES

None RuralCong TotalCong UrbanCong


165 170 175 180

UT BI Loss Cost

Year

Loss Cost


80 100 120 140

Change Point

ES

None RuralCong TotalCong UrbanCong


160 165 170 175 180
First Change Point
Second Change Point
Third Change Point

WY Comp Frequency

Year

Frequency

0.05 0.10 0.15 0.20

Year

Frequency

190 200 210

Year

Frequency

1000 2000 3000

Year

Severity

10.0 10.5 11.0 11.5

Year

Severity

58 60 62 64

Year

Loss Cost

250 500 750

Year

Loss Cost

58 60 62 64

Change Point

ES

None RuralCong TotalCong UrbanCong

Change Point

ES

None RuralCong TotalCong UrbanCong

Change Point

ES

None RuralCong TotalCong UrbanCong

Change Point

ES

None RuralCong TotalCong UrbanCong

Change Point

ES

None RuralCong TotalCong UrbanCong