Analyzing the Effect of Autonomous Vehicles on Auto Insurance Premiums

Completed for the Certificate in Scientific Computation

Fall 2017

Ramzi Saud

B.A. Economics

Economics Department

College of Liberal Arts

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Thomas Struppeck  
Lecturer

Mathematics

**ABSTRACT** The growth of the autonomous vehicle industry will have a staggering effect on the auto insurance industry. Insurance premiums are a transfer of risk between the insured and the insurer. Because fully implemented autonomous vehicles (AV’s) are safer than human drivers the probability of the insured filing a claim and the size of the claim decrease. With this decline in frequency and severity, premiums will go down as well.

This paper predicts premiums for differing levels of autonomy based on the classification system published by SAE International. Premiums are categorized by gender, vehicle type, and 6 age groupings. Pure premiums from 2009 are calculated by combining frequency of accidents and severity of accidents, where severity is estimated comprehensive costs. A Gamma log-link GLM regression of the insured’s characteristics on pure premium predicts future pure premiums for 4 safety features. Actual premiums are then calculated by using the ratio between total revenue and profits from Progressive’s 2009 Tax Report. Crash data is obtained from the 2009 National Automotive Sampling System General Estimates System (GES) collected by the National Highway Traffic Safety Administration. Vehicle miles traveled (VMT) by age, gender, and vehicle type is obtained from the 2009 Nationwide Household Transportation Survey (NHTS). Population data was estimated using the ‘survey’ package in R. Present premiums are calculated using the Consumer Price Index published by the Bureau of Labor Statistics.

**Key Terms:**

Frequency: Claims/Exposures

Severity: Losses/Claims

Pure Premium: Severity \* Frequency

Actual (Commercial) Premium: Pure Premium \* ()

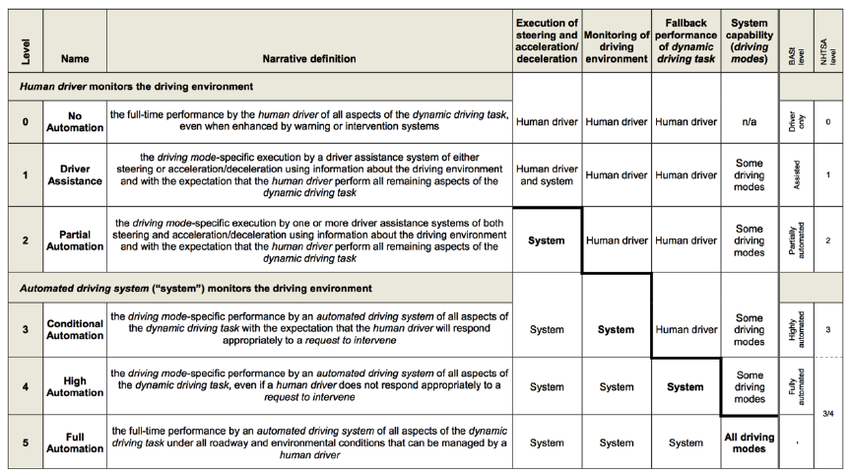
Lane Departure Warning: Warns driver when vehicle begins to move out of lane

Forward Collision Warning: Warns driver of impending forward collision

Autonomous Emergency Braking: Vehicle automatically breaks to avoid or mitigate a crash

Blind Spot Detection: Warns a driver when a car enters blind spot of driver

* Levels of Automation:

**Figure 1: Definitions of automation levels**

**Introduction**

The auto insurance industry is facing the greatest upheaval since its inception. Manufacturers like Volvo, Tesla, and Ford anticipate levels 3 and 4 vehicles to be commercially available and on roads by 2020.[[1]](#footnote-1) KPMG estimates the auto insurance industry will be reduced by 70% by 2050.[[2]](#footnote-2) Insurers need to be ready to immediately adapt to the changes in the industry or risk being pushed out of the market by competitors offering appropriately cheaper rates. For level 5 AV’s it is likely that liability will shift from the driver to the manufacturer, but between now and then insurers must adjust their rates to account for differing levels of autonomy. These levels will be between 2, 3, and 4 where liability is not completely clear. Insurance claims data and crash data including information on safety features of vehicles are not publicly available, and exposures for vehicles which currently have high levels of autonomy are low.

The predicted results of this research are that the actual premiums for all of those insured will go down regardless of age, gender, or vehicle type. It is expected that the significance of those characteristics in explaining the variation in the actual premiums will go down as well. Premiums for those vehicles with safety features which reduce crash frequencies by a larger proportion than other safety features should accordingly be lower.

To account for the lack of exposure data, VMT in millions is used as the exposure for frequency, and estimated crash counts from the GES are used for number of claims. An added benefit utilizing VMT is the ability to effectively analyze the differences in exposure for vehicle types. The VIN’s provided by GES only contain 12 digits instead of the full 17 in order to ensure privacy of those included in the survey, the last 5 digits of the VIN provide essential information on the vehicles which would allow identification of safety features in the vehicles. Without proper identification of vehicles with safety features there is no way to estimate the reduction in crash frequency using public data. The Highway Loss Data Institute (HLDI) collects insurance data from many insurers, allowing their researchers to estimate effects of safety features on crash frequencies. This paper uses reduced crash rates from Jessica Cicchino’s research on lane departure warnings, blind spot detection, and front crash warnings with and without autonomous braking.[[3]](#footnote-3)[[4]](#footnote-4)[[5]](#footnote-5)

This work gathers that there is a significant reduction in premium rates across all driver characteristics. Safety features that reduce crash rates at a greater frequency are estimated to reduce premiums by a greater effect than other features. It would be worthwhile to investigate the comprehensive costs of accidents when these safety features are present in order to more accurately predict premiums. The reduction in significance of the driver characteristics alone is found to be negligible. It is possible that they would be less significant for vehicles where multiple safety features are present at once.

**METHODS**

**Data Sets** The 2009 National Household Travel Survey (NHTS)[[6]](#footnote-6) and General Estimates Survey (GES)[[7]](#footnote-7) data sets were used for this research. The 2009 NHTS data is the most recent data currently available. GES data is collected annually, but the 2009 set is used to be consistent with the NHTS data.

The NHTS dataset provides comprehensive data on travel patterns in the United States. contains data from 150,147 households, 308,901 individuals, and 309,163 vehicles. There are 4 separate files in the NHTS data, with focuses on person, household, vehicle, and travel day. The data includes weights on households, persons, travel day, and travel period. In order to conduct analysis on national VMT, travel day weights were used.

The GES dataset is a nationally representative probability sample from police accident reports in the United States. GES includes data on fatal, injury, and property-damage crashes. The GES contains data on 5,505,000 unique crashes with 9,640,000 vehicles. Each crash has a weighting factor to be used to scale up crash counts to a national total.

**Processing Data**

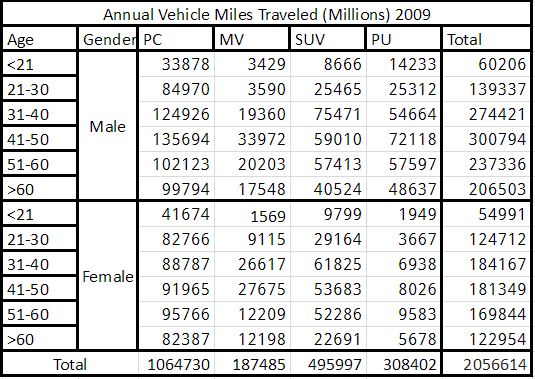
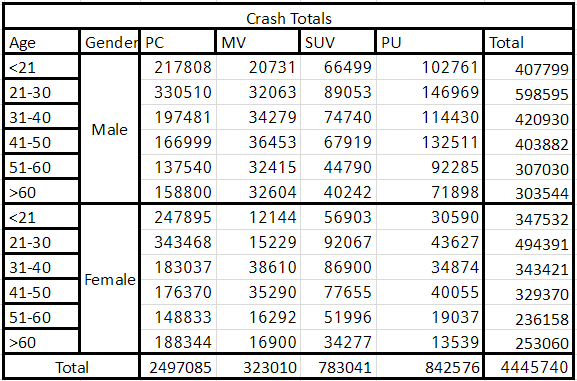
Both the GES and NHTS datasets have separate files, and some data is not relevant to this research. So some work was involved to merge and subset the data for analysis. The NHTS files were all merged by HOUSEID and TDCASEID. Survey data cannot be subsetted as regular data would be as the population estimates will be inaccurate. Using the ‘survey’ package in R, the subset() command is used to maintain the original design. The cohorts of interest are age, gender, and vehicle type. Also, only records on drivers were used, as this research is interested in driver exposure. The subsetted groups were male and female drivers of all ages driving passenger cars, minivans, SUV’s, or pickup trucks. Pickup trucks for this research are defined as pickup style trucks weighing less than 10,000lbs. From the GES data, the files person, vehicle, and crash were merged by CASENUM. When data was missing for age, gender, and vehicle type, imputed values were used, calculation of imputed values can be viewed in the 2009 GES and NHTS user manuals. The same subsetted groups were used for the GES as the NHTS. The code in figure 2 illustrates the creation of a survey design and appropriate subsetting practice.

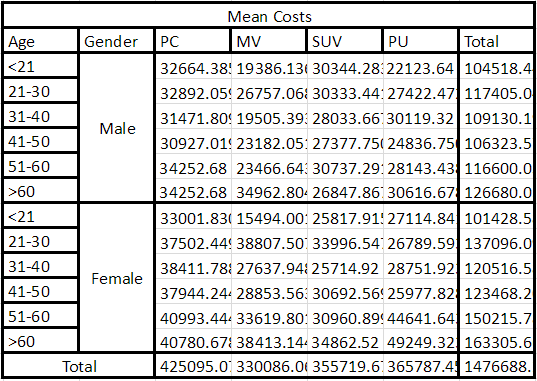
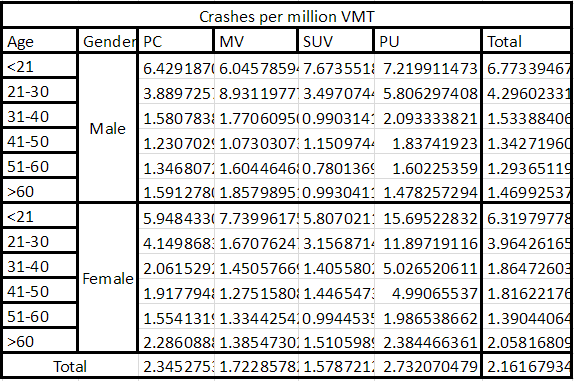
dailyVMT.w <- svydesign(ids=~1, data=dailyVMT, weights=dailyVMT$WTTRDFIN  
dailyVMT.young.male.passengercar <- subset(dailyVMT.w, ~I(SEX == 01 & AGE < 21 &   
+ VEHTYPE ==01 & DRVR\_FLG == 1))

(**Figure 2:** coding survey design with ‘survey’ package)

**Pure Premium Calculations**

The first step to calculating the pure premium is calculating the frequency of claims. For the purposes of this investigation each accident record is considered as an insurance claim. VMT, in millions, by groupings aggregated from the NHTS is used as the unit of exposure which can be viewed in table 1. Crash counts by groupings aggregated from the GES is used for crash counts, which can be viewed in table 2. Table 3 shows the frequency of crashes for each grouping which is calculated by dividing table 2 by table 1.

**Table 1 Table 2**

**Table 3 Table 4**

For severity, comprehensive costs estimated by Miller and Zaloshnja[[8]](#footnote-8) are matched to the GES data by injury severity on the KABCO classification scale used by police officers on their accident reports. These costs are then scaled to the national level to observe the distribution of mean costs by age, gender, and vehicle type in table 4. The costs are converted to 2016 dollars by multiplying by the inflation rate between 2016 and 2001.

Multiplying the frequency by the severity gives mean expected losses per crash, or pure premium. Each cell in table 4 is the amount a person with those characteristics is expected to lose in a crash.

**Pure Premium by Safety Features**

To calculate pure premiums for each safety feature, crash counts were reduced in correspondence with the estimated reduction in crashes. For forward collision warning (FCW) with and without autonomous emergency braking (AEB) all forms of crashes are reduced by 12% and 6% respectively. So to calculate pure premiums for cars with only FCW the crash totals would be reduced by 12% and then calculated in the same way as without the features. Lane departure warning (LDW) reduced sideswipe and head on crashes by 11%, and blind spot monitoring reduced lane-change crashes by 14%.

**Modelling Pure Premium**

Typically, auto insurance claims data has many extreme cases, and when total claims are the exposure unit there are many people insured with zero claims. In this case, the extreme cases create a heavy right skew on the pure premium histogram and a fat positive tail on a QQplot from a standard additive linear regression. The continuous response variable is premiums, the categorical explanatory variables are sex and body type, and the continuous explanatory variable is age. Sex and body type are called as factor variables for all models for simple interpretation of the coefficients.

In an effort to normalize the distribution logarithmic and square-root transformations on premiums are attempted. The square-root transformation does not have a desirable effect, while the logarithmic transformation seems to normalize the data with the possibility of it becoming bimodal. The linear model in figure 3 is attempted, while the significance and dispersion parameters are fine, the residuals violate the assumption of homoskedasticity. Including interaction effects between all variables does not rectify the violation.

svyglm(log(premium)~SEX + AGE + BDYTYP, design=master.w)  
(**Figure 3:** linear model with log transformation)

Instead of normalizing the distribution, generalized linear models are attempted. The histogram suggests that the premiums could be gamma distributed, so variations with the gamma family are attempted. Gamma family regressions with log, identity, and inverse link are all tested with and without interaction effects. To determine which link fits the best, the Preigbon link test is used residual deviance is compared. The t-value for the Pregibon test is insignificant for only the log link, indicating the correct link. Using the log-link, coefficients are interpreted through the formula in figure 4.



(**Figure 4:** interpretation of coefficients for gamma log-link)

When interaction effects are included, deviance is reduced, but it becomes more difficult to interpret the predictors. The trade-off in goodness-of-fit for interpretation of the model is not enough to choose to use interaction effects of predictors. It should be noted that the predictors driving a minivan, and minivan \* AGE are less significant, though still significant at the 95% confidence interval, in the model with interaction effects. All other predictors are significant at the 99% confidence interval for all models. When predicting premiums for vehicles with different safety features, there was no change in the significance of the predictors as was hypothesized.

**Adjusting for Actual Premium**

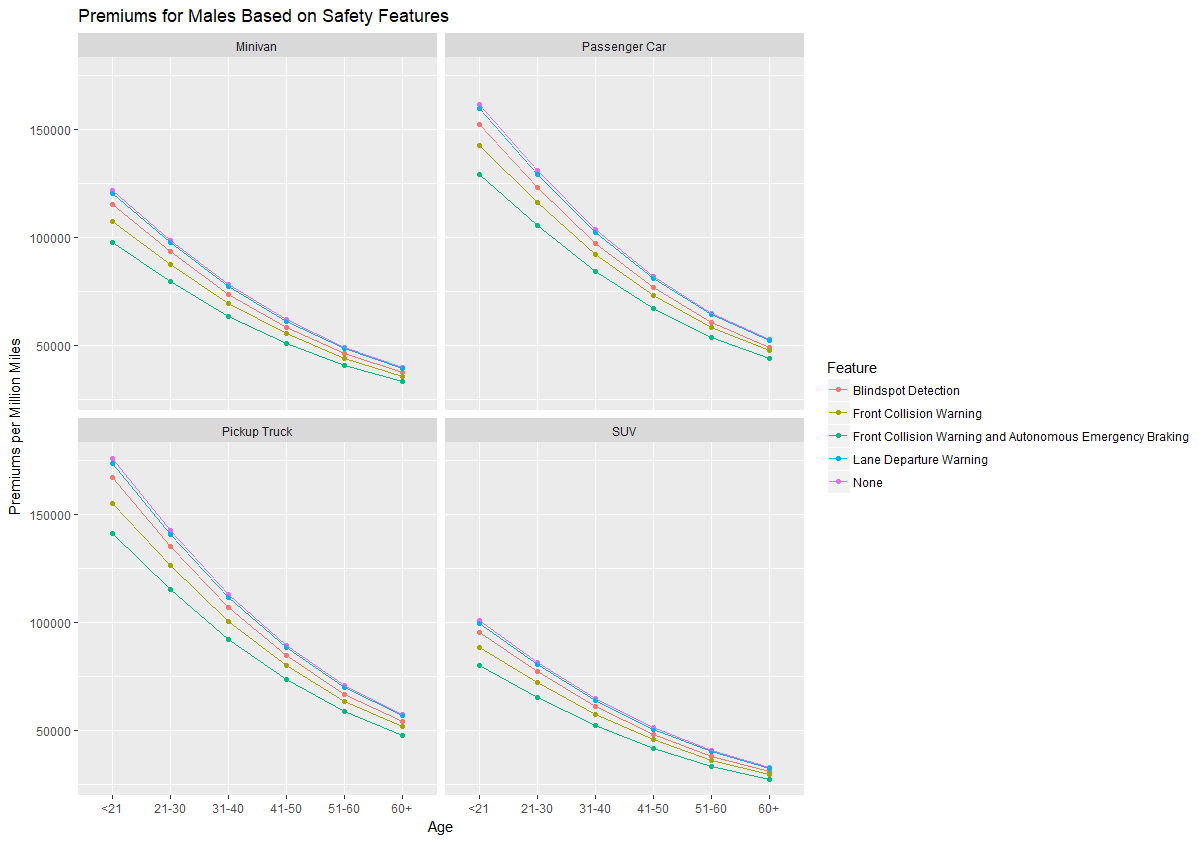
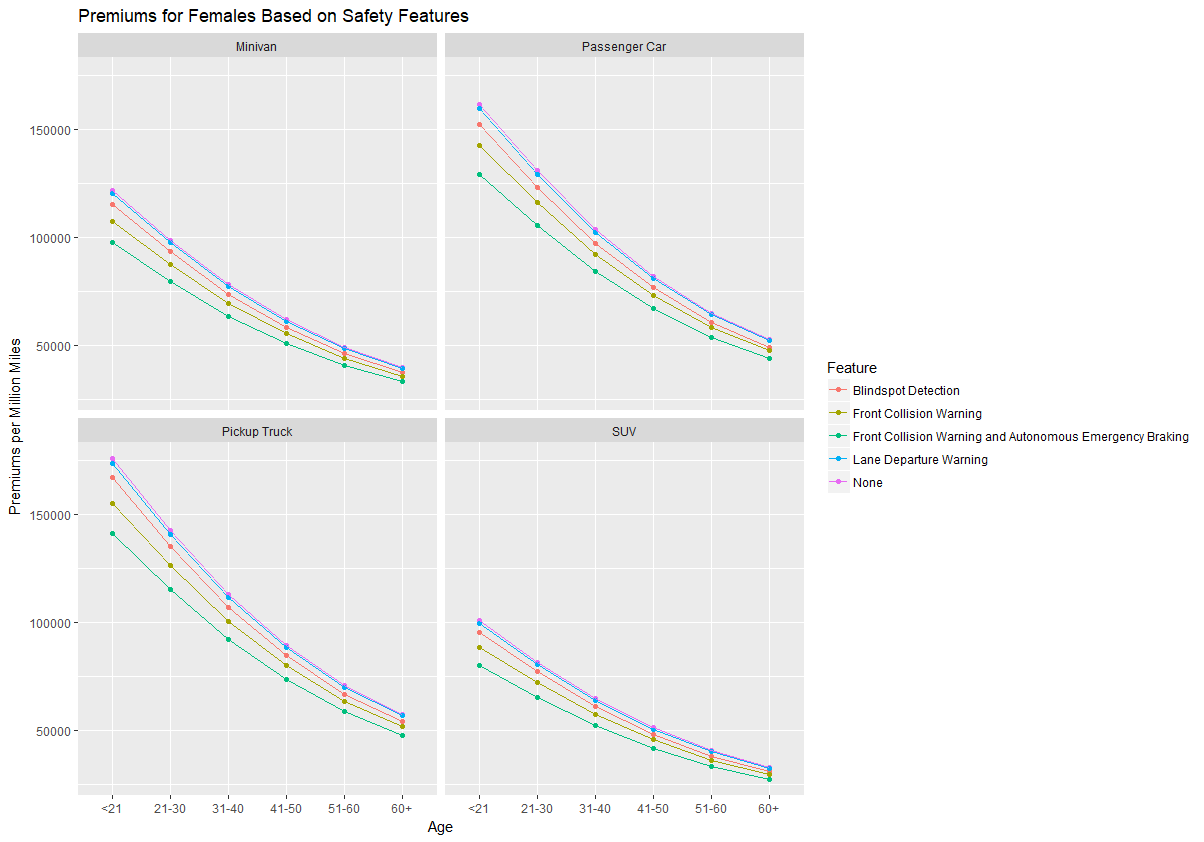
Pure premiums are equivalent to the expected loss cost from an accident. Insurers must charge on top of the expected loss cost in order to make a profit. To determine the profit to actual premium ratio, Progressive’s 2009 annual financial report is used. From commercial auto, Progressive’s total revenue was $1,623.3 million and the pretax profit is $229.8 million, which is a 14.16% profit margin. Scaling up the predicted pure premiums by 14.16% results in the actual predicted premium.

Actual premiums can be converted to become more meaningful to consumers using a per-mile premium system. The pure premium is divided by one million and multiplied by average annual miles driven. This is not done until the end of the analysis because it makes the rates difficult to interpret when they are in the magnitude of millionths. This predicted premium has greater variance for risk-pools as it is a per-mile premium; benefits and drawbacks of per-mile premiums are discussed thoroughly by Edlin.[[9]](#footnote-9) Because this model is not widely used by insurers, the analysis focuses on percent changes in premiums without using the per-mile system.

**RESULTS AND DISCUSSION**

As evidenced in table 1, males drive 45% more than females, and people younger than 21 drive less than other age groups. These findings are in accordance with cultural norms where males drive instead of females, and young people do not work much if at all so they drive less. Passenger cars are driven 7% more than all other types of cars combined. Due to such high exposures, production of safety features should be focused on preventing head on collisions and lane change accidents which are the most severe for passenger cars. Crash rates for people younger than 21 driving all types of cars are 58% higher than those for 21-30 year olds, and 410% higher than all other age groups. Because young people are more prone to crash for both genders and all vehicle types due to their inexperience, their premiums are reduced by factors less than those of older individuals. It seems likely that with full eyes-off automation, young people have the greatest potential for savings on premiums in the future as their risk profile will be reduced the most. Females younger than 21 who drive pick-up trucks have the highest crash rates at 15.70 crashes per million miles driven, and pick-up trucks in general have the lowest change in crash rates with safety features installed. This could be evidence of a trend in the decreased production of pick-up trucks as they would be riskier than other vehicles even with automation.

Predicted actual premiums with and without safety features are visualized in figured 5 and 6. Blind spot monitoring, FCW alone, FCW with AEB, and lane departure warning reduce premiums by an estimated 6%, 12.2%, 22.7%, and 1.2% respectively, on average. Premiums for all genders and vehicle types are reduced by a significant amount for all safety features except for lane departure warnings. For lane departure warning and blind spot monitoring, older individuals’ premiums are reduced at a greater rate than their younger cohorts. While for FCW alone and with AEB premiums are reduced at a greater rate for younger individuals. This could be due to younger people engaging in more risky driving behavior such as speeding or drunk driving[[10]](#footnote-10) which would increase both the frequency and severity of crashes.

**Figure 5**

**Figure 6**

FCW with AEB has the greatest effect on premiums, and is the feature with the highest level of autonomy. Combining AEB with FCW decreases premiums by an extra 10.5% which shows that when the vehicle’s system is able to react to its environment with a dynamic driving task then there is significantly less risk in insuring the vehicle. FCW and FCW with AEB reduce premiums by a greater magnitude than blind spot monitoring and lane departure warnings. Likely because the head on accidents prevented are typically more severe than other accidents. Further study into how many fatalities are prevented due to AEB should yield promising results. Rollover accidents have high costs, but none of the studied safety features have a significant effect on reducing rollovers.

Vehicles with AEB would fall under the classification of a level 4 vehicle, while FCW alone, lane departure warning, and blind spot monitoring would be features of a level 3 vehicle as they monitor the driving environment without performing any dynamic driving tasks. For level 3 vehicles the average predicted premium across all cohorts is $70,910 per million miles driven, and $955.6 under the per-mile premium system. For level 4 vehicles the predicted premium the predicted premium is $61,206.29 per million miles driven, and $824.82 under the per-mile premium system.   
  
**CONCLUSION**

Higher autonomy in vehicles decisively lowers the risk of the insured, and for every autonomous safety feature included in a vehicle the premium should be significantly reduced. Further research is required to determine if there is also a reduction in accident severity. The prediction that cars becoming more autonomous will reduce premiums is correct. Firms that do not lower their premium prices will be pushed out of the auto insurance industry by firms who do appropriately reduce their prices to account for the reduction in accidents across all categories. The predictors age, sex, and vehicle type did not have any noticeable reduction in significance or in explanation of the variance of costs from accidents.

**LIMITATIONS**

Several limitations were encountered through the course of this investigation. The most significant limitation is the lack of real insurance data. If there was data with insurance claims exposures, crash counts, and included safety features in vehicles the results would be more accurate and provide more insights. Without included safety features in each vehicle it is not valid to combine reductions in crash rates when adding a safety feature. For example, lane departure warnings and blind spot monitoring prevent similar accidents, so without complete data reductions cannot be determined for vehicles with both lane departure warnings and blind spot monitoring. It would be useful to have data on a wider variety of features which can perform dynamic driving tasks such as lane departure warning with lane assistance systems or steering assist. Gender, age, and vehicle type explain a significant portion of the variation in premiums, but driver history is an important predictor which is not possible to ascertain. Another limitation due to the lack of data is the modelling of survey data. There is no method to apply a Box-Cox transformation on survey estimates, which could potentially provide a more suitable model than the Gamma-log.

**ACKNOWLEDGEMENTS**

Appreciated is the educational guidance of Thomas Struppeck and the support of Vicki Keller and the University of Texas’s Statistics and Data Sciences department.

**BIBLIOGRAPHY**

Cicchino, Jessica. “Effectiveness of Forward Collision Warning Systems with and without Autonomous Emergency Braking in Reducing Police-Reported Crash Rates.” Highway Loss Data Institute. January 2016.

Cicchino, Jessica. “Effects of Lane Departure Warning on Police-Reported Crash Rates.” Highway Loss Data Institute. August 2017.

Cicchino, Jessica. “Effects of Blind Spot Monitoring Systems on Police-Reported Lane-Change Crashes.” Highway Loss Data Institute. August 2017.

Edlin, Aaron. *Per-Mile Premiums for Auto Insurance*. UC Berkeley Department of Economics. January 27, 1999.

ftp://ftp.nhtsa.dot.gov/GES/GES09/

http://nhts.ornl.gov/download.shtml

KPMG LLP. *The Chaotic Middle the Autonomous Vehicle and Disruption in Automobile Insurance*. June 2017

Miller, Ted., and Eduard Zaloshnja. *Comprehensive and Human Capital Crash Costs by Maximum Police-Reported Injury Severity Within Selected Crash Types*. Annu Proc Assoc Adv Automot Med. 2004; 48: 251–263.

Smart, Diana., and Suzanne Vassallo. “In the Driver’s Seat II: Beyond the Early Driving Years.” No. 17. April 2010.

Stoll, John. “GM Executive Credits Silicon Valley for Accelerating Development of Self-Driving Cars.” The Wall Street Journal. May 10, 2016.

1. Stoll, John. “GM Executive Credits Silicon Valley for Accelerating Development of Self-Driving Cars.” The Wall Street Journal. May 10, 2016. [↑](#footnote-ref-1)
2. KPMG LLP. *The Chaotic Middle the Autonomous Vehicle and Disruption in Automobile Insurance*. June 2017 [↑](#footnote-ref-2)
3. Cicchino, Jessica. “Effectiveness of Forward Collision Warning Systems with and without Autonomous Emergency Braking in Reducing Police-Reported Crash Rates.” January 2016. [↑](#footnote-ref-3)
4. Cicchino, Jessica. “Effects of Lane Departure Warning on Police-Reported Crash Rates.” August 2017. [↑](#footnote-ref-4)
5. Cicchino, Jessica. “Effects of Blind Spot Monitoring Systems on Police-Reported Lane-Change Crashes.” August 2017. [↑](#footnote-ref-5)
6. http://nhts.ornl.gov/download.shtml [↑](#footnote-ref-6)
7. ftp://ftp.nhtsa.dot.gov/GES/GES09/ [↑](#footnote-ref-7)
8. Miller, Ted., and Eduard Zaloshnja. *Comprehensive and Human Capital Crash Costs by Maximum Police-Reported Injury Severity Within Selected Crash Types*. Annu Proc Assoc Adv Automot Med. 2004; 48: 251–263. [↑](#footnote-ref-8)
9. Edlin, Aaron. *Per-Mile Premiums for Auto Insurance*. UC Berkeley Department of Economics. January 27, 1999. [↑](#footnote-ref-9)
10. Smart, Diana., and Suzanne Vassallo. “In the Driver’s Seat II: Beyond the Early Driving Years.” No. 17. April 2010. [↑](#footnote-ref-10)