# Harnessing Predictive Models to Enhance Road Safety: An Econometric Investigation into Risk Compensation & A Test of Machine Learning Model Efficacy

Zachary Cohen

Presented to the Economics Field Group
in partial fulfillment of the requirements for a
Bachelor of Arts degree in Economics with Honors

Pitzer College
Claremont, California

April 8th, 2024

## ACKNOWLEDGMENTS

I would like to extend my gratitude and appreciation to Dr. Linus Yamane and Dr. Zachary Dodds for their assistance and mentorship during the undertaking of this project. Both Professor Yamane and Professor Dodds graciously volunteered their time to provide feedback and guidance, which was instrumental in the success of this paper.

I can accredit Professor Yamane for my eagerness and curiosity in the field of economics, beginning with his Econometrics course I took during the spring of my sophomore year. I would have failed the course if not for bi-weekly office hours visits, during which Professor Yamane was patient and generous with my barrage of questions.

To Professor Dodds, I owe a deepened appreciation for computer science and its many applications. Professor Dodds was my first computer science professor with his introductory course Intro to Computer Science at Harvey Mudd. Without our pre-lecture conversations every day before class this semester in his financial technology course, my understanding of the inner workings of machine learning techniques would have been diminished.

# Table of Contents

# Chapter 1:
## Introduction

Prior to the COVID-19 pandemic, road use in the US was steadily rising, reaching 3.3 trillion vehicle miles traveled (VMT) in 2019, according to the Federal Highway Administration. With more road use, the measures we take to ensure road safety become increasingly important. The introduction of the National Traffic and Motor Vehicle Safety Act of 1966 was the first act of U.S. legislation that required automotive manufacturers to abide by safety standards, such as the lap belt, which protected drivers and occupants from the risks of traffic accidents. Today, these safety standards have become universally adopted and more robust, and thus, examinations into their effectiveness are progressively more warranted.

It is widely agreed upon in the literature that automotive safety devices, particularly seatbelts, are effective in saving lives in the event of an accident. However, whether safety regulations are related to the rate at which crashes happen is less understood. Some, such as Peltzman (1975), propose that there may be risk compensation effects related to the use of seatbelts. Risk compensation theory postulates that as the perceived cost of risk for some activity decreases, an individual's appetite for risk increases. Peltzman hypothesized and found that seatbelts lowered drivers' perceived cost of risk and thus increased their appetite for driving intensity.

Peltzman's results have been highly scrutinized, with many, such as Robertson (1977), and Cohen et al. (2001), claiming he used improper statistical techniques and proxy measures. Even if Peltzman's conclusions were valid, I hypothesize that today, there are no risk compensation effects associated with seatbelts; wearing a seatbelt is such habitual and benine behavior that it is unlikely to affect driving habits. The first chapter of this paper provides an up-to-date empirical analysis on the life-saving potential of seatbelts conditional on being in an accident, as well as an investigation into the existence of any risk compensation effects pertaining to seatbelt usage. To investigate seatbelt effectiveness in the event of an accident, I control for other risky behaviors and sex-based differences, revealing significant heterogeneous effects amongst men and women. My general conclusions support the consensus in the literature that seatbelts are an effective life-saving tool in the event of an accident and show that risky behaviors, such as the use of illicit substances and speeding, are associated with increased odds of death. My results fail to make any definitive conclusions on risk compensation as it pertains to safety devices but illuminate areas where further research may improve to uncover any such effects.

Beyond an investigation into the presence of risk compensation behavior, in Chapter 2, I leverage automotive fatality data to assess and compare different modeling approaches. Economic research techniques most commonly utilize econometric models such as the ordinary least-squared regression or logistic style regressions. While these techniques illicit highly interpretable results, the individual nature of the models can impose an overly strict set of assumptions which, if violated by the data, can produce spurious results. More nuanced ensemble-based machine learning methods, such as the random forest, can yield more reliable predictions. The objective of this research presented in Chapter 2 is to evaluate and compare the efficacy of ensemble-based machine learning models and individual econometric models in classification tasks, specifically assessing the performance of logit, probit, and random forest models.

The predictive task for which the models will be tested is to predict, given a set of driver and vehicle characteristics, whether a driver or passenger of a motor vehicle dies as a result of a crash. With rising rates of road travel and high rates of automotive incidents, confident predictions pertaining to on-road fatalities have life-saving implementations.

I use a comprehensive array of metrics to compare model performance across random forest, probit, and logit models, ultimately revealing that the performance of the random forest model surpasses that of the logit and probit in every aspect. A comparison of confusion matrices and a McNemar test further corroborates the superior performance of the random forest model. Beyond performance results, interpretability and inference capability across the models are also assessed, revealing the black box and difficult-to-interpret nature of the random forest.

Ultimately, I conclude that the random forest models’ comprehensive outperformance over individual models shows the potential of ensemble-based machine learning approaches in classification performance. However, there are drawbacks regarding interpretability, indicating further areas for exploration and the need for the use of econometric models for inference.

# Chapter 2:

Are Seatbelts Safe: Risk Compensation and The Efficacy of Automotive Safety Regulation in The U.S.

## I. Literature Review

### *A. Automotive Safety Regulation and Restraint Use*

The discourse on automobile safety regulation has evolved over time, with early literature offering varied opinions on its effectiveness. From the introduction of the National Traffic and Motor Vehicle Safety Act of 1966 to contemporary understandings, this review traces the trajectory of automotive safety literature. Attention must be paid to findings related to the efficacy of automobile safety due to life-saving policy implications. In 2022, the National Highway Traffic Safety Administration (NHTSA) reported that 42,795 people died in motor vehicle traffic crashes, and the US Transportation Secretary referred to traffic deaths as a "national crisis" (NHTSA). With many factors being relevant in the pursuit of safer roads and lowering the vehicle-related death toll, how we direct our resources is a question of crucial importance. If, for instance, it is found that seatbelts are an effective tool for reducing automotive fatalities, then there should be resources dedicated to ensuring the use of such tools. Reducing automotive deaths not only has positive social impacts but also could drive positive economic impacts in terms of reduced healthcare costs. For instance, Conner et al. (2010) find that in Ohio, a 10% increase in seatbelt usage would lead to medical cost savings of $15,376,787.

The consensus amongst the bulk of literature on automotive safety regulations, such as the seatbelt, is that regulation poses an effective tool to reduce vehicle-involved deaths. Early writing, by Huelke & Gikas (1966) finds that seatbelts have the potential to reduce fatalities by 40%. Modern examinations of automotive safety regulations have confirmed their life-saving potential. Carpenter & Stehr (2008) found that between 1991 and 2005, seatbelt laws were effective in increasing seatbelt use and subsequently reducing fatalities by 8%. Additionally, in a comprehensive study on the impact of state laws on vehicular fatality rates, Notrica et al. (2020) discovered that seatbelt regulations significantly decrease fatality rates amongst those under 65. Specifically, fatalities dropped by 11% for those aged 56-65, 16% for the 16-20 age bracket, and 24% for individuals between 21 and 55. Subsequently, driver age will be included as a control variable in my model.

The literature also suggests the presence of heterogeneous effects of sex on seatbelt efficacy and crash risk. Cullen et al. (2021) use a binomial regression to model the effects of sex on crash risk. They find that men are at higher risk for accidents compared to women at all ages. The conclusions given by Cullen et al. suggest that sex-based differences exist in both crash involvement and injury severity, where injuries for women were more severe. However, the use of hospital admissions rates to conclude that injuries for women are more severe may have led to a confounding effect from gender bias within the care system. Cullen et al. suggest the possibility that there is a lower threshold for the admittance of women into hospitals. My research will, therefore, omit admissions rates and instead use fatalities, which assume a finite value of 0 or 1. In my analysis of seatbelt efficacy, in terms of the probability of death conditional on being in an accident, I include a variable for sex and similarly uncover differential effects of belt usage for men and women.

Similar to Cullen et al., Lardelli-Claret et al. (2011) conclude that amongst all age groups, excluding their oldest age cohort, men are overrepresented in crash risk compared to women. Given the suggestion and consensus in the literature that men exhibit a higher risk for accidents than women, it might also be expected that there is a difference in driving behavior. It could thus be the case that accident severity may be different between men and women, causing a difference in injuries. Forman et al. (2019) find differences in injury risk amongst belted occupants; women are at greater risk for moderate and serious injury even after controlling for age, height, BMI, and vehicle model year. Notably, Forman et al. propose that sex itself has an effect on injury tolerance, where women are more vulnerable than men. Given the sex-based differences in the literature, my examination will take a sex variable into account when considering the effectiveness of seatbelts as a life-saving tool.

Going against the widespread consensus that automotive safety regulation is an effective tool to reduce vehicle-related fatalities, Peltzman (1975) takes the opposing stance. Peltzman argues that there is a moral hazard problem where safety regulation increases drivers' appetite for risky driving behavior, in turn offsetting the lifesaving potential of devices and shifting the burden of damages to pedestrians. Peltzman proposed that there is a direct relationship between driving intensity and the probability of death. Peltzman finds that while seatbelts initially lower the probability of death at any given driving intensity, with the advent of seatbelts, demand for driving intensity increases. Peltzman did not conclude definitively on the intuition that drives this result but theorized that drivers had offset the safety benefits of safety devices with an increased appetite for risky driving behavior. Peltzman presents time series evidence that supports this theory. Ultimately, due to the offsetting effect seatbelts have on fatalities, Peltzman concludes that safety regulation is ineffective in that it has no effect on vehicle-related death rates. Peltzman was one of the early pioneers of risk compensation theory, which states that there is an inverse relationship between the cost of risk and risky behavior. His application of this theory to automotive safety has been dubbed the Peltzman effect. Affirming references to the Peltzman effect as it pertains to automotive safety are rare; most include a harsh evaluation of the methods and theories Peltzman proposed in his 1975 work.

Questionable methods of empirical investigation have led others to be skeptical of the results presented by Peltzman. Robertson (1977) goes as far as to say Peltzman had "mistakenly claimed support" for his hypothesis of increased demand for risky driving on account of inappropriately calculated fatality rates and proxy measures. Additionally, Cohen et al. (2001), note that Peltzman, as well as the few others who support his theory of the Peltzman effect, fail to account for a possible endogeneity issue of seatbelt usage. Robertson (1977) supports this notion by adding how Peltzman failed to report "rudimentary" tests on the validity of his models, thus failing to identify such problems. Having adjusted for endogeneity, Cohen et al. (2001) found that seatbelt usage does significantly reduce fatalities among vehicle occupants and does not incur damage to pedestrians and cyclists. Peltzman also used the pre-post comparison method to estimate the counterfactual scenario where drivers are unbelted, taking advantage of data from the period preceding the implementation of safety regulation. The pre-post method, in this case, is an invalid comparison method due to the lack of control for time-related differences. My analysis does not use any pre-post comparison and rather focuses on data from just one year.

More general research on risk compensation and moral hazard in the context of risk-intensive activities has yielded mixed results. For example, Scott et al. (2007), in an assessment of risk compensation amongst alpine ski and snowboarders, find no evidence of risk compensation amongst helmet users. Alternatively, Weisburd (2015) found significant evidence pointing towards the presence of a risk compensation effect amongst automobile users. Weisburd noted that the adoption of automotive insurance increases the probability of a crash. Specifically, Weisburd found that a "$100 discount in the cost of an accident increases the probability of an accident by 1.7 percentage points."

Similar to Cohen et al. (2001), Robertson (1977), and Notrica et al. (2020), I hypothesize that there is no risk compensation or Peltzman effect happening with automobile safety devices and that instead, they act as an effective tool to reduce automotive fatalities. To evaluate this theory, this study will expand on existing work with more up-to-date cross-sectional data analysis on the effects of automotive safety devices on crash severity as well as on crash frequency to examine any possible risk compensation effects.

The discourse on automotive safety reveals a spectrum of perspectives. The majority of the literature on the topic highlights the positive implications of safety devices; however, there also exists a counter-narrative that proposes a net-negative effect of safety devices. Contemporary investigations validate the benefits of safety regulations and undermine the work of Peltzman in 1975. Utilizing a cross-sectional data methodology on a fresh dataset, this study seeks to add depth to the existing literature on automotive safety by evaluating the life-saving potential of safety devices and examining the potential existence of any risk compensation effects. The value of this work is to inform future policy choices on automotive safety with life-saving potential.

### *B. Other Notable Influences on Crash Risk and Injury Severity*

To pursue an analysis of automotive safety regulation efficacy in terms of the seatbelt and its lifesaving potential, adjustments must also be made for the potential influence of other risk factors. Research has pointed to determinants beyond restraint use as significantly influencing accident frequency and severity. Behaviors such as speeding, drinking, and drug usage are all thought to be important predictors of accident risk and severity. The NHTSA claims that speeding reduces the ability to steer, extends stopping time, and increases the risk of collisions due to unpredictable driving behavior. Likewise, Doecke et al. (2020), in an assessment of the relationship between impact speed and injury, find speed to have a highly significant and positive relationship with the risk of serious injury among vehicle occupants.

The documentation of the relationship between drinking and driving is more multifacited than that of speeding but universally agreed upon as a significant influence on driving behavior. On the more intuitive side, it is proposed that drinking increases the risk of high-severity accidents; Levitt & Porter (2001) note that drivers with any alcohol in their system are seven times more likely to cause a fatal crash. They also find that drivers with a blood alcohol content (BAC) above 0.10 are 13 times more likely to cause a fatal crash. These conclusions are supported by Drummer et al. (2020), who find that there is a positive relationship between BAC and whether or not a driver is considered responsible for a crash, measured via an 8-point culpability assessment. Specifically, they find that BAC begins to have a significant increase in culpability above levels of 0.05. However, there is also literature that suggests alcohol may have a diminishing effect on morality rates. Freidman (2012) finds that increases in blood alcohol concentration are associated with a significant decrease in in-hospital mortalities, but fails to identify the culpable biomechanism.

Drug usage, like alcohol use, is thought to have impairing effects, altering drivers' response time, thus having a possible association with crash risk. Via culpability analysis, where the proportion of drivers who were found to be responsible for a crash (culpable drivers) using a drug is compared to those not using a drug, Drummer et al. (2020) find that the effects of illicit drug presence (mostly THC or Methamphetamine), without the presence of alcohol, to be a 10-fold increase in the odds of being responsible for a crash.

The literature on drug usage is often concerned with making a discernment between the magnitude of the effect of drug usage compared to that of alcohol usage on injury severity and crash risk. The analysis perused in my report, although primarily concerned with the fact that these behaviors seem to be significant predictors of crash risk and injury severity, will also examine the differing effects of the two behaviors. Due to the significance of alcohol, drug usage, and speeding in driving behavior documented in the literature, all variables will be accounted for when making assessments about injury severity.

## II. Theoretical Background

The literature on automotive safety presents two conflicting views: automotive safety regulation is an effective tool to reduce automotive-related fatalities, or alternatively, automotive safety regulation has no effect or an increasing effect on automotive-related fatalities. The first stance, the more widely supported of the two, is affirmed by many empirical assessments, such as that undertaken by Cohen et al. (2001), who found that seatbelts effectively reduce automotive fatalities. The second stance has been spearheaded by Peltzman (1975), who found the seatbelt and automotive regulations to have a negligible effect on automotive safety. Peltzman claimed that automotive safety regulation increased drivers' appetite for risky driving, offsetting the lifesaving benefits of belts and creating more damage for pedestrians.

Both hypotheses are based on the basic demand model, which states that as the price of an item increases, the quantity of demand for that item decreases and vice versa. In regards to automotive safety regulation, seatbelts can alter the perceived cost of an accident and thus change the demand for risky driving behavior. Peltzman proposed the now dubbed “Peltzman effect,” which supposed that the advent of safety regulation increased the demand for risky driving behavior, thus increasing the number of automotive-related incidents. While the number of automotive-related accidents increased, Peltzman found that deaths of drivers and vehicle occupants stayed the same as levels before safety regulation. The burden of the increased crash incidence was felt by pedestrians, who are unable to partake in automobile safety device usage.

Peltzman’s theory, or the Peltzman effect, is an adaptation of risk compensation theory (RCT), which postulates that individuals change their behavior according to levels of perceived risk. Specifically, as the consequence of risky behavior decreases, individuals may be incentivized to engage in more risky behavior under the guise that they are protected against consequences. In the context of automotive safety, RCT suggests that drivers might experience higher levels of perceived safety due to safety regulations such as the seatbelt, which may empower them to drive in a riskier manner. According to RCT, safety regulations may alter an accident's perceived consequences, which might seem less severe due to the added protection. This theory supports Peltzman's arguments, suggesting that while regulations might decrease the direct harm of an accident to the driver, they might inadvertently promote behaviors that could lead to more accidents, affecting unprotected parties like pedestrians.

Peltzman's findings are unlikely to hold today due to a change in driving norms and faulty empirical analysis. The primary factor contributing to the widespread denunciation of Peltzman's findings is a lack of empirical validity; Peltzman failed to run the proper tests and used questionable proxy measures (Robertson, 1977). Even if Peltzman's results were to be interpreted as valid, his findings would still be unlikely to hold with modern data. The primary difference being the salience of automotive safety regulations. Peltzman’s study pulled data from just after the advent of the automotive regulations, meaning that safety devices were new for drivers. The event of putting on a seatbelt was most likely a very noticeable experience, thus making it more reasonable for Peltzman to hypothesize an effect on driving behavior. Today, seatbelts are so commonplace that the act of putting one on is so benign that it is likely to happen subconsciously, thus most likely not altering a driver’s appetite for risky driving.

I expect that seatbelts are an effective tool for reducing automotive fatalities and do not alter drivers' demand for risky driving. My expectations align with the findings of recent work within the automotive sector, such as that conducted by Carpenter & Stehr (2008), who find that an increase in seatbelt usage leads to a decrease in automotive-related fatalities. These expectations are also congruent with research on risk compensation in other fields. For instance, (Scott et al., 2007) found no evidence of risk compensation amongst helmet users in alpine sports. This paper seeks to test for the presence of a Peltzman effect, or risk compensation theory, via the use of modern data and the introduction of relevant proxy variables to assess the relationship between the relative cost of accident and accident probability.

## III. Data

### *A. Descriptive Statistics*

The data is divided into two buckets: Table 1 shows an individual-level dataset used to examine the determinants of death conditional on being in an accident. Table 2 shows a state-level dataset used to examine influences on the frequency of automotive accidents, with the intent to test for the existence of any risk compensation effects.

Table 1 reports basic descriptive statistics on 32,225 drivers involved in motor vehicle accidents, resulting in fatalities. Data was compiled from the Fatality Analysis Reporting Systems (FARS) 'Vehicle' and 'Person' tables from 2019. FARS records an observation for every motor vehicle incident where there is a fatality, meaning that if, for example, two vehicles crash and there is a fatality in one, observations are recorded for both. The 'Vehicle' table consists of one record for every vehicle involved in an accident where there is a fatality for a given year. The 'Person' table contains 1 entry for every person (drivers, occupants, and pedestrians) involved in a vehicle accident in a given year. Each accident is given a unique identifier number called the 'State Case', and each vehicle involved in an accident is given an ID number under 'Vehicle Number.' Both the 'Vehicle' and 'Person' tables contain the 'State Case' and 'Vehicle Number' variables, making it possible to join the two. Before joining, because there could be more than one person for every vehicle, only the drivers were selected from the 'Person' table to ensure an equal number of observations in each table.

**TABLE 1**
DESCRIPTIVE STATISTICS: INDIVIDUAL LEVEL

| VARIABLES | Mean | SD | Min | Max |
|---|---|---|---|---|
| Driver fatality | 0.552 | 0.497 | 0 | 1 |
| Restraint use | 0.616 | 0.486 | 0 | 1 |
| Alcohol use | 0.188 | 0.391 | 0 | 1 |
| Drug Use | 0.079 | 0.271 | 0 | 1 |
| Speeding related | 0.183 | 0.387 | 0 | 1 |
| Driver age | 43.949 | 18.330 | 9 | 100 |
| Driver sex | 0.738 | 0.440 | 0 | 1 |
| Car model year | 2008 | 7.544 | 1930 | 2020 |
| Driver height | 68.601 | 3.911 | 37 | 83 |
| Driver weight | 186.436 | 46.556 | 50 | 550 |
| Speed limit | 49.668 | 13.269 | 0 | 80 |
| Rush hour time | 0.427 | 0.495 | 0 | 1 |
| Daytime lighting | 0.701 | 0.458 | 0 | 1 |
| Cold weather | 0.219 | 0.413 | 0 | 1 |
| License status | 0.884 | 0.320 | 0 | 1 |
| Airbag deployment | 0.325 | 0.468 | 0 | 1 |
| Observations | 32,225 | 32,225 | 32,225 | 32,225 |

**Table 1:** Table values are descriptive statistics for all drivers in all vehicles involved in a crash in 2019. 'Driver fatality,' 'Restraint use,' 'Alcohol use,' and 'Speeding related' are all {0, 1} indicators where 1 indicates an affirmative value. 'Driver sex' is a dummy variable where 1 indicates male and 0 female. Rush hour time' indicates if a crash occurred during rush hour (1 for yes, 0 for no). 'Daytime lighting' is a {0, 1} indicator for whether the accident occurred in daylight (1 for yes). 'Cold weather' is a {0, 1} indicator for whether the weather was cold (1 for yes). 'License status' is a {0, 1} indicator for whether the driver has a valid license (1 for yes). Continuous variables are measured as follows: 'Driver age' in years, 'Car model year' as the calendar year, 'Driver height' in inches, 'Driver weight' in pounds, and 'Speed limit' in miles per hour. 'Observations' represent the number of individual driver cases analyzed.

Column 1 of Table 1 reports the mean values for each variable, and column 2 reports the standard deviation. Columns 3 and 5 report the minimum and maximum values. The majority of drivers (55%) in the sample died as a result of the crash. 61% of drivers wore seatbelts, while 39% of drivers either improperly wore a belt or did not wear one. 18% of drivers had a positive blood alcohol content (BAC), and 8% of drivers were subject to some level of drug involvement at the time of the incident. A minority of drivers (18%) in the sample were speeding at the time of the accident. 74% of the sample is male. The average sample subject is 43 years old, 5'7" tall, and weighs about 186 pounds. Interestingly, the youngest 'driver' in the sample was 9 years old. While the 9-year-old in question was not actually a driver, she was sitting in the driver's seat of a parked vehicle that was stuck by another. A large majority (88%) of the sample has a valid driver's license, while 12% of the sample does not. Most (57%) of crashes occurred in non-rush hour conditions when it was light outside (70%). 78% of crashes happened when it was not cold outside. The airbag was deployed in 33% of incidents.

Table 2 reports descriptive statistics for accident data across 51 US states, including D.C., where an accident had at least 1 fatality (later used as data for the model in Table 4). Measures for the fatality rate of drivers (Driver fatality rate) and pedestrians (Pedestrian fatality rate) are constructed via a joining of the 'Accident' table from the National Center for Statistics and Analysis Fatality Analysis Reporting System (FARS) and vehicle travel data from the Bureau of Transportation Statistics (BTS).

First, data from the FARS 'Accident' table is sorted into groups by state and summarized via a count of deaths for each state. The joining of vehicle miles traveled (VMT) state-level data from the BTS allows for the construction of a rate measure for fatalities, which is fatalities divided by 100 million VMT. The average number of fatalities per 100 million VMT across all 51 observations is 0.887 for drivers and 0.212 for pedestrians. These measures account for the frequency of a fatality occurring per a standardized unit of 100 million VMT.

**TABLE 2**

DESCRIPTIVE STATISTICS: STATE LEVEL

| VARIABLES | Mean | SD | Min | Max |
|---|---|---|---|---|
| Driver fatality rate | 0.887 | 0.265 | 0.346 | 1.370 |
| Pedestrian fatality rate | 0.212 | 0.094 | 0.040 | 0.427 |
| Seatbelt usage rate | 0.891 | 0.056 | 0.707 | 0.971 |
| Drunk drivers | 0.274 | 0.069 | 0.136 | 0.528 |
| Mean household income | 69,721 | 11,715 | 44,790 | 95,570 |
| Population Density | 423.635 | 1574.301 | 1.300 | 11,280 |
| Seatbelt law | 0.980 | 0.140 | 0 | 1 |
| Seatbelt law type | 0.686 | 0.469 | 0 | 1 |
| Observations | 51 | 51 | 51 | 51 |

**Table 2:** Table values are descriptive statistics for average values of each variable for 50 US states and Washington D.C. Pedestrian and driver fatality rates are the number of fatalities for every 100 million vehicle miles traveled in a state

Seatbelt usage rates for each state are joined on with NHTSA data from an annual Traffic Safety Facts Report. The mean seatbelt usage rate across all states is 89%, with a minimum usage rate of 71% in New Hampshire, where there is no seatbelt law. The maximum belt usage rate is 97% in Hawaii. Information on drunk driving is also obtained from FARS; 'Drunk drivers' measures the percent of fatal accidents in a state where alcohol was involved (27%). Rhode Island has the highest rate of drunk drivers in fatal accidents (53%), while Washington DC has the lowest at 14%. To account for the relative cost of a crash, median household income data is joined onto the dataset from the United States Census Bureau Historical Income Tables. The mean household income across all states in 2019 was $69,721. Dummies for seatbelt law and seatbelt law type are constructed from the Governors Highway Safety Association (GHSA). Seatbelt law accounts for whether a state has or does not have a law. Seatbelt law-type documents whether a state has primary or secondary enforcement seatbelt laws; 69% of states have primary enforcement seatbelt laws, meaning a citation can be conducted solely on the basis of a driver not wearing a belt.

### *B. Visual Trends*

**FIGURE 1**

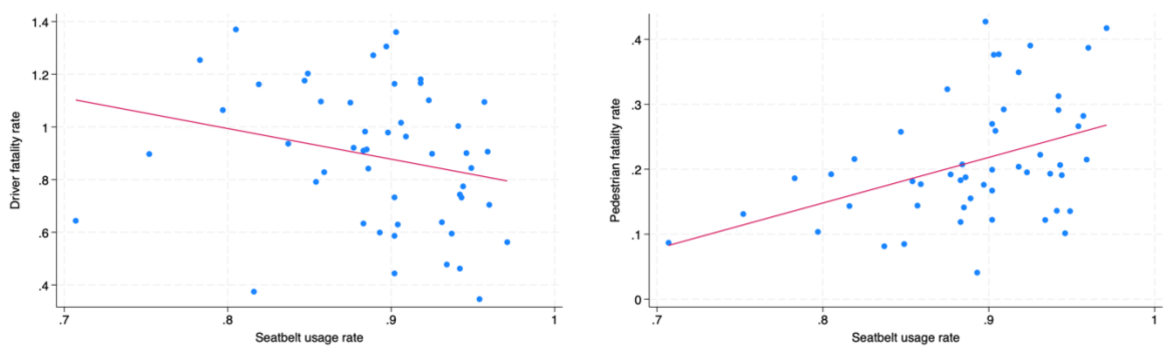
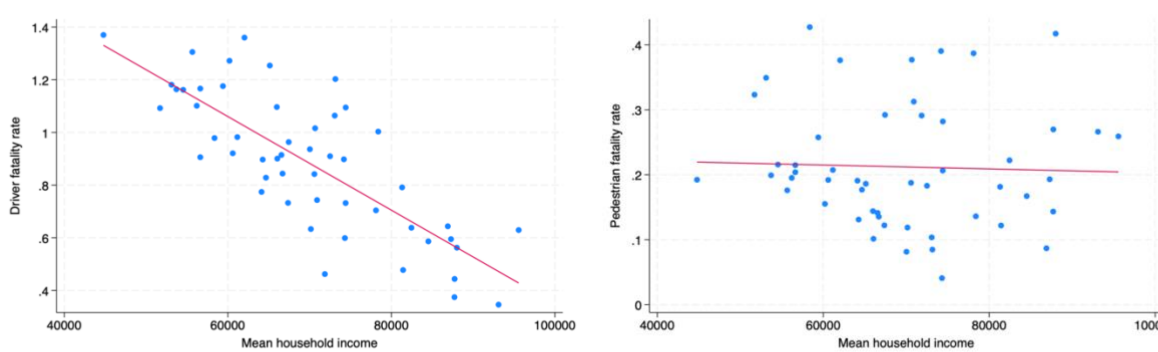



**FIGURE 2**



From the right side of Figure 1, there is a positive visual relationship in the data between seatbelt usage and pedestrian fatalities, indicating the possible presence of a ‘Peltzman effect’ within the dataset. On the left side of Figure 1, there is a negative visual trend between seatbelt usage and the rate of driver fatalities, indicating possible support for the life-saving potential of seatbelts. A negative relationship between mean household income and the rate of driver fatalities is shown in Figure 2. A negative relationship between income and the driver fatality rate goes against the demand model that Peltzman proposed, where, as the relative cost of accidents decreases in relation to household income, demand for risky driving behavior increases. Here, we see that as the relative cost of accidents decreases, so does the rate of driver fatalities. Additionally, from Figure 2, there seems to be no observable pattern between mean household income and the rate of pedestrian fatalities. Although these visual relationships give insights into possible trends, they do not indicate causality or significant effects. A more thorough econometric analysis will be perused in the following section to support the high-level insights gleaned from preliminary charts and summary statistics.

## IV. Results

### *A. Determinants of Driver Death Probability Conditional on Being in an Accident*

Table 3 reports results from a probit style regression on the effects of various behavioral choices on a {0, 1} indicator for whether a driver has died as a result of an accident (1 indicating death and 0 indicating no death). In this case, the event of a fatality is used as a proxy to assess the severity of injury, where death indicates peak severity, and no death indicates less severe injuries. The output is aimed at assessing the mechanisms and behaviors that play a significant role in determining injury, conditional on being in an accident. Results were not sensitive to the specification of probit compared to logit, hence, the probit specification was chosen due to a higher intellectual appeal. Results are shown with and without controls for various driver and road condition characteristics (Table 3.A reports results from a probit regression which shows that a large majority of the controls are highly significant, indicating strong effects on injury severity, thus their inclusion in the estimation of marginal effects). Column 3 illustrates marginal effects, which allows for the interpretation of coefficient magnitude.

To contextualize the marginal effects estimates in column 3 of Table 3, a baseline probability of death for the average driver in the sample is constructed. To uncover what the probability of death is, given being in an accident, for the driver with average characteristics amongst the sample, mean values for each variable are substituted into equation 2. The summation of the products of mean values and coefficient estimates yields a z-score, which can then be cross-checked against the standard normal table to assess the probability of that value occurring within the standard normal distribution. In the sample, the driver with average characteristics has a 58% chance of dying, conditional on being in an accident. This estimate is consistent with the mean fatality rate amongst the sample of 55%, as shown in Table 1.

***Equation 1 (Column 1):***

$$\phi(Y_i) = \beta_0 + \beta_1 RestraintUse + \beta_2 AlcoholUse + \beta_3 DrugUse + \beta_4 SpeedingRelated + e_i$$

***Note.*** *Probit specification without the inclusion of any control variables.*

***Equation 2 (Column 2):***

$$\phi(Y_i) = \beta_0 + \beta_1 RestraintUse + \beta_2 AlcoholUse + \beta_3 DrugUse + \beta_4 SpeedingRelated + Controls + e_i$$

***Note.*** *Probit specification with the inclusion of controls referenced via 'Controls' as shown in Table 3.A.*

The presence of heterogeneous effects of belts can also be tested via this method. Given being in an accident, the driver with average characteristics who is wearing a seatbelt has a 40% chance of death. This estimate is a 52% decrease compared to the probability of dying for those not wearing a seatbelt (83%). Additional heterogeneous effects are present across genders; females wearing seatbelts in the event of an accident are 45% less likely to die compared with males wearing seatbelts (80% chance of death for males and 44% chance of death for females). The disparity across gender groups is present but with diminished effects when not wearing belts. In the event of an accident, when not wearing a seatbelt, male drivers are 12% more likely to die compared to females.

Within gender variation also reveals information about seatbelt efficacy for each group. Seatbelt usage for the average male in the sample decreased the probability of death by 18%, whereas seatbelt usage decreased the probability of death for females by 49%.

Columns 1 and 2 from Table 3 illustrate that restraint use, alcohol use, drug use, and speeding are all significant in their effects on injury severity. To interpret the magnitudes of estimates, a marginal effects specification was undertaken. The first estimate in column 3 shows that restraint use decreases the probability of death by 35.35 percentage points ($p < 0.01$) amongst drivers, given the event of getting in an accident. This is a 60% decrease from the expected probability of death for the average driver in the sample of 58%. This result intuitively makes sense and is generally consistent with the consensus amongst the literature on automotive safety; Carpenter & Stehr (2008) find that increases in seatbelt usage lead to an 8% decline in fatalities. The findings here also align with more dated work, such as that of Huelke & Gikas (1966), who estimate the fatality-reducing potential of seatbelts to be near 40%. Understandably, wearing a belt reduces the likelihood of death, whereas not wearing one or improper use increases the odds of death.

**TABLE 3**

DETERMINANTS OF DRIVER DEATH (Y = 1 death, 0 no death)

| | (1)<br>(Probit) | (2)<br>(Probit) | (3)<br>(Marginal Effects) |
|---|---|---|---|
| Restraint use | -1.140*** | -1.228*** | -0.353*** |
| | (0.0170) | (0.0182) | (0.00405) |
| Alcohol use | 0.605*** | 0.665*** | 0.191*** |
| | (0.0233) | (0.0248) | (0.00693) |
| Drug use | 0.235*** | 0.217*** | 0.062*** |
| | (0.0311) | (0.0321) | (0.00920) |
| Speeding related | 0.528*** | 0.633*** | 0.182*** |
| | (0.0226) | (0.0236) | (0.00657) |
| Constant | 0.680*** | 52.393*** | |
| | (0.0154) | (2.237) | |
| Controls | | Y | Y |
| Observations | 32,225 | 32,225 | 32,225 |
| Pseudo R-Squared | 0.200 | 0.264 | 0.264 |

**Table 3:** Estimates from probit regressions on probability of death where the dummy variable is a 0 or 1 indicator where 1 indicates death. Controls seen in table 3.A. Standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Moving on in column 3, the estimate on alcohol use shows that drivers who are drinking are 19.1 percentage points more likely to die given that they get into an accident ($p < 0.01$), a 32% increase upon the expected probability of death for the average driver. This estimate is not consistent with the theories of relaxation effects associated with alcohol consumption lowering the probability of injury, which suppose alcohol to have a negative effect on the odds of death. For instance, Friedman (2012) finds that increases in blood alcohol concentration are associated with a significant decrease in in-hospital mortalities. The positive estimate for alcohol in this analysis is likely to be picking up the effect that alcohol could have on accident severity. Inebriated drivers are likely to drive erratically, thus crashing more often and having more severe accidents. Therefore, when there is alcohol involved, since accidents are likely to be more severe, drivers are more likely to die, negating any possible relaxation effects.

Similar results are shown for the estimated effects of drug use, illustrating a 6.24 percentage point increase in the odds of death ($p < 0.01$) or an 11% increase in the odds of death for the average person in the sample. While the effects of drug use are smaller in magnitude, the intuition behind the positive relationship is similar to that behind the effects of alcohol use; drug-related incidents are likely to be more severe and frequent due to non-typical driving behavior, thus increasing the likelihood of death. Similar results are found in the estimate for speeding-related effects. Accidents that are speeding-related increase the probability of driver fatality by 18.2 percentage points ($p < 0.01$), a 31% increase over the expected probability of death.

### *B. Factors Influencing the Frequency of Fatal Accidents*

Table 4 moves forward to examine influences on the frequency of getting into fatal accidents for drivers (column 1) and pedestrians (column 2). The key difference between the examination here as it compares to that of Table 3 is that getting in an accident is no longer given. Therefore, estimates indicate kinds of behaviors that change the likelihood of getting into an accident and dying, in contrast to above, where estimates indicate the relationship between a behavior and the probability of death conditional on being in an accident.

Column 1 shows the estimated effects of restraint use, alcohol use, household income, seatbelt law types, and population density on an outcome variable measuring the probability of getting into a fatal accident. The response variable is calculated as the total number of fatalities within a state, normalized by the total vehicle miles driven in that state, and then scaled to reflect the number per 100 million VMT. In column 1, the dependent variable is reflective of all driver fatalities, whereas in column 2, the count is restricted to pedestrian fatalities. Splitting the analysis into pedestrian and driver categories shows how the effects of safety devices and behavioral choices affect the likelihood of death amongst both groups differently.

The results in column 1 indicate that seatbelt usage rate has no significant relationship with the frequency of fatality for drivers. Only household income has a significant effect on the probability of getting into a fatal accident, where a $1.00 increase in household income is associated with 0.0000164 fewer fatalities per 100 million vehicle miles driven ($p < 0.01$). In other words, to reduce fatalities by 1 fatality per 100 million vehicle miles, average household income would have to increase by $60,975. According to the results in column 1, no other variables have a significant effect on the probability of being in a fatal accident. While not significant, from interpreting the estimated effect of seatbelt usage rate, one would conclude that an increase in seatbelt usage rate decreases the probability of getting in a fatal accident for drivers. Likewise, although not significant, indicating no evidence of any relationship, if interpreting the estimated effect of drunk driving, one would see that it is negative, indicating that drinking and driving decreases the probability of getting into a fatal accident.

**TABLE 4**

FREQUENCY OF FATALITY AMONGST PEDESTRIANS AND DRIVERS

| | (1)<br>Driver Fatality Rate | (2)<br>Pedestrian Fatality Rate<br>(Robust errors) |
|---|---|---|
| Seatbelt usage rate | -0.031 | -0.272 |
| | (0.381) | (0.174) |
| Drunk drivers | -0.724 | 0.579** |
| | (0.519) | (0.231) |
| Mean household income | -1.64e-05*** | -5.59e-07 |
| | (2.24e-06) | (1.06e-06) |
| Seatbelt law type | 0.019 | 0.009 |
| | (0.0640) | (0.0321) |
| Population density | -1.74e-05 | 7.54e-07 |
| | (1.69e-05) | (4.67e-06) |
| Constant | 2.679*** | -0.198 |
| | (0.458) | (0.210) |
| Observations | 51 | 51 |
| R-squared | 0.644 | 0.226 |

**Table 4:** Estimates from linear regressions on frequency of fatal accidents where the dependent variable is a a measure of the number of deaths for every 100 million vehicle miles driven. Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Non-statistically significant estimates are likely to be non-representative of any real-world effects. The non-significant estimates in Table 3 should thus not be interpreted as true representations of real-world effects. However, while insignificant, these results could possibly reflect some issues within the model. Specifically, the suggestion that alcohol use leads to a decreased probability of driver fatality is counterintuitive to the otherwise well-understood effects and dangers of drunk driving. Given that there are only 51 observations, one for each state, the model here is unlikely to pick up any representative effects.

Column 2 of Table 4 shows the estimated effects of restraint use, alcohol use, household income, and seatbelt law types on pedestrian fatalities. The results indicate no significant relationship between seatbelt usage rate and the frequency of pedestrian fatalities. No estimates except the effects of drunk driving return as significant. The results in Table 4, column 2 show a coefficient of 0.579 ($p < 0.05$) on drunk driving, indicating a positive relationship between the percentage of drunk drivers in a state and the frequency of pedestrian fatalities. Since the rate of drunk drivers is expressed as a percentage, a one percentage point change corresponds to a 0.01 change in the rate when expressed as a proportion. Therefore, a one percentage point increase in the rate of drunk drivers is associated with an increase of 0.00579 pedestrian fatalities per 100 million VMT.

## V. Discussion

### *A. Overview*

The results of this study demonstrate the life-saving potential of seatbelts but fail to explain any association between the effects of automotive safety regulations on accident frequency. In other words, the models in this paper explain, conditional on being in an accident, whether seatbelts help but fail to explain whether seatbelts affect the rate at which people get into accidents. Additionally, the results support commonly observed positive associations of risky behaviors such as alcohol consumption, drug usage, and speeding on the severity of accidents.

### *B. Probability of Death Conditional on Being in a Severe Accident*

The findings in this paper on the effects of seatbelt usage are consistent with the general consensus in the literature, which supports the effectiveness of belts as a life-saving tool. From the estimates in Table 3, I find that the probability the driver with average characteristics in the sample dies, conditional on being in a fatal accident, is 58%. I find that drivers not wearing belts have an 83% chance of dying compared with a 40% chance of death for those wearing belts. The estimate on restraint use in column 3 of Table 3 is statistically significant ($p < 0.01$) and shows that there is a negative association between restraint use and the probability of death amongst drivers, given the event of getting in an accident. The result indicates restraint use is associated with a 35-percentage point decrease in the likelihood of death, conditional on getting into an accident. This is a 60% decrease in the expected probability of a fatality for the average driver in the sample of 58%. This decrease is generally consistent with the findings of literature of old and recent. Huelke & Gikas (1966) estimate that seatbelts have the potential to reduce automotive fatalities by 40%. More recently, Nortica et al. (2020) also find that seatbelts are negatively related to automotive fatalities. The study design does not allow for a causal interpretation of the relationship between seatbelts and the probability of death, but there is a clear negative association.

While the results, on average, support that seatbelts are an effective tool to reduce the probability of fatality conditional on being in an accident, the effects of belts across sex are seen to differ in magnitude but not direction. Overall, for both men and women, seatbelt usage reduces the probability of death, but more so for women. Women wearing seatbelts in accidents where there is at least one fatality have a 95% reduction in the relative probability of death compared to women not wearing belts. Seatbelts have the same effect on men but on a much more marginal scale, reducing the probability of death by 23%. This 72-percentage point difference between seatbelt effectiveness for men and women could be attributed to sampling bias or a difference in the driving behavior of men and women. It could be possible that there is some sort of selection bias attributed to random chance where the males in the sample are involved in more severe crashes. However, it could also be the case that males exhibit certain driving behaviors, which put them more at risk for severe crashes. In the sample, the average male is likely to die 87% of the time, and the average female is likely to die 62% of the time. These estimates are generally representative of trends observed by other studies, which conclude that men are overrepresented in crash risk compared to women (Lardet-Claret et al., 2011). While there has not been much investigation into why this difference exists, it could be due to a difference in driving tendencies based on sex. From my results, it seems as if men are getting into more severe accidents, and thus, seatbelts offer marginal protection for men against greater or more severe impacts.

The results I find in regard to seatbelt efficacy as it pertains to women go against the literature on sex-based differences in belt efficacy. My results support that seatbelts work better for women than they do for men. As expressed above, the magnitude of this difference is 72 percentage points. This goes against the findings proposed by Forman et al. (2019), who find that women are more vulnerable to seatbelt-related injuries than men, especially in the lower body.

While the intuition behind this result is more muddy, it could be the case that seatbelts are designed for a specific body mold, which does not represent the female bone structure. Our sample, however, does not support this theory. Again, this could be due to sampling bias, where the men in the sample seem to be getting into more severe accidents. Thus, seatbelts work worse among men compared to women. There could also be a broader bias in this assessment attributed to the fact that all of the data comes from incidents where there was at least one fatality. This means that, by nature, almost all of the crashes in the sample are very severe. The effects of belts for women and men in less severe accidents may be different and should thus be examined in future work.

Results in my examination point towards alcohol usage as a significant predictor of the probability of death, conditional upon being in a severe accident. The analysis indicates a statistically significant ($p < 0.01$) impact of alcohol consumption on mortality: for individuals involved in accidents, the likelihood of death when drinking is involved is 32% higher compared to the baseline probability of dying for the average person in the sample of 58%. It's difficult to compare these findings with other studies in the literature due to the limited research on the specific effects of alcohol on injury severity in the event of an accident. Substantial research exists showing a positive relationship between alcohol consumption and the risk of accidents. Both Levit and Porter (2001), and Drummer et al. (2020) find drivers with positive BAC are more likely to be responsible for accidents. These results, while not illustrating directly what I have concluded, propose the underlying effect that the results in my paper are most likely capturing: drivers who have positive BAC are crashing more often, meaning that there is likely some reckless or abnormal behavior they are exhibiting. Thus, drivers with alcohol in their systems may be more likely to get in severe accidents, therefore making them more likely to die in all accidents. While some studies, such as that of Friedman (2012), find that inebriation, in the event of an injury, is likely to decrease the odds of a fatality, my results sensibly do not illustrate this effect because there is an offsetting difference in the kinds of accidents drunk drivers are getting in.

I find similar significant findings regarding drug use as it pertains to the probability of death conditional on being in an accident where there is a fatality. Drug use correlates with an 11% increase in the probability of death in accidents involving a fatality compared to the baseline probability for the average individual in the study ($p < 0.01$). This indicates that drug use is a significant factor in accident-related fatalities. Similar to alcohol, the literature on the relationship between drug usage and injury severity given being in an accident is sparse. However, the literature does support the theory that drivers under the influence of drugs are more likely to be responsible for an accident (Drummer et al., 2020). It's not likely that being under the influence of drugs itself is increasing the probability of death, given being in an accident, but rather that drug usage is causing different kinds of accidents. Like drivers under the influence of alcohol, drivers involved with drugs are likely to drive more erratically, thus possibly causing more severe accidents. Therefore, my estimate that drugs are positively associated with the odds of fatality conditional on being in a crash makes sense. However, I fail to discern with certainty why that relationship exists or whether it is causal.

Similar to the estimated effects of drugs and alcohol, I estimate speeding to be highly statistically significant and positively associated with the probability of driver fatality conditional on being in an accident ($p < 0.01$). Those who are speeding have a probability of death, which is 31% higher than the average driver in the dataset. This result is consistent with a shared understanding of the dangers of speeding, as well as findings within the literature, such as Doecke et al. (2020), who find that speeding increases the odds of serious injuries among vehicle occupants.

The results here show that alcohol use, drug use, and speeding are associated with a higher risk of death conditional upon being in an accident where there is at least one death. More simply put, the use of illicit substances, or the decision to engage in risky driving behavior, is related to an increased risk of death when in severe accidents. Current road safety protocols highly discourage the use of alcohol and drugs while driving. This emphasis seems to be accurately informed and aligns with our findings; I find that both the use of alcohol and drugs increases the odds of death, but the economic significance of the estimated effect of alcohol is larger than that of drugs. Therefore, a larger regulatory focus on alcohol is adequate and could be a more cost-effective solution given the popularity of alcoholic beverages. It may also make sense from a regulatory stance to investigate patrolling procedures that lower the instances of speeding. Programs such as automated speeding cameras, which offer a cost-effective solution and decrease the necessity for manned patrol vehicles, may be effective in doing so.

### *C. Influences on Crash Frequency*

Moving on to the discussion on factors influencing the frequency of fatal accidents for drivers and pedestrians, the models presented in this paper fail to make any reliable predictions. A combination of data inaccessibility, small sample size, and oversimplistic proxy measures may have led to unreliable, difficult-to-interpret coefficients.

According to the models presented in Table 4, there is no significant relationship between the use of seatbelts and the frequency of driver fatalities or pedestrian fatalities. This result goes against the work of Peltzman and the proposed 'Peltzman effect' where belted drivers supposedly take more risk, causing a rise in pedestrian fatalities. Here, although there is no statistically significant relationship, if one were to interpret the estimated effects of seatbelt usage as significant, one would conclude that seatbelt usage is associated with a decrease in the frequency of both pedestrian and driver fatalities. The negative association of seatbelt usage and fatality frequency, if it were to be interpreted as significant, aligns with the work of Cohen & Eniav (2001), who find that seatbelt usage is negatively associated with the frequency of fatality for drivers and has no damaging effects on pedestrians. However, the non-significant results in the model are more likely representative of the hypothesis that seatbelts have no effects on driving behavior and that there is no existence of risk-compensation behavior, as also concluded by Robertson (1977) and Notrica et al. (2020).

While the estimates on seatbelt usage support the hypothesis that there is no risk-compensating behavior, the estimates on the effects of drunk driving and mean household income indicate some inconsistencies. The model in Table 4 shows that drunk driving has a statistically insignificant relationship with the frequency of fatalities for drivers. However, from interpreting the coefficient as significant, one would conclude that drinking has a negative relationship with the frequency of fatalities; in other words, increasing the number of drunk drivers would decrease the frequency of fatal accidents. Even while statistically insignificant at any reasonable level, this estimate is nonsensical and illuminates potential inconsistency within the model. According to the literature, as well as a common understanding of the dangers of drunk driving, it would be nonsensical to conclude that there is a real negative relationship between drinking and the frequency of accidents. Both Porter (2001) and Drummer et al. (2020) find that drunk drivers are more likely to cause and be responsible for accidents. Additionally, in my observations from Table 3 on the effects of alcohol on the probability of death conditional on being in an accident, I conclude that there is most likely a positive influence on accident severity related to drinking, which is biasing the estimated effects of alcohol usage upwards. Thus, concluding that alcohol decreases the instances of fatal accidents for drivers would be invalid.

Moving to the estimated effects of alcohol usage on pedestrian fatality frequency, a statistically significant and positive coefficient indicates that increases in the rate of alcohol consumption are related to an increase in the frequency of pedestrian fatalities in automotive accidents ($p < 0.05$). This result does make intuitive sense: unpredictable and risky driving behavior associated with drunk driving is likely to lead to increased crash incidence, thus increasing the number of pedestrian deaths. However, given that the model also supports an association between alcohol and decreased fatal crash incidence for drivers, a significant relationship between alcohol and pedestrian fatalities is not enough to confirm support for the Peltzman hypothesis.

Both estimates for the effects of household income on pedestrian fatalities and driver fatalities point to evidence against the existence of any risk compensation or Peltzman effects. The nonsignificant estimate on the association between mean household income and pedestrian fatality rate, if interpreted as a true effect, represents that as the relative cost of accidents decreases, there is no change in the pedestrian death toll. Looking at the association between mean household income and the frequency of driver fatalities, I find that increases in income are associated with decreases in the frequency of fatality. Although the economic significance of this estimate is small, it is significant at the 1% level. This result goes against the Peltzman hypothesis of a risk-compensating effect, where the perceived cost of accidents, relative to one's income, is decreasing, one's appetite for risk is increasing. Here, I find that as the relative cost of accidents decreases, or as median household income gets larger, the instances of fatal accidents decrease.

While the relationship between household income and the frequency of driver fatality is clear and statistically strong, scrutiny must be used when interpreting the result. I propose mean household income as a proxy measure for the relative cost of a crash, where, as household income is increasing, the relative cost of a crash is decreasing. One might argue that rising income is not necessarily related to a decrease in the relative cost of a crash but rather an increase: as one's income rises, so might the opportunity of the cost of a crash where if the driver is hospitalized or unable to work, the value of forgone income becomes larger. Additionally, with higher income, a driver might have a more expensive vehicle for which the cost of repairs may be greater. Thus, I can conclude there is a real relationship between mean household income and the frequency of accidents, but I cannot make conclusions on how this relationship relates to risk compensation due to the vague nature of household income as a proxy measure for the relative cost of a crash.

## VI. Conclusion

This study has covered several critical aspects of automobile safety, particularly the efficacy of seatbelts in saving lives and the complex relationships between risky behavior and accident severity. The evidence found supports the life-saving potential of seatbelts, conditional on being in an accident, showing a substantial decrease in the probability of death for belted individuals in severe accidents. This aligns with the consensus in the safety literature and confirms the importance of seatbelt use as a critical safety measure. Furthermore, this research also sheds light on the differential impact of seatbelts across sexes, finding that women experience a greater relative reduction in the probability of fatality than men when wearing belts. The heterogeneous findings across sexes challenge existing work within the literature and prompt the need for a reevaluation of sex-based differences in the efficacy of safety mechanisms.

The associations between alcohol and drug use and the severity of accidents are significant and reinforce previous common conceptions about the perilous outcomes stemming from driving under the influence. While this paper does not suggest a causal relationship, the significant correlation between substance use and increased probability of death in severe accidents should not be ignored.

While I find significant interpretable results as it pertains to the efficacy of automotive safety conditional on being in an accident, the findings on the influences of seatbelt usage, household income, and alcohol consumption on the frequency of accidents do not yield interpretable results. The small sample size and difficult-to-interpret proxy measures make for a statistically invalid model. While the results are non-significant, the study still raises questions about the potential existence of risk-compensating behaviors, promoting the need for further investigation with more sophisticated data methodology over a broader set of observations.

The insights gained from my work demonstrate robustness in some areas but also show limitations due to data constraints and potential biases. Thus, this examination should be used as a precursor for more detailed and sophisticated efforts into the complex network of variables that are associated with the incidence and severity of automotive accidents. Further research would do well to create a dataset that contains not only positive classification examples for when a driver does get into an accident but also negative examples for when a driver does not get into an accident. This would allow for discernment in the kind of behavioral characteristics that differ between those getting into a crash and those who do not. Additionally, a dataset where not all crashes involve a fatality may lead to a more relevant and broader set of applicable conclusions about what happens in all crashes rather than just those that are severe. Through such observations, a more comprehensive analysis can be undertaken to develop more effective safety regulations to reduce the on-road death toll.

# Chapter 3:
## Comparative Analysis of Ensemble vs. Individual Models in Predicting Automotive Fatality Outcomes.

### I. Literature Review

Academic discussions pertaining to the use of predictive modeling techniques to assess various topics in automotive safety are plentiful. As discussed in the prior chapter, on-road fatalities are a leading cause of death in the United States, and thus, research related to road safety has lifesaving implications. Beyond the existing body of research, which employs econometric models to understand traffic fatality patterns, ensemble-based machine learning models offer a more nuanced method of inference, possibly yielding better predictive results and a tool to make roads safer. This review examines the possible benefits and drawbacks of different machine learning and econometric models as they apply to the automotive safety context. The literature also reveals best practices for measuring and comparing model performance. The findings from this review will serve as the foundational methodology for the empirical analysis conducted in this study.

Research on automotive safety, which applies machine learning techniques, has returned positive results when comparing ensemble-based models to individual regression models. Ji and Levinson (2020) test different approaches to predicting injury severity, comparing the predictive accuracy of probit models to several different machine learning models. Their results show that a random forest model had the highest predictive accuracy, followed by the probit model. Similarly, Mamlook et al. (2020) conclude that random forest models have the highest predictive accuracy in the context of their investigation of risk factors contributing to crash severity amongst the elderly. Thus, I plan to compare the performance of a random forest approach to probit and logit-style regression models.

The literature on variable selection techniques and parameter fine-tuning yields a dichotomy of understandings. I plan to take an approach that combines best practices among the reviewed studies. Ji and Levinson ran multi-class models with seven different possible injury levels as the dependent variable. They found that while their model was competent in correctly classifying common cases (lower-level injuries), it was less so when working with less common cases (severe injuries). Alternatively, Mamlook et al. (2020) use an approach with a binary dependent variable, 0 indicating less severe injury and 1 indicating severe injury or death. While the distribution of the dependent injury severity variable for Mamlook et al. was heavily skewed towards less severe injuries, similar to Ji and Levinson's, their model did not perform worse when classifying the less common case. Like Mamlook et al., to work around the issue of poor classification results with less common cases, my model will use a binary dependent variable where 1 indicates fatality and 0 indicates no fatality.

Performance measurement is of the most contentious topics within the machine learning literature. A variety of different measures and techniques can be applied to assess model performance, the aptitude of which depends on the specific application and the size of the dataset. Mamlook et al. (2020) use accuracy, precision, recall, F1 score (harmonic mean of precision and recall), and receiver operating characteristics (ROC) to evaluate model performance. Additionally, they use an 80-20 train test split. Both Najada & Mahgoub (2016) and Ji & Levinson (2020) use a similar set of performance measurements. In a larger systematic review of machine learning model best practices, Raschka (2018) also recommends a 2-way holdout, or train test split, for larger datasets. Raschka also proposes using a McNemar test to statistically compare the performance of two classification models. According to most common practice, I plan to use accuracy, precision, recall, F1, and receiver operating characteristics (ROC) scores to assess model performance. Beyond that, I will also employ the McNemar test, as recommended by Raschka (2018).

Parameter selection, while important, is useless without proper data preprocessing and cleaning.  Ji and Levinson pull data from the Crashworthiness Data System (CDS) from the NHTSA. Like Ji and Levinson, I plan to use the NHTSA for crash data. For reliability reasons, I will pull from the Fatality Analysis Reporting System (FARS) instead of the CDS. Police-reported data on injury severity levels contained in both CDS and FARS can be subject to reporting biases and inconsistencies. Police reports on fatalities are much more reliable due to the binary nature of death. My research will subsequently use Fatality as a dependent variable, eliminating reporting bias and working around poor model accuracy with less common cases.

The findings in the literature relating to the efficacy of machine learning models as they apply to predicting automotive incident outcomes are overwhelmingly positive. Ji & Levinson (2020), Mamlook et al. (2020), and Najada & Mahgoub (2016) all find that random forest models outperform individual logistic or probit regressions when predicting accidents and injury severity. Additionally, each study concludes that random forests predict on-road outcomes with relatively high accuracy (over 80% in each case). The findings from the reviewed studies suggest that ensemble-based machine learning models, specifically the random forest approach, provide superior predictive performance compared to traditional individual econometric models in the context of automotive fatalities. Thus, I hypothesize that a random forest approach has the potential to predict automotive fatalities reliably.

## II. Data

### *A. Descriptive Statistics*

Table 5 shows descriptive statistics for 39,510 motor vehicle occupants, excluding motorcyclists, who were involved in accidents where there was at least one fatality. Similar to Chapter 1, Data was compiled from the Fatality Analysis Reporting Systems (FARS) 'Vehicle' and 'Person' tables from 2019. Entries with missing values were removed.

Column 1 reports the mean values for each variable, and column 2 reports the standard deviation. Most subjects (53.1%) in the sample died as a result of the crash. The mean injury level was 1.585, indicating that most subjects were either minorly injured or seriously injured. 67.6% of subjects wore seatbelts, while 32.84% of subjects either improperly wore a belt or did not wear one. 16.9% of subjects had either positive blood alcohol content (BAC), and 5.67% of drivers were subject to some level of drug involvement at the time of the incident. 17.5% of subjects were in speeding-involved accidents. The average age in the sample is 39 years old, a distribution of ages in the sample is shown via a histogram on the left of Figure 1.A. The average subject in the sample is about 5 feet 7 inches tall and weighs 184 pounds. The average speed limit where fatal accidents occur in the sample is around 50 miles an hour. 42.3% of the subjects were involved in an accident during rush hour time, and 69.9% were in an accident during the day, whereas 30.1% were in the dark. 23.1% of the subjects were involved in accidents in the cold. The average car model year amongst the subjects in the sample is 2008. Model years across the sample are shown on the right side of Figure 1 below. 87.9% of the subjects have a valid license, while 12.1% do not. A minority of subjects were in accidents where the airbag deployed (33.9%), while 66.1% of subjects did not have an airbag deploy. 59.6% of subjects were in car-to-car collisions. 68.9% of the subjects were drivers of motor vehicles, and 87% were in the front seat. 9% of subjects were ejected, and 9.6% were in larger-than-average vehicles.

**TABLE 5**

DESCRIPTIVE STATISTICS

| VARIABLES | Mean | SD | Min | Max |
|---|---|---|---|---|
| Fatality *(1 = Fatality)* | 0.531 | 0.499 | 0 | 1 |
| Restraint Use | 0.676 | 0.468 | 0 | 1 |
| Alcohol Use | 0.169 | 0.374 | 0 | 1 |
| Drug Use | 0.0567 | 0.231 | 0 | 1 |
| Speeding | 0.175 | 0.380 | 0 | 1 |
| Age *(Years)* | 39.90 | 20.68 | 0 | 105 |
| Sex (1 = male) | 0.647 | 0.478 | 0 | 1 |
| Height *(Inches)* | 68.30 | 3.975 | 37 | 82 |
| Weight *(lbs.)* | 184.1 | 47.05 | 58 | 550 |
| Speed Limit | 50.39 | 13.30 | 0 | 80 |
| Rush Hour Time | 0.423 | 0.494 | 0 | 1 |
| Daytime Lighting | 0.699 | 0.459 | 0 | 1 |
| Cold Weather *(Temp < 32 degrees)* | 0.231 | 0.422 | 0 | 1 |
| Model Year | 2,008 | 7.373 | 1930 | 2020 |
| License Status *(1 = valid)* | 0.879 | 0.326 | 0 | 1 |
| Airbag Deploy *(1 = yes)* | 0.339 | 0.473 | 0 | 1 |
| Driver | 0.689 | 0.463 | 0 | 1 |
| Front Seat | 0.870 | 0.336 | 0 | 1 |
| Car Collision | 0.596 | 0.491 | 0 | 1 |
| Ejection | 0.0908 | 0.287 | 0 | 1 |
| Large Size Car | 0.0966 | 0.295 | 0 | 1 |
| Observations | 39,510 | 39,510 | 39,510 | 39,510 |

**Table 5:** Table values are descriptive statistics for all drivers and passengers in vehicles involved in a crash in 2019. ‘Fatality,’ ‘Restraint use,’ ‘Alcohol use,’ ‘Drug Use,’ and ‘Speeding’ are all {0, 1} indicators where 1 indicates an affirmative value. 'Sex' is a dummy variable where 1 indicates male and 0 female. ‘Speed Limit’ is the speed limit measured in MPH at the spot of the crash. ‘Rush Hour Time’ indicates if a crash occurred during rush hour (1 for yes, 0 for no). ‘Daytime lighting’ is a {0, 1} indicator for whether the accident occurred in daylight (1 for yes). ‘Driver is a {0, 1} variable for whether the subject is the driver of a motor vehicle, 1 for yes. Front seat is a {0, 1} indicator for a subject being in either of the front seats, 1 for yes. ‘Car Collision’ indicates whether the incense was between two cars, 1 for yes. Ejection indicates whether the subject was ejected, 1 for yes. ‘Large Size Car’ indicates whether the car was an abnormal size, such as a truck or larger. Observations' represent the number of individual driver cases analyzed.

Figure 2 shows a correlation matrix among all the dependent variables. When implementing linear regression models, variables with high correlation coefficients can cause multicollinearity problems. Ensemble models, such as the random forest model used in this research, are less prone to multicollinearity issues in model training. Still, highly correlated variables can cause uncertainty when calculating variable importance scores. The correlation matrix shown in Figure 3 does not present any glaring issues, but several variables have correlation coefficients in-between 0.4 and 0.6 indicating moderate correlation. Height and Sex have a correlation coefficient of 0.412, Weight and Height have a correlation coefficient of 0.475, Light Condition and Rush Hour have a correlation coefficient of 0.563, and Front Seat and Driver have a correlation coefficient of 0.574. Moderate correlation may present problems, but since it is not extreme, meaning it does not exceed 0.6, the models estimates should be unbiased.

**FIGURE 3**
CORRELATION MATRIX

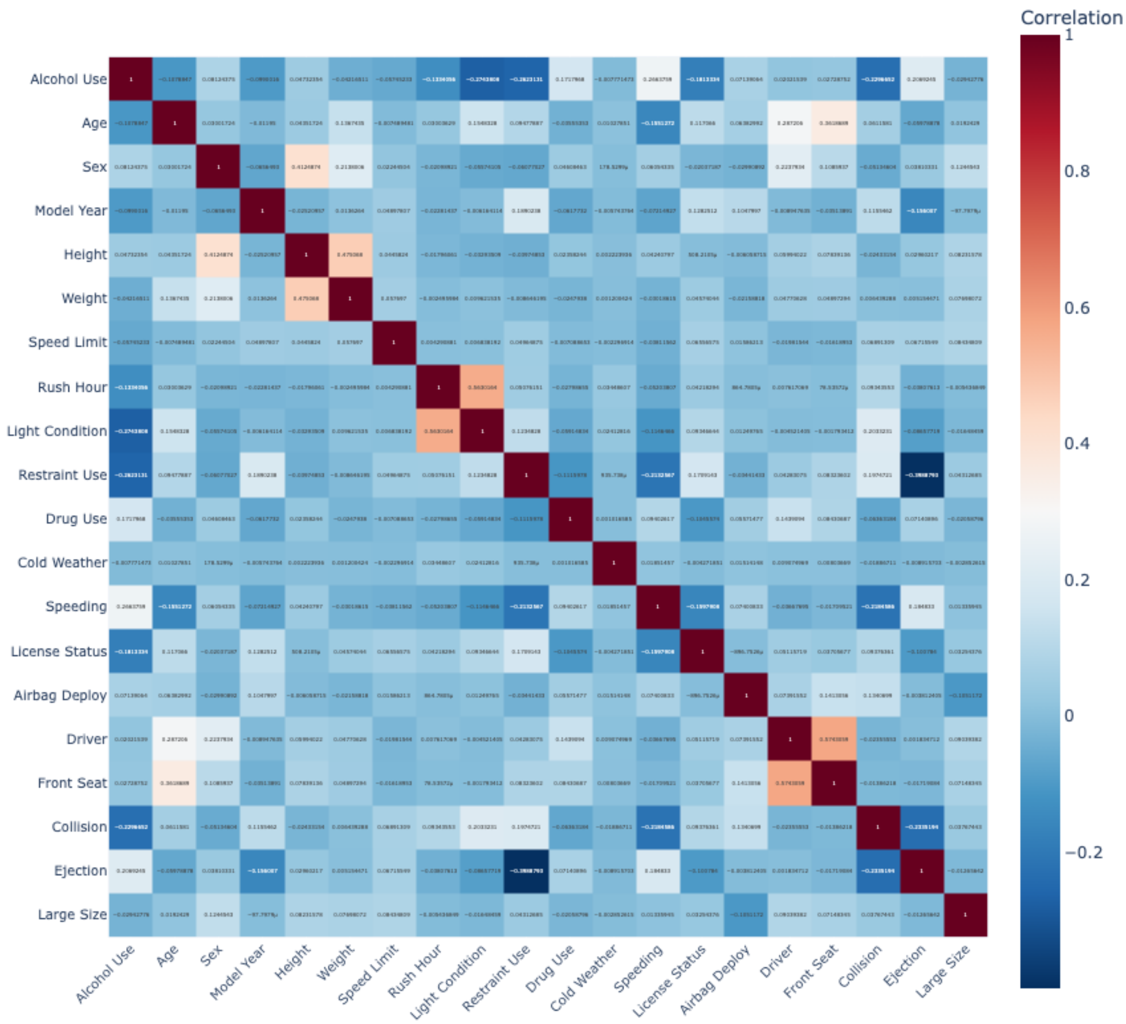



***Figure 3:*** *Correlation matrix for all independent variables as shown in Table 5*

### *B. Train-Test-Split*

The data must be split into training and validation datasets to train a machine learning model. The process of splitting data in this manner is commonly referred to as a "Train-Test-Split." First, the target, or the dependent variable, must be isolated from the features or the independent variables.

In my case, that means splitting the 'Fatality' data from the rest of the predictors. Once the target variable has been isolated, the resulting datasets will be divided into user-specified proportions via a random sampling process. A typical split ratio is 80% train 20% test, such as the one used in this paper. After the data has been split, the model can be trained on the training data and tested on the validation data. I use the scikit-learn machine learning library's built-in Train-Test-Split function in Python to implement this process, specifying an 80-20 ratio. Figure 4 shows a visualization of the Train-Test-Split workflow.

**FIGURE 4**
TRAIN TEST SPLIT

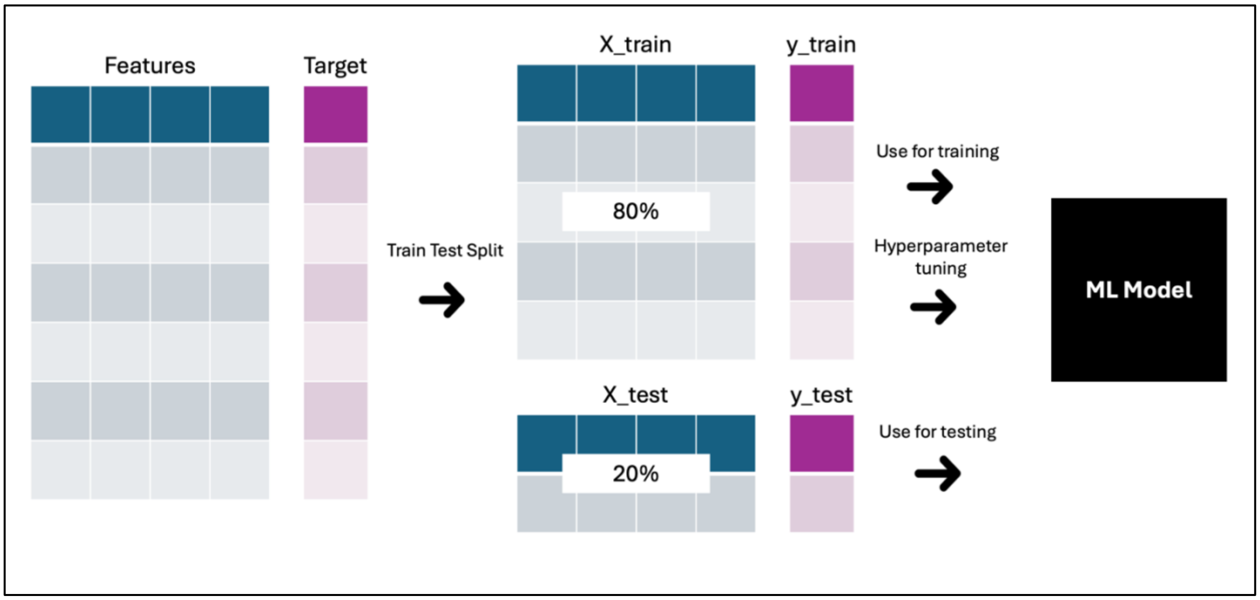



***Figure 4:*** *Train-Test Split-Workflow. This diagram illustrates the process of preparing a dataset for machine learning. The dataset is initially divided into 'Features' (in blue) and 'Target' (in purple), representing independent and dependent variables.*

## III. Results & Discussion

### *A. Overview*

To test and compare models, I first implement and optimize a random forest classification model utilizing scikit-learn, a Python machine learning library. I then compare the random forest model performance to the performance of logit and probit models from the statsmodels module in Python. Several metrics will be utilized to assess performance across models: accuracy, precision, recall, F1 score, ROC-AUC, and log loss (derivation documented in Table 4). In addition to performance metrics, confusion matrices will be utilized to compare model classification performance, as well as the McNemar test which will provide a statistical supplement for model comparison. Finally, feature importance metrics will be compared to coefficient estimates to contrast model interpretability.

### *B. Background*

The comparison of individual econometric models, such as the probit and logit, to an ensemble-based machine learning model, the random forest, is warranted due to underlying differences in assumptions that may affect classification performance. Before examining performance differences, one must understand how an ensemble-based machine learning model works and how that differs from an individual econometric model. The distinction between ensemble and individual is in the composition of the model: individual models are composed of a single mathematical function and set of assumptions, which are used to generate a set of estimates or predictions. Ensemble models, such as Brieman's random forest (2001), combine the predictions of multiple individual models.

**FIGURE 5**
DECISION TREE MODEL

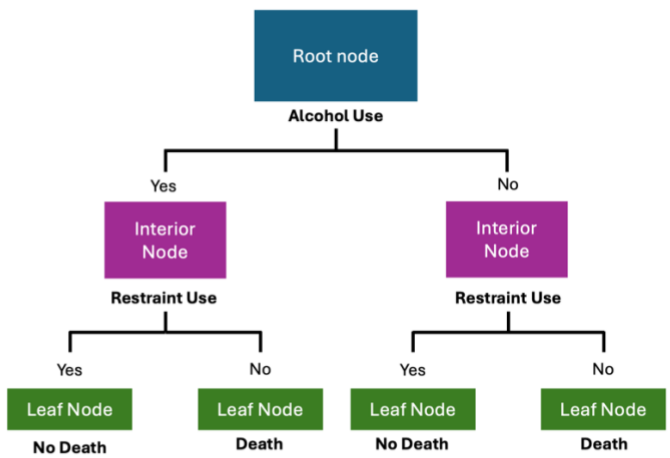



***Figure 5:*** *Decision Tree Model. A simplified visualization of how a decision tree generates predictions.*

The random forest is built from an ensemble of decision trees, from which predictions are aggregated. The decision tree is a simple classification model closely resembling the human decision-making process. Figure 5 illustrates a simple decision tree that predicts death or no death in an automotive accident. This simplified example does not represent what an entire tree would look like but offers a visualization for understanding. The root node will contain 100% of the data, with some entries being deaths and some being survival instances. Upon splitting by alcohol use, each interior node will contain a subset of the data, each node having a different distribution of deaths and no deaths (the node with positive examples of alcohol use will likely have more deaths). Another split then made on restraint use produces the leaf nodes, where the prediction will be made.

**FIGURE 6**
RANDOM FOREST MODEL

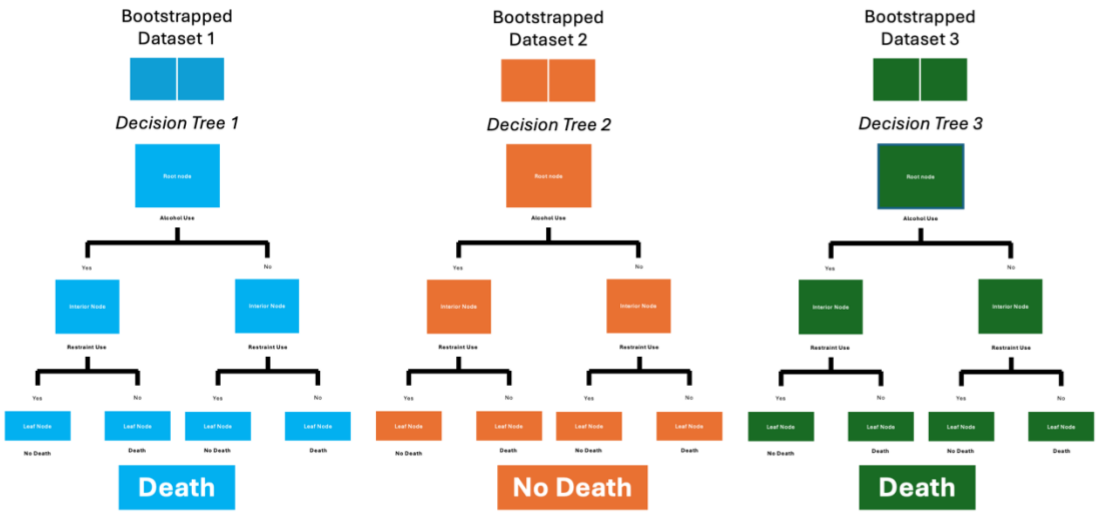



***Figure 6:*** *Random Forest Model. A simplified visualization of the random forest model*

In a random forest, each individual decision tree is created via a bagging (bootstrap-aggregating) process that creates bootstrap samples from the original data, makes a prediction from each sample, and averages the predictors to make a final prediction (Biau & Scornet, 2016). Each decision tree in the forest is constructed using a unique bootstrap sample, meaning the data for each tree is created using a random sampling process, which results in multiple datasets that statistically resemble the original. Additionally, the nodes, or features from which each tree splits, differ for each tree in the forest, reducing the correlation between the trees and hence reducing variance within the model. The splits in each tree are determined via the CART-split criterion (Classification And Regression Trees). CART uses Gini impurity (for classification problems) or prediction squared error (for regression problems) to determine the best split at the decision tree node. Gini impurity (the measure used in my implementation) measures how impure a tree is or how uniform in class a resulting group is from a split. More simply stated, the CART method (for classification problems) attempts to reduce impurity by creating splits, resulting in groups that comprise mostly the same class (Biau & Scornet, 2016).

Figure 5 illustrates a simple diagram of a random forest where three decision trees are trained using three distinct bootstrapped datasets. Each tree makes a prediction: death or no death, thus contributing a vote towards the final classification decision. The final prediction, in the case of a classification problem, will be based on a majority vote amongst the decision trees, in this case, yielding a prediction of death.

Unlike individual models such as the probit or logit, a random forest model makes no assumptions about the functional form of variable relationships, the distribution of errors, or interactions between variables. A probit model, for example, is constricted via a single equation, which attempts to fit a smooth line, following the shape of a cumulative normal distribution, through all of the data. This works well when the relationship between the independent and dependent variables has a particular functional form. However, variable relationships often follow unpredictable patterns, which probit models fail to account for by default. Alternatively, a decision tree functions by making a series of splits, where each split attempts to separate the data into pure or homogenous groups. A decision tree follows the functional form of the data rather than expecting the data to follow one pattern or fit into one equation. A random forest, which consists of the aggregated predictions of many decision trees, thus has the capability to handle complex non-linear relationships.

Additionally, individual models such as the probit model make an assumption about the distribution of errors. The probit model equation assumes that errors are distributed normally, indicating they follow a bell-shaped curve. Similarly, the logit model assumes that errors follow a logistic distribution. This works well when the errors are, in fact, distributed normally or logistically, but if there are outliers or the errors are non-normal, then the model will fail to return accurate predictions. A random forest model makes no such assumption of normality, meaning the model can handle any distribution of error within the data.

Random forest models are also better equipped to handle variable interaction effects. Each split within a decision tree includes a random subset of variables. The nature of multivariate splits means that variable interactions are inherently considered when multiple trees are aggregated, given all of the different combinations of feature pairs in splits across trees. Therefore, the ensemble as a whole can account for complex variable interactions without the need for explicit user identification beforehand. Conversely, a probit or logit model assumes by default that the variables do not have interactions, meaning that to capture interaction effects, the user must specify the interaction before running the regression.

Because random forests are able to better cope with complex variable relationships, non-normal error distributions, and complex interactions, they should return better classification results than probit and logit models. The next sections illustrate this theory and contain a comparison of classification results between the ensemble based random forest, and the individual probit and logit models.

**TABLE 6**
INDIVIDUAL REGRESSION OUTPUTS

| VARIABLES | Logit | Probit | Marginal Effects |
|---|---|---|---|
| Restraint Use | -1.223*** | -0.720*** | -0.206*** |
| | (0.0298) | (0.0175) | (0.00467) |
| Model Year | -0.0531*** | -0.0314*** | -0.00899*** |
| | (0.00178) | (0.00104) | (0.000287) |
| Age | 0.0240*** | 0.0141*** | 0.00404*** |
| | (0.000669) | (0.000393) | (0.000107) |
| Airbag Deploy | 1.072*** | 0.634*** | 0.181*** |
| | (0.0274) | (0.0161) | (0.00434) |
| Alcohol Use | 1.106*** | 0.639*** | 0.183*** |
| | (0.0414) | (0.0238) | (0.00663) |
| Speeding | 1.018*** | 0.588*** | 0.168*** |
| | (0.0385) | (0.0222) | (0.00619) |
| Constant | 107.9*** | 63.81*** | |
| | (3.581) | (2.092) | |
| Observations | 39,510 | 39,510 | 39,510 |

***Table 6:*** *logit and probit regression estimates with only a small subset of the variables. Standard errors in parentheses*
**** p<0.01, ** p<0.05, * p<0.1*

### *C. Model Training*

First, logit and probit models are implemented using the Train-Test-Split method as described in the descriptive statistics. The default models in the statsmodels Python module package lack a constant term. Without a constant term, the a logit or probit model could not accurately make predictions on any data for which the regression line does not cross the origin at (0, 0). Therefore, a constant term is added to the training and validation datasets utilizing the statsmodels ‘add_constant’ feature. With the added constant to the dataset, probit and logit models can be fit using the new training data. In their default states, probit and logit regression models estimate probabilities rather than make finite classifications, such as a random forest classifier, which returns 0 or 1 classification outcomes. Probabilistic predictions must be transformed into classifications to compare a linear regression to the random forest classifier. In this case, classifications are binary, with 1 indicating death and 0 indicating no death. A threshold of 0.5 is set when making predictions with the model to transform the model-predicted probabilities into classifications. This works such that for any predicted outcome above 0.5, a classification result of 1 is returned and vice versa. Table 6 shows a truncated output, for simplicity, of both logit and probit regressions (the full output is shown in Table 6.A).

Second, the random forest classification model is implemented using the scikit-learn 'RandomForestClassifier' package in Python. The random forest model assumes a set of default hyperparameters that can be optimized via several different automated methods. Hyperparameters are a set of user-specified model features that are not learned from the data. Grid search and random search are two of the most widely used methods to find the most optimal hyperparameter settings. Grid search is a technique that takes from a grid of predefined hyperparameters and tests every combination of values to find the most optimal one. Grid search is the most thorough method of hyperparameter tuning due to its exhaustive nature, but it is computationally expensive.

**TABLE 7**
RANDOM SEARCH RESULTS

| Hyperparameter | Best Value |
|---|---|
| bootstrap | TRUE |
| max_depth | 50 |
| min_samples_leaf | 1 |
| min_samples_split | 2 |
| n_estimators | 476 |
| Accuracy | 0.789 |

***Table 7:*** *Hyperparameter best value results from Random Search method on RF model.*

Random search is a less computation-intensive alternative that selects random combinations of hyperparameters until finding the best one. For comparison, in my case, the random search method with the parameters specified in Table 7 finished in 4m 42s, and the grid search failed to finish before 25m 7s. In my random search, I specify 'n_estimators,' 'max_depth,' 'min_samples_leaf,' 'min_samples_split,' and 'bootstrap,' meaning that the random search will look for the best combination of values for each. The random search will operate within a specified range; for example, I specify a range between 100 and 500 for 'n_estimators.' Random search has a parameter called 'n_inter,' which is set to 100 for this case, indicating that 100 different parameter combinations will be tested, where a cross-validation approach is used for each. Cross-validation splits the dataset into n folds (in this case, 5), where a model is trained on four folds and validated on the 5$^{th}$. This process is repeated for all folds, eventually culminating in an average performance measure across all of the folds. The random search process returns a value for each initially specified parameter, for which the resulting model can then be trained. Table 7 illustrates the returned hyperparameter values and the model's accuracy with cross-validation training.

### *D. Model Performance Assessment*

After the models have been trained and optimized, they can be evaluated using the metrics specified in Table 8. Accuracy is the most general measure of model performance, which takes the ratio of correct predictions to the total number of samples. Precision gauges the accuracy of positive predictions; that is, it measures the ratio of true positives (correct positive predictions) to all predicted positives, indicating a model's ability to minimize false positives. Recall, or sensitivity, assesses the true positive rate by calculating how well the model identifies positives out of the total actual positive cases.

**TABLE 8**
PERFORMANCE MEASURMENTS

| **Measure** | **Formula** |
|---|---|
| Accuracy | $\frac{\text{True Positives (TP)} + \text{True Negatives (TN)}}{N}$ |
| Precision | $\frac{\text{TP}}{\text{TP} + \text{False Positives (FP)}}$ |
| Recall/Sensitivity | $\frac{\text{TP}}{\text{TP} + \text{False Negatives (FN)}}$ |
| F1 Score | $2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$ |
| ROC-AUC | $\text{AUC} = \text{Probability}(\text{TPR} > \text{FPR})$ |
| Log Loss | $\text{Log Loss} = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log(p_i) + (1 - y_i)\log(1 - p_i)]$ |

***Table 8.*** *Equations for each performance measurement used in model assessment*

The F1 Score is a balanced combination of precision and recall, taking the harmonic mean of the two. The ROC-AUC score is comprised of two metrics: the Receiver Operating Characteristic Curve (ROC) and the Area Under the ROC Curve (AUC). The ROC is a plot of the true positive rate against the false positive rate. The area under the curve measures the model's performance across all classification thresholds. An ROC-AUC score of 1 represents perfect prediction performance, whereas 0.5 indicates random classification. The final measurement is log loss, which considers the confidence of predictions. The log loss metric highly penalizes predictions that are confident and wrong, providing insight into the model's certainty in its classifications.

All three models performed reasonably well under each metric. The random forest model outperformed the probit and logit models for each metric. The probit and logit models had almost identical performance, with the logit model slightly outperforming the probit. The performance of the two is so similar that for the rest of the performance comparisons, just the logit model will be referenced.

**TABLE 9**

PERFORMANCE MEASUREMENT (on validation data)

| | Accuracy | Precision | Recall | F1 Score | ROC AUC | Log Loss |
|---|---|---|---|---|---|---|
| Random Forest | 0.794 | 0.809 | 0.801 | 0.805 | 0.876 | 0.440 |
| Probit Model | 0.752 | 0.787 | 0.733 | 0.759 | 0.830 | 0.503 |
| Logit Model | 0.754 | 0.788 | 0.734 | 0.760 | 0.831 | 0.502 |

***Table 9:*** *performance measurement results for random forest, probit, and logit models.*

Table 9 shows the scores of the three models under six performance metrics. The random forest model made correct predictions on 79.4% of the sample, whereas the logit model made correct predictions 75.4% of the time. In terms of precision, the random forest is shown to make correct predictions on the positive class 80.9% of the time, while the logit model is correct on positive class predictions 78.8% of the time. For recall, a measure of how well the model identifies all positives out of all of the actual positives, the random forest has a score of 80.1%, beating the logit model (73.4%) by 6.7 percentage points. The F1 score captures the harmonic mean of precision and recall. The random forest model scored 0.805, whereas the logit model scored 0.760, indicating that the random forest has a better balance between precision and recall. The random forest model has an AUC of 0.876, which is higher than the logit model's AUC of 0.831, indicating superior distinguishing capability between positive and negative classes. Finally, the random forest model also returned superior performance under the log loss method of evaluation, which assesses a model's prediction certainty. In this case, a low measure is good; random forest returned a score of 0.440 compared to 0.502 for logit, indicating better predictive confidence. Overall, the random forest model implemented in this context comprehensively outperforms the two individual linear models.

Confusion matrices, as shown in Figure 7, offer a visualization of model performance, which can be helpful for comparison purposes. From the figure below, one can see that the random forest model has far fewer instances of false predictions when compared to the logit model on the right. The random forest returned a false negative 837 times (bottom left box), while the logit did so 1115 times. The random forest returned a false positive 794 times when the logit did so 830 (top left box). While this visualization can be useful, no statistical measure is provided to compare the two matrices.

**FIGURE 7**
CONFUSION MATRIX COMPARISON

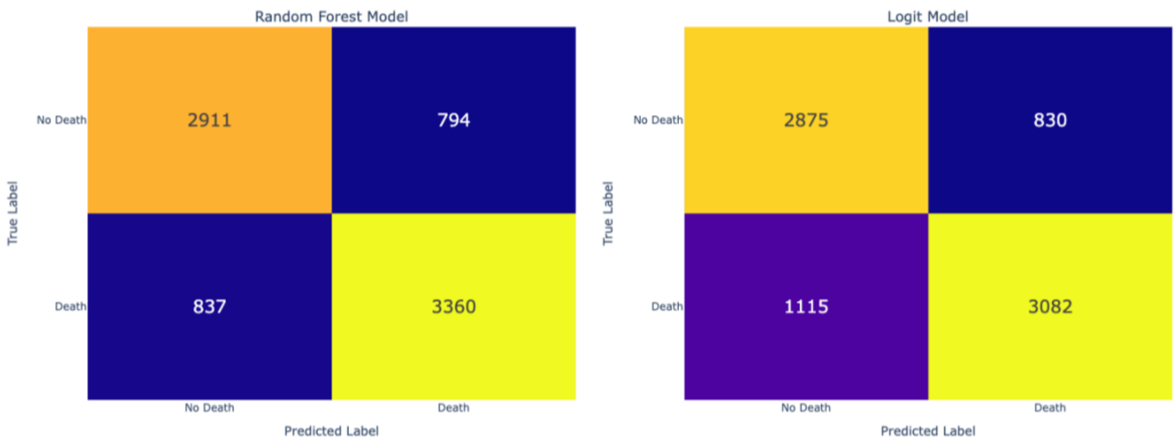



***Figure 7:*** *Confusion matrices comparing the performance of random forest and logit models in predicting mortality outcomes, with counts of true and false predictions for validation data.*

The McNemar test, as shown in Figure 8 offers a supplement to the traditional confusion matrix with a statistical measure of difference in terms of model performance. The top left corner and the bottom right corner show the number of classifications for which the random forest and probit models agree. The top left shows that there are 5557 instances in the validation data where both the models correctly predict a class. The bottom right cell represents 1231 instances where both models incorrectly make a classification. The other two cells represent areas of divergence or where the models disagree. The top right box shows that there were 714 instances of the random forest, making the correct classification where the probit was incorrect. The bottom left cell shows that there were far fewer instances (400) where the random forest was wrong, and the probit was correct. Results from the statistical test comparing these differences (equation 1.A) return a chi-squared statistic of 93.65, indicating significance at the 1% level ($p < 0.01$). This means that I cannot support the null hypothesis that the models perform the same, indicating the random forest model performs significantly better than the probit model (and, by proxy, the logit model).

**FIGURE 8**

MCNEMAR TEST CONTINGENCY TABLE

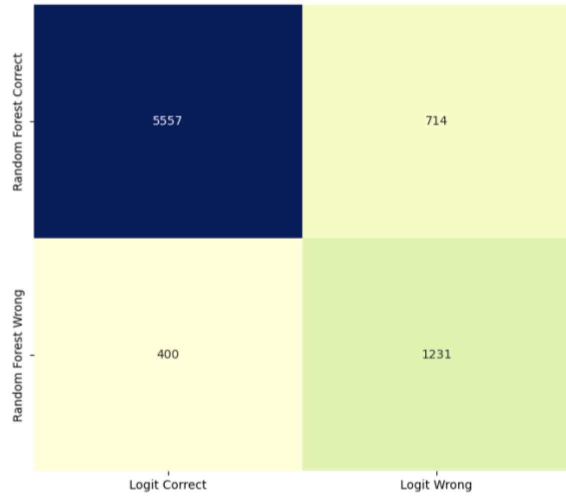



***Figure 8:*** *McNemar test contingency table (p <0.01)*

### *E. Inference*

Probit and logit models make for simple inference and interpretation. Take, for example, the coefficients on the probit marginal effects regression from Table 6. Each coefficient estimate represents a predicted percentage point change in the response variable. For instance, the coefficient on restraint use is -0.206 ($p < 0.01$), indicating that restraint use is associated with a 20.6 percentage point decrease in the likelihood of death conditional on being in a severe accident. Linear regressions also allow for significance calculation via a t-statistic. From this, one can deduce that the relationship between belt usage and death is statistically significant at the 1% level.

A weakness of ensemble-based machine learning models such as the random forest is their lack of interpretability. Many, such as Biau & Scornet (2016), say machine learning models function like a ‘black box’ where the combination of complex components such as bagging and CART-split criterion make for very difficult complex mathematical interpretation. Random forest models, however, provide feature importance measures, which can give more insight into feature relationships with the target. Unfortunately, feature importance scores are somewhat vague and can also be easily skewed by a number of biases.

**FIGURE 9**
RANDOM FOREST: MDA FEATURE IMPORTANCE

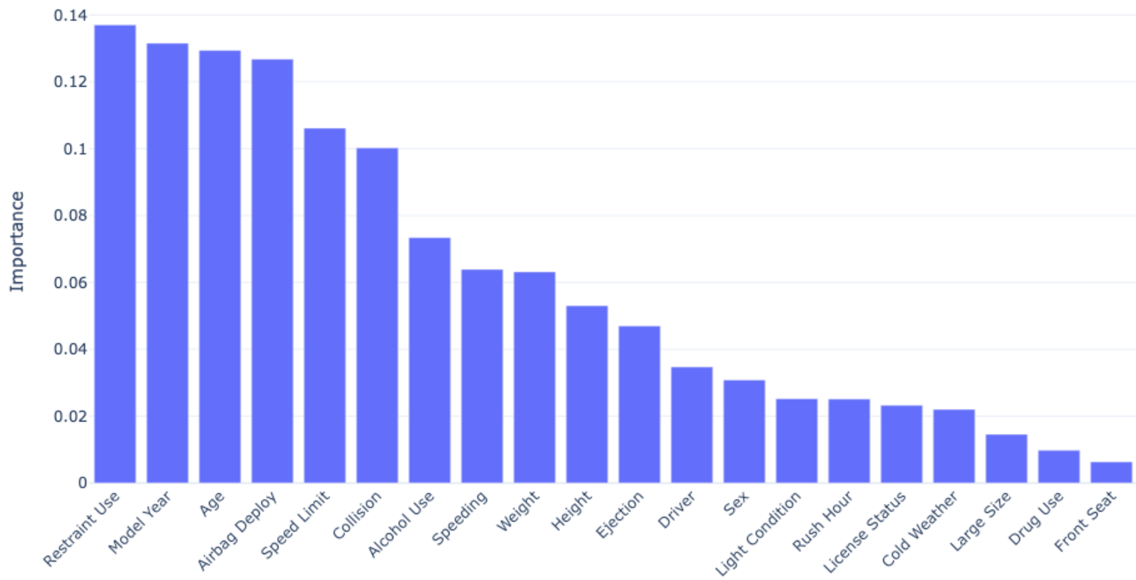



***Figure 9:*** *random forest feature importance diagram sorted in descending order using the MDI method of importance.*

There are two primary ways to calculate feature importance: mean decrease in impurity (MDA – Figure 9) and mean decrease in accuracy (MDI – Figure 9.A). MDA, the measure used here, is the less sensitive of the two to bias, such as class imbalance and variable type (categorical vs continuous). MDI feature importance is based on the Gini method of impurity, where MDA is derived from the total average loss in node impurity from splitting on a certain variable (Biau & Scornet, 2016). MDA is a more complex method of variable importance calculation based on the idea that shuffling the values of a variable in random permutations should not have an effect on its importance if the variable is truly important. Thus, the decrease in importance should be small for variables with true importance upon being reshuffled. The units for MDA (shown on the y-axis of Figure 8) represent the average accuracy decrease attributed to each feature's shuffling. MDA might be less sensitive to class imbalance or feature scale diversity due to separate evaluation of each feature: MDA examines each feature equally by assessing how its random shuffling affects model performance, independent of how often a class appears or the scale of measurement. MDI, however, can be biased towards overestimation for features that appear often because they have more opportunities to contribute towards impurity.

While MDA is a less sensitive, possibly more robust alternative to MDI, the measure still has inconsistencies, which make it unreliable. Additionally, MDA scores are made relative to all features and do not provide feature-specific measurements. Multicollinearity problems could muddy MDA results. If two features are correlated, the shuffling of one might not affect model accuracy, thus artificially inflating the importance of its related counterpart.

## IV. Conclusion

The random forest model's comprehensive outperformance of probit and logit models points to the superior classification potential of ensemble-based models. However, the comparative analysis in this study also reveals tradeoffs between predictive accuracy and interpretability between econometric and machine learning models.

In my research, I find that the random forest method provides a statistically significant improvement in predictive performance over individual models. The random forest model's superiority in handling complex, non-linear relationships is illustrated via the McNemar test, which confirms the model's significant performance benefits. However, while the random forest provides a highly confident classification tool, there are significant drawbacks when comparing it to individual models regarding interpretability. The random forest model has a 'black-box' nature, making results difficult to interpret. Tools such as feature importance scores allow for inference but are vague and suspectable to bias. For clear interpretation of results, individual models such as the probit or logit still demonstrate superiority, returning easy-to-interpret coefficients instead of vague feature importance scores.

While my work shows that machine learning approaches have efficacy in modeling traffic safety, I also reveal areas for further exploration in interpretability limitations. Future work would do well to take a deeper look into returning interpretable results in order to take advantage of the predictive power ensemble-based models present. Interpretable and accurate machine learning approaches to traffic safety problems have a wide range of life-saving and cost-reducing implementations. For example, insurance companies could utilize models such as the one presented in this paper to adjust and assess insurance premiums more accurately. Additionally, policymakers could adjust road infrastructure investments according to machine learning models to create safer road conditions. The lifesaving potential of ensemble-based models is clear and should be explored further.

# Chapter 4:
## Summary & Conclusion

In summary, I find no significant results pointing to the presence of risk compensation behavior as it pertains to automotive safety devices, specifically seatbelts. However, I do support the efficacy of seatbelts as a life-saving device and also uncover increased odds of death associated with risky behaviors such as alcohol and drug consumption. My lack of findings pertaining to risk compensation can be attributed to a small sample size and positive bias towards severe accidents within the data. Future work would benefit from a larger and more diverse dataset, from which findings related to compensation would be more apparent.

The second chapter of my work demonstrates the predictive power of ensemble-based machine learning models, such as the random forest, compared to individual econometric models when working with automotive fatality data. I find that the random forest is a stronger classification model compared to individual models such as the probit and logit. The superior performance of the random forest can likely be attributed to its aptitude for handling complex patterns and relationships within the data. My work stands as a testament to the benefits and a case for the further use of ensemble-based techniques for understanding automotive patterns

## References

Al Mamlook, R. E., Abdulhameed, T. Z., Hasan, R., Al-Shaikhli, H. I., Mohammed, I., Tabatabai, S., & 2020 IEEE International Conference on Electro Information Technology (EIT) Chicago, IL, USA 2020 July 31 - 2020 Aug. 1. (2020). 2020 ieee international conference on electro information technology (eit). In *Utilizing machine learning models to predict the car crash injury severity among elderly drivers* (pp. 105–111). essay, IEEE.

Al Najada, H., & Mahgoub, I. (2016). Big vehicular traffic Data mining: Towards accident and congestion prevention. In *2016 International Wireless Communications and Mobile Computing Conference (IWCMC)* (pp. 256-261). IEEE.

Biau, G., & Scornet, E. (2016). A random forest guided tour. *Test*, *25*, 197-227.

Breiman, L. (2001). Random forests. *Machine learning*, *45*, 5-32.

Bureau of Transportation Statistics (2023). Roadway Vehicle-Miles Traveled (VMT) and VMT per Lane-Mile by functional Class 2000-2021.

Carpenter, C. S., & Stehr, M. (2008). The Effects of Mandatory Seatbelt Laws on Seatbelt Use, Motor Vehicle Fatalities, and Crash-related Injuries Among Youths. *Journal of Health Economics*, 27(3), 642–62. https://doi.org/10.1016/j.jhealeco.2007.09.010.

Cohen, A., Einav, L., & John M. Olin Center for Law, Economics, and Business. (2001). *The effects of mandatory seat belt laws on driving behavior and traffic fatalities* (Ser. Discussion paper, no. 341). Harvard Law School, John M. Olin Center for Law, Economics, and Business.

Conner, K. A., Xiang, H., & Smith, G. A. (2010). The impact of a standard enforcement safety belt law on fatalities and hospital charges in Ohio. *Journal of Safety Research*, *41*(1), 17–23. https://doi.org/10.1016/j.jsr.2009.10.005.

Cullen, P., Möller, H., Woodward, M., Senserrick, T., Boufous, S., Rogers, K., ... & Ivers, R. (2021). Are there sex differences in crash and crash-related injury between men and women? A 13-year cohort study of young drivers in Australia. *SSM-population health*, *14*, 100816.

Doecke, S. D., Baldock, M. R. J., Kloeden, C. N., & Dutschke, J. K. (2020). Impact speed and the risk of serious injury in vehicle crashes. *Accident; analysis and prevention*, *144*, 105629. https://doi.org/10.1016/j.aap.2020.105629.

Drummer, O. H., Gerostamoulos, D., Di Rago, M., Woodford, N. W., Morris, C., Frederiksen, T., Jachno, K., & Wolfe, R. (2020). Odds of culpability associated with use of impairing drugs in injured drivers in victoria, australia. *Accident Analysis and Prevention*, *135*. https://doi.org/10.1016/j.aap.2019.105389.

Federal Highway Administration. U.S. Driving Last Year Was Lowest in Two Decades, New Data Show. *U.S. Department of Transportation.*

Forman, J., Poplin, G. S., Shaw, C. G., McMurry, T. L., Schmidt, K., Ash, J., & Sunnevang, C. (2019). Automobile injury trends in the contemporary fleet: Belted occupants in frontal collisions. *Traffic injury prevention*, *20*(6), 607-612.

Friedman L. S. (2012). Dose-response relationship between in-hospital mortality and alcohol following acute injury. Alcohol (Fayetteville, N.Y.), 46(8), 769–775. https://doi.org/10.1016/j.alcohol.2012.08.006.

Governors Highway Safety Association (2023). Seat Belts

Huelke, D. F., Gikas, P. W. (1968). Causes of Deaths in Automobile Accidents." American Medical Association. Pp. 1100-1107.

Ji, A., & Levinson, D. (2020). Injury severity prediction from two-vehicle crash mechanisms with machine learning and ensemble models. *Ieee Open Journal of Intelligent Transportation Systems*, *1*. https://doi.org/10.1109/OJITS.2020.3033523

Lardelli-Claret, P., de Dios Luna-del-Castillo, J., Jiménez-Mejías, E., Pulido-Manzanero, J., Barrio-Anta, G., García-Martín, M., & Jiménez-Moleón, J. J. (2011). Comparison of two methods to assess the effect of age and sex on the risk of car crashes. *Accident analysis & prevention*, *43*(4), 1555-1561.new lits

Levitt, S. D., & Porter, J. (2001). How Dangerous Are Drinking Drivers? *Journal of Political Economy*, *109*(6), 1198–1237. https://doi.org/10.1086/323281.

National Center for Statistics and Analysis. (2023). Fatality Analysis Reporting System (FARS) auxiliary datasets 1975-2021 (Report No. DOT HS 813 417). National Highway Traffic Safety Administration.

National Highway Traffic Safety Administration. Traffic Safety Facts. *U.S. Department of Transportation.*

National Highway Traffic Safety Administration (2019). Traffic Safety Facts: Seatbelt Use in 2019. *U.S. Department of Transportation.*

Notrica, D. M., Sayrs, L. W., Krishna, N., Rowe, D., Jaroszewski, D. E., & McMahon, L. E. (2020). The impact of state laws on motor vehicle fatality rates, 1999-2015. *The journal of trauma and acute care surgery*, *88*(6), 760–769. https://doi.org/10.1097/TA.0000000000002686

Peltzman, S. (1975). The Effects of Automobile Safety Regulation. *Journal of Political Economy*, *83*(4), 677–725. http://www.jstor.org/stable/1830396.

Raschka, S. (2018). Model evaluation, model selection, and algorithm selection in machine learning. *arXiv preprint arXiv:1811.12808*.

Robertson L. S. (1977) A Critical Analysis of Peltzman's "The Effects of Automobile Safety Regulation." *Journal of Economic Issues*, 11(3), 587-600. https://doi.org/10.1080/00213624.1977.11503463.

Scott, M. D., Buller, D. B., Andersen, P. A., Walkosz, B. J., Voeks, J. H., Dignan, M. B., & Cutter, G. R. (2007). Testing the risk compensation hypothesis for safety helmets in alpine skiing and snowboarding. *Injury prevention : journal of the International Society for Child and Adolescent Injury Prevention*, *13*(3), 173–177. https://doi-org.ccl.idm.oclc.org/10.1136/ip.2006.014142.

Territories (Report No. DOT HS 813 307). U.S. Department of Transportation.

United States Census Bureau (2023). Historical Income Tables: Households; Table H-8. Median Household Income by State.

Weisburd, S. (2015). IDENTIFYING MORAL HAZARD IN CAR INSURANCE CONTRACTS. *The Review of Economics and Statistics*, *97*(2), 301–313. http://www.jstor.org/stable/43556176.

# Appendix

**TABLE 3.A**
CONTROLS: DETERMINANTS OF DRIVER DEATH
(Y = 1 death, 0 no death)

| | (2)<br>Probit | (3)<br>Probit (marginal effects) |
|---|---|---|
| Driver age | 0.0144***<br>(0.000461) | 0.00413***<br>(0.000127) |
| Driver sex | -0.137***<br>(0.0233) | -0.0394***<br>(0.00669) |
| Driver height | -0.00340<br>(0.00282) | -0.000976<br>(0.000811) |
| Driver weight | 0.000338*<br>(0.000200) | 9.69e-05*<br>(5.74e-05) |
| Speed limit | 0.0110***<br>(0.000609) | 0.00316***<br>(0.000173) |
| Airbag deployment | 0.492***<br>(0.0173) | 0.141***<br>(0.00480) |
| Cold weather | -0.0420**<br>(0.0192) | -0.0121**<br>(0.00552) |
| Daytime lighting | 0.0573***<br>(0.0222) | 0.0165***<br>(0.00638) |
| License status | -0.165***<br>(0.0272) | -0.0474***<br>(0.00780) |
| Car model year | -0.0262***<br>(0.00111) | -0.00753***<br>(0.000312) |
| Rush hour time | -0.0400**<br>(0.0194) | -0.0115**<br>(0.00556) |
| Constant | 52.39***<br>(2.237) | |
| Observations | 32,225 | 32,225 |

**Note.** Estimates from Probit regressions on injury severity where the dummy variable is a 0 or 1 indicator where 1 indicates death. 'Driver sex' is a dummy variable where 1 indicates male and 0 female. Rush hour time' indicates if a crash occurred during rush hour (1 for yes, 0 for no). 'Daytime lighting' is a {0, 1} indicator for whether the accident occurred in daylight (1 for yes). 'Cold weather' is a {0, 1} indicator for whether the weather was cold (1 for yes). 'License status' is a {0, 1} indicator for whether the driver has a valid license (1 for yes). Continuous variables are measured as follows: 'Driver age' in years, 'Car model year' as the calendar year, 'Driver height' in inches, 'Driver weight' in pounds, and 'Speed limit' in miles per hour. 'Observations' represent the number of individual driver cases analyzed.
Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

**TABLE 6.A**

REGRESSION OUTPUTS

| VARIABLES | Logit | Probit | Marginal Effects |
|---|---|---|---|
| Restraint Use | -1.223*** | -0.720*** | -0.206*** |
| | (0.0298) | (0.0175) | (0.00467) |
| Alcohol Use | 1.106*** | 0.639*** | 0.183*** |
| | (0.0414) | (0.0238) | (0.00663) |
| Drug Use | 0.448*** | 0.250*** | 0.0716*** |
| | (0.0598) | (0.0344) | (0.00982) |
| Speeding | 1.018*** | 0.588*** | 0.168*** |
| | (0.0385) | (0.0222) | (0.00619) |
| Age | 0.0240*** | 0.0141*** | 0.00404*** |
| | (0.000669) | (0.000393) | (0.000107) |
| Sex | -0.175*** | -0.106*** | -0.0304*** |
| | (0.0289) | (0.0171) | (0.00489) |
| Height | -0.0245*** | -0.0145*** | -0.00415*** |
| | (0.00378) | (0.00223) | (0.000637) |
| Weight | 0.000558* | 0.000330* | 9.45e-05* |
| | (0.000302) | (0.000178) | (5.11e-05) |
| Speed Limit | 0.0224*** | 0.0132*** | 0.00379*** |
| | (0.000961) | (0.000559) | (0.000157) |
| Rush Hour Time | -0.0642** | -0.0394** | -0.0113** |
| | (0.0296) | (0.0176) | (0.00502) |
| Daytime Lighting | 0.160*** | 0.0958*** | 0.0274*** |
| | (0.0343) | (0.0202) | (0.00578) |
| Cold Weather | -0.0179 | -0.00919 | -0.00263 |
| | (0.0290) | (0.0171) | (0.00491) |
| Model Year | -0.0531*** | -0.0314*** | -0.00899*** |
| | (0.00178) | (0.00104) | (0.000287) |
| License Status | -0.336*** | -0.180*** | -0.0516*** |
| | (0.0418) | (0.0242) | (0.00691) |
| Airbag Deploy | 1.072*** | 0.634*** | 0.181*** |
| | (0.0274) | (0.0161) | (0.00434) |
| Driver | -0.537*** | -0.320*** | -0.0915*** |
| | (0.0334) | (0.0197) | (0.00560) |
| Front Seat | -0.556*** | -0.328*** | -0.0939*** |
| | (0.0470) | (0.0277) | (0.00789) |
| Car Collision | -0.542*** | -0.335*** | -0.0960*** |
| | (0.0272) | (0.0161) | (0.00453) |
| Ejection | 3.128*** | 1.547*** | 0.443*** |
| | (0.143) | (0.0599) | (0.0169) |
| Large Size Car | -0.427*** | -0.256*** | -0.0734*** |
| | (0.0444) | (0.0260) | (0.00741) |
| Constant | 107.9*** | 63.81*** | |
| | (3.581) | (2.092) | |
| Observations | 39,510 | 39,510 | 39,510 |

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

**EQUATION 1.A**
MCNEMAR TEST EQUATION AND TEST STATISTICS

$$\chi^2 = \left|\frac{(|B - C| - 1)^2}{B + C}\right|, \quad \chi^2 = 93.65, \quad p < 0.001$$

**FIGURE 1.A**
AGE & MODEL YEAR HISTOGRAM

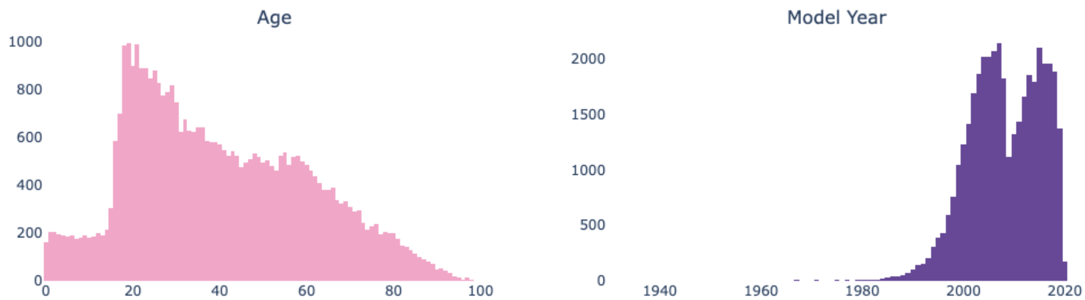



***Figure 1:*** *Histogram of age and car model year*

**FIGURE 2.A**
PROBIT MODEL CONFUSION MATRIX

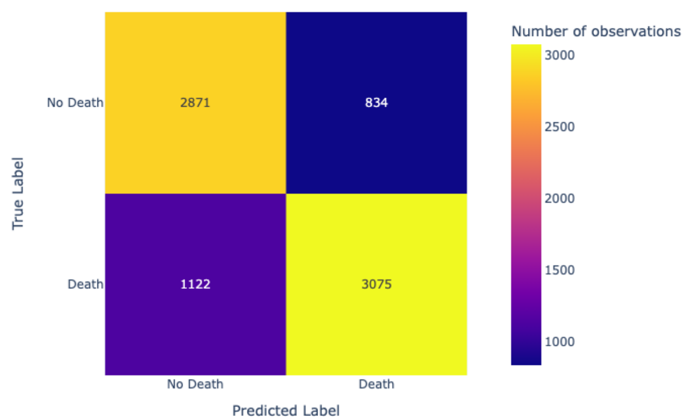


**FIGURE 9.A**
RANDOM FOREST MDI FEATURE IMPORTANCE