Final Project Report

An exploration of contemporary issues in highway safety, evolving transportation alternatives, and activity and travel behavior modelling

Prepared for **Teaching Old Models New Tricks (TOMNET) Transportation Center**

By,

Michael Maness
Assistant Professor
Email: manessm@usf.edu | ORCID: 0000-0001-5780-8666
Department of Civil and Environmental Engineering
University of South Florida, 4202 E. Fowler Avenue, Tampa, FL 33620

Fred Mannering
Professor, Executive Director
Email: flm@usf.edu | ORCID: 0000-0002-2803-4582
Department of Civil and Environmental Engineering
University of South Florida, 4202 E. Fowler Avenue, Tampa, FL 33620

With

Abdul Pinjari, Yu Zhang, Nawaf Alnawmasi, Natalia Barbour, Mouyid Islam, Chandra Bhat, Venky Shankar, Mohamed Abdel-Aty, Trang Luong, Divyamitra Mishra

December 2020

TECHNICAL REPORT DOCUMENTATION PAGE

1. Report No. N/A	**2. Government Accession No.** N/A	**3. Recipient's Catalog No.** N/A	
4. Title and Subtitle An exploration of contemporary issues in highway safety, evolving transportation alternatives, and activity and travel behavior modelling		**5. Report Date** December 2020 **6. Performing Organization Code** N/A	
7. Author(s) Michael Maness, https://orcid.org/0000-0001-5780-8666 Fred Mannering, https://orcid.org/0000-0002-2803-4582 Abdul Pinjari, https://orcid.org/0000-0002-3056-4259 Yu Zhang, Nawaf Alnawmasi, https://orcid.org/0000-0001-5753-3025 Natalia Barbour, https://orcid.org/0000-0002-0787-3993 Mouyid Islam, https://orcid.org/0000-0001-9068-9187 Chandra Bhat, https://orcid.org/0000-0002-0715-8121 Venky Shankar, https://orcid.org/0000-0002-1771-9957 Mohamed Abdel-Aty, https://orcid.org/0000-0002-4838-1573 Trang Luong, https://orcid.org/0000-0002-4233-1532 Divyamitra Mishra, https://orcid.org/0000-0002-9192-9936		**8. Performing Organization Report No.** N/A	
9. Performing Organization Name and Address University of South Florida, 4202 E. Fowler Avenue, Tampa, FL 33620		**10. Work Unit No. (TRAIS)** N/A **11. Contract or Grant No.** 69A3551747116	
12. Sponsoring Agency Name and Address U.S. Department of Transportation, University Transportation Centers Program, 1200 New Jersey Ave, SE, Washington, DC 20590		**13. Type of Report and Period Covered** Research Report (9/2019 – 8/2020) **14. Sponsoring Agency Code** USDOT OST-R	
15. Supplementary Notes N/A			
16. Abstract This report addresses a multitude of contemporary issues in highway safety, evolving transportation alternatives, and activity and travel behavior modelling. The report begins by studying issues relating to big data, traditional data and the tradeoffs between prediction and causality in highway-safety analysis, showing that a combination machine learning methods and advanced statistical methods are the most promising directions for future research. The report then moves on to study ridehailing-usage rate where a multitude of factors were found to influence usage including self-assessed health, high body mass index, and registration for other shared mobility services were all found to play roles in ridehailing usage. The report then studies the emerging phenomena of carsharing and specifically the renting of personal vehicles, identifying important factors that influence the success of such programs. The report then presents a study of aggressive driving and a study of the factors affecting work zone safety to highway work zone safety, showing the importance of capturing unobserved effects and considering the temporal stability of model-estimated parameters. The report then shifts to the study of electric vehicles and an assessment of the zero-price effect to estimate a monetary value of free charging. Next, bikesharing behavior during holidays with one of the findings being that federal holidays negatively affect member ridership and positively affect non-member ridership. Finally, social capital is explored in relation to leisure activity behavior. By addressing critical contemporary modeling issues, this report provides practical insights as well as serving as a basis for future research.			
17. Key Words Econometric Modeling, Ride-Hailing, Bike Sharing, Highway Safety, Social Capital, Renting of Personal Vehicles		**18. Distribution Statement** No restrictions.	
19. Security Classif.(of this report) Unclassified	**20. Security Classif.(of this page)** Unclassified	**21. No. of Pages** 296	**22. Price** N/A

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

DISCLAIMER

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated in the interest of information exchange. The report is funded, partially or entirely, by a grant from the U.S. Department of Transportation's University Transportation Centers Program. However, the U.S. Government assumes no liability for the contents or use thereof.

ACKNOWLEDGMENTS

This study was funded by a grant from A USDOT Tier 1 University Transportation Center, supported by USDOT through the University Transportation Centers program. The authors would like to thank the TOMNET, USDOT, and University of South Florida for their support of university-based research in transportation, and especially for the funding provided in support of this project.

TABLE OF CONTENTS

List of Tables

List of Figures

EXECUTIVE SUMMARY

This report addresses a multitude of contemporary issues in highway safety, evolving transportation alternatives, and activity and travel behavior modelling. The report begins by studying issues relating to big data, traditional data and the tradeoffs between prediction and causality in highway-safety analysis. The discussion provided shows that the nature of the data, and the implementation target of the analysis means that analysts must often tradeoff the predictive capability of the resulting analysis and its ability to uncover the underlying causal nature of crash-contributing factors when using traditional models, data-driven models, heterogeneity models, or causal inference models. It is indicated that a combination machine learning methods and advanced statistical methods are the most promising directions for future research. The report then moves on to study ridehailing-usage rates where a multitude of factors were found to influence usage including self-assessed health, high body mass index, and registration for other shared mobility services were all found to play roles in ridehailing usage. The finding show that that ridehailing usage tends to be driven by a wide variety of individual characteristics and lifestyle choices.

The report then studies the emerging phenomena of carsharing and specifically the renting of personal vehicles, identifying important factors that influence the success of such programs, which suggest that supply-side determinants (determining who would be willing to rent their vehicles) are critical to future carsharing success. The report then presents a study of aggressive driving and a study of the factors affecting work-zone safety to highway work zone safety. The finding of this portion of the report underscore the importance of capturing unobserved effects (which would include attitudes toward safety, and other unobserved factors that would influence injury severities). In addition, the findings show the importance of explicitly considering the temporal stability of model-estimated parameters.

The report then shifts to the study of electric vehicles and an assessment of the zero-price effect to estimate a monetary value of free charging. It is argued in the findings that for early analysis of free charging policies and pricing structures, the cost-effectiveness of the policy should be a greater focus than the exact structure of the policy, so a looser mean-focus and distribution-focused approach should bring greater value. Next, the report shifts to bikesharing behavior during holidays with one of the findings being that federal holidays negatively affect bikesharing member ridership and positively affect non-bikesharing member ridership. The findings on bikeshare ridership patterns during special calendar days have implications for the management of bikeshare systems, local economies, and public health. Because of increased non-member ridership on holidays, municipalities and bikeshare systems can concentrate information and advertising campaigns around non-users on holidays. Finally, social capital is explored in relation to leisure activity behavior. The findings contribute to a growing interest of considering the effects of social network characteristics on activity-travel behavior. As social capital has distinct impacts even among homogeneous groups, transportation modelers can derive more refined characteristics from social capital measures to build more socially and behaviorally realistic models.

By addressing critical contemporary modeling issues, this report provides practical insights into several emerging modeling issues in the transportation field. Insights that can form the basis for effective transportation policies.

Chapter 1

Introduction

Several critical issues have emerged in recent years in the fields of highway safety, alternative transportation modes, and activity and travel behavior modeling. Regarding highway safety, there is currently an ongoing methodological debate about the use of data-driven methods (machine learning, etc.), conventional statistics, statistical models that address unobserved heterogeneity, and causality models. This debate has obvious transferability to other fields such as activity and travel behavior modeling. Next, mobility on demand (ride hailing), bikesharing, electric vehicles, and person-to-person carsharing have become disruptive behaviors in travel demand, and this report will provide studies of these behaviors as well. Returning to matters of highway safety, there have been several recent studies that indicate that driver behavior is changing continuously over time in response changing vehicle technologies, changing behavior and utilization of social media and texting as well as other temporally shifting factors. This has profound implications for highway safety and the development of safety policies and countermeasures. The intent of the safety portion of this study is to explore the temporal instability of driver behavior from various perspectives including the possible temporally shifting effects of aggressive driving and driving behavior in work zones, two elements of driver behavior that are believed to be highly unstable over time. Statistical evidence of possible changes in the effects of these elements over time can help guide public policy and effect mitigation. Finally, this report offers a social capital theory explanation for variety seeking in leisure activity behavior.

The project report begins by studying issues relating to big data, traditional data and the tradeoffs between prediction and causality in highway-safety analysis (Chapter 2). The analysis of highway accident data is largely dominated by traditional statistical methods (standard regression-based approaches), advanced statistical methods (such as models that account for unobserved heterogeneity), and data-driven methods (artificial intelligence, neural networks, machine learning, and so on). These methods have been applied mostly using data from observed crashes, but this can create a problem in uncovering causality since individuals that are inherently riskier than the population as a whole may be over-represented in the data. In addition, when and where individuals choose to drive could affect data analyses that use real-time data since the population of observed drivers could change over time. This issue, the nature of the data, and the implementation target of the analysis imply that analysts must often tradeoff the predictive capability of the resulting analysis and its ability to uncover the underlying causal nature of crash-contributing factors. The selection of the data-analysis method is often made without full consideration of this tradeoff, even though there are potentially important implications for the development of safety countermeasures and policies. This chapter provides a discussion of the issues involved in this tradeoff with regard to specific methodological alternatives and presents researchers with a better understanding of the trade-offs often being inherently made in their analysis.

Chapter 3 considers various issues related to mobility on demand. The recent growth in the popularity of mobility-on-demand (ridehailing) has substantially disrupted the transportation market by providing a variety of new transportation options. While new mobility-on-demand options have significantly impacted some traditional transportation services (such as taxis), the factors that determine usage rates of new ridehailing options are not fully understood. The intent of the current report is to develop a statistical model of individuals' usage rates of ridehailing services. Using a sample of recently collected data, a mixed logit model (multinomial logit model with random parameters) of ridehailing-usage rate was estimated and, in addition to traditional socio-demographic factors, several travel and health-related variables were found to play statistically significant roles for ridehailing usage. Specifically, age, gender, income, household size, vehicle ownership, typical parking time, and the nature of commutes were some of the significant variables found in model estimation results. In addition, self-assessed health, high body mass index (BMI), and registration for other shared mobility services were all found to play roles in ridehailing usage. The results suggest that ridehailing usage tends to be driven by a wide variety of individual characteristics and lifestyle choices.

Chapter 4 looks at the emerging phenomena of carsharing and specifically the renting of personal vehicles. The renting of personal vehicles for monetary compensation (peer-to-peer carsharing or abbreviated as P2P carsharing) has become increasingly popular in the U.S. In applications, the fleet of peer-to-peer carsharing vehicles typically consists of personally owned vehicles identified and coordinated by a third-party company. However, little is known about the attitudes, perceptions, and decision process through which individuals decide to offer their car for rent in such peer-to-peer carsharing. To explore individuals' attitudes and perceptions regarding the act of supplying a personal vehicle to peer-to-peer vehicle fleet, a stated preference survey was designed and disseminated between February and April of 2018 where survey respondents were asked how likely they would be to rent their car (extremely unlikely, unlikely, unsure, likely, extremely likely). The survey questionnaire collected detailed socio-demographic information, as well as data on travel behavior and travel patterns. These data were then used to estimate a random parameters ordered probit model of their likelihood of renting their car. Some of the variables found statistically significant determinants of the willingness to rent a personal vehicle were gender, age, income, household composition, vehicle ownership, living location with respect to a grocery store, and participation in other shared mobility modes. The above findings and especially the gender and income related variables were found to complement prior literature and offered additional layer of understanding of the factors determining the supply side of peer-to-peer carsharing. The findings of this study offer some initial insights into the factors that may determine the success or failure of this novel transportation alternative.

Chapter 5 turns to the area of safety and specifically modeling issues relating to aggressive driving. Aggressive driving has become a national traffic-safety concern, with increasing congestion and other stress-inducing factors making it more likely drivers take out their frustrations by driving aggressively. Looking at single-vehicle crashes, this study investigates differences between resulting crash-injury severities when aggressive and non-aggressive driving behavior is observed, and how these differences changed over time. Using three years of crash data from 2015 to 2017 (inclusive), random parameters multinomial logit models with unobserved heterogeneity in means and variances were estimated. The available crash data included a wide variety of factors known to influence driver-injury severity including data related to the crash, vehicle, driver, spatial and temporal characteristics, roadway attributes, and traffic volume. Model estimates show that there were significant differences in driver-injury severities resulting from aggressive and non-aggressive driving, and that the effect of factors that determine injury severities changed significantly over time (statistically significant temporal instability). However, it is noteworthy that crashes involving non-aggressive drivers had many explanatory variables that produced temporally stable marginal effects, whereas crashes involving aggressive drivers had only one such variable (restraint belt usage). This suggests the possibility that temporal instability found in many recent safety studies may be driven by a subset of the crash population, and that there may be temporal stability in many crashes. Exploring this possibility is a promising direction for future empirical investigation.

Chapter 6 continues the safety emphasis by looking at issues relating to work zone safety, a critical issue with likely nationwide infrastructure initiatives. In the state of Florida, work-zone related crashes and their resulting injury severities have been increasing recently, particularly over the 2015 to 2017 time period. In the current study, we seek to provide insights into the factors that have been influencing this trend. Using work zone data from the 2012 to 2017 time period, resulting driver-injury severities in single-vehicle work zone crashes were studied using random parameters logit models that allow for possible heterogeneity in the means and variances of parameter estimates. The available data included a wide variety of factors known to influence driver injury severity including data related to the crash characteristics, vehicle characteristics, roadway attributes, prevailing traffic volume, driver characteristics, and spatial and temporal characteristics. The model estimates produced significantly different parameters for each of the year from 2012 to 2017, and a fundamental shift in unobserved heterogeneity, suggesting statistically significant temporal instability. In addition, in several key instances, the marginal effects of individual parameter estimates show marked differences between one year and the next.

However, these findings may not be the sole result of variations in driver behavior over time as has been argued in past research that has found temporal instability. This is because each work zone has a unique set of characteristics and, with the sample of work zones changing from one year to the next as highway maintenance and construction is undertaken in different locations, this work-zone sample variation could be a substantial source of the observed temporal instability.

Chapter 7 shifts to a behavioral analysis of the zero-price effect phenomenon. Over the past decade, electric vehicles have become a viable alternative to standard combustion engine vehicles. Prior research has shown that a short-term free public charging program could possibly increase plug-in electric vehicle sales, decrease oil consumption, and decrease greenhouse gas emissions. To deepen the understanding of consumer behavior relating to free charging, this research aims to analyze the zero-price effect to estimate a monetary value of free charging. To arrive at accurate estimation, data from stated preference survey were used to estimate latent class models of attribute non-attendance. The values calculated via different computations methods were then compared. The national mean zero-price effect for public charging ranged from $0.95 to $1.40 across the models. Because the collected sample was correctly weighted and national representativeness was achieved, the findings from this work can help to assess policies which offer free public charging infrastructure. This work can give more insights into how much value drivers may place on zero-cost vehicle charging.

Chapter 8 provides an analysis of bikeshare behavior during holidays. Bikeshare provides important first mile last mile, commuting, circulation, and sightseeing options in many cities. Bikeshare can also be healthy and convenient for users. Throughout the year, holidays occur which change typical bikeshare activity patterns. Existing literature shows mixed results relating to the ridership impacts of holidays, as some research shows that these days may result in higher ridership, while others show no effect. Because of variations in system locations and modeling methods, it is difficult to determine the reasons for these mixed results. To control for these aspects, this project consists of a multi-city study of the effect of holidays on system-level ridership using a log-linear regression model with robust standard errors. The results show the impacts of holidays on bikeshare system ridership for different user types among systems in the Washington D.C., Chicago, Boston, Los Angeles, and Minneapolis metro areas. Several hypotheses are built and tested for examining the expected effects of holidays on bikeshare usage. A major finding from this study is that federal holidays negatively affect member ridership and positively affect non-member ridership. It was also found that different federal holidays have dissimilar effects on total ridership. These findings could be useful for bikeshare agencies to plan, reposition fleets, and improve system operation.

In Chapter 9, social capital is explored in relation to leisure activity behavior. Transportation research has paid deeper attention to exploring factors affecting leisure activities and their induced travel. Motivated by the influence of social capital on leisure activity behavior, this report proposes a theory that leisure activity variety is an instrumental outcome and thus mostly affected by instrumental social resources. The theory underlines two hypotheses that 1) social capital is an integral determinant of leisure activity participation, and 2) having access to instrumental social support promotes instrumental outcomes demonstrated by increased leisure activity variety. This theory was comprehensively tested on the number of different unique leisure activities collected from

1,297 survey respondents. To the authors' knowledge, this refined and specially designed survey is the first in the transportation literature to use both position generator and resource generator to measure social capital. Results from negative binomial regression models demonstrated that instrumental support indeed had the largest influence on predicting activity variety outcome. This study's findings helped to reduce the biases and unobserved heterogeneity across various socioeconomic attributes. As social capital has distinct impacts even among homogeneous groups, transportation modelers can derive insights from social capital measures to build more socially and behaviorally realistic models.

Chapter 2

Big Data, Traditional Data and the Tradeoffs between Prediction and Causality in Highway-Safety Analysis

Fred Mannering, Chandra Bhat, Venky Shankar, Mohamed Abdel-Aty

2.1. Introduction

The implicit assumption in traditional statistical analyses is that an appropriately estimated model will both uncover causal effects and have the highest possible prediction accuracy. But the recent development and application of data-driven methods, as well as issues of causality in traditional statistical modeling, suggest that safety analysts must often, even if not always, make a trade-off between prediction accuracy and uncovering underlying causality. That is, models that predict well may not be the best at uncovering causality, and models that are good at uncovering causality may not be the best for practical prediction purposes.

There are four general methodological approaches that are potentially suitable for the analysis of transportation safety data: traditional statistical models, endogeneity/heterogeneity models, data driven methods, and causal inference models.[1] Each of these models have an implicit trade-off between practical prediction accuracy and their ability to uncover underlying causality. Traditional statistical models, such as those in the Highway Safety Manual (AASHTO, 2010), use conventional statistical methods with limited data (data that is readily available to most safety practitioners) to predict the effect

[1] Causal inference models have become a key analytic approach in the economics field and have been gaining in interest among transportation researchers. However, the complexities of applying the approach in the complex behavioral arena of transportation-related decision making are an ongoing concern (Brathwaite and Walker, 2018).

of various safety improvements on accident risk. The traditional literature (such as that supporting the Highway Safety Manual) claims predictive capabilities and causal explanations, but generally lacks fundamental support for these claims via assessments of parameter bias (for example, potential biases in parameter estimates and estimates of standard errors). Predictive capabilities of traditional highway-safety models are typically based on assessment of aggregate counts (total count of accidents for example), and there is scant support for true tests of predictability (such as tracking observational predictions against observed counts several years ahead of the estimated models). In fact, claims of predictive ability in many traditional models are limited in credibility, in large part due to temporal instability in parameters (Mannering 2018). Similarly, claims about causal ability in the traditional safety literature are limited because the true range of influential factors on accident likelihoods is unknown. Missing data problems, problems of consistency of measurement, and variation in unobserved effects due to economic, socio-demographic and vehicle characteristics amplify the potential bias in estimation.

To address some of the limitations above, endogeneity models (see Bhat et al., 2014) and heterogeneity models (see Mannering et al., 2016 for a thorough review) have been developed to extend traditional safety models by using advanced statistical and econometric methods. Endogeneity models account for the potential endogeneity of a safety-related variable when attempting to extract the "true" causal effect of the variable on a primary safety outcome variable of interest, after accommodating "spurious" associative effects or correlation effects between the variables. Unobserved heterogeneity models control for unobserved factors that may influence the likelihood and resulting injury severities in accidents. Endogeneity models and heterogeneity models are stylized, in that they are based on relatively limited datasets where the range of the potential endogenous and explanatory variables is much larger than widely available transportation highway data. A richer set of variables can potentially improve predictive capability and understanding of causality; however, the increased model complexity creates an additional burden on model transferability and predictive validation. Model complexity also poses challenges in estimation due to computational constraints. Estimation of highly complex endogeneity models and heterogeneity models involves simulation-based methods or analytic approximation methods due to the numerical integration needed to capture unobserved effects. While there has been substantial progress in such methods in the recent past (see, for example, Bhat, 2018), the required estimation techniques can still present dimensionality challenges for large accident datasets.

Data-driven methods include a wide range of techniques including those relating to data mining, artificial intelligence, machine learning, neural networks, support vector machines, and others. Such methods have the potential to handle extremely large amounts of data and provide a high level of prediction accuracy. On the downside, such methods may not necessarily provide insights into underlying causality (truly understanding the causal effects of specific factors on accident likelihoods and their resulting injury probabilities).[2]

Finally, causal-inference models explicitly recognize that accidents are only observed for a portion of the driving population and that this can lead to erroneous interpretations of findings (more on this below). Causal-inference models have rarely been

[2] Some insight into the influence of specific variables in data-driven methods can made through simulation and calculating factors such as Gini Index, but this may not necessarily provide insight into underlying causality.

applied in the accident analysis literature, but such approaches in other fields base these models on time series data to identify causal effects. However, causal-inference models have weak predictive capabilities because, among other reasons, they typically are not based on individual-accident level data and thus address a limited number of explanatory variables. Besides, the time-series nature of these models, while supposedly providing more basis for inferring causality, raises additional issues about the possible presence of uncontrolled factors that change during the intervening periods of time thereby potentially tainting the presence and estimated extent of causation.

Figure 2.1 presents a graphic of the trade-offs associated with these methods regarding predictive capability, causal inference capability and big data suitability (the ability of the methods to address problems that involve large amounts of data.) The choice of one method over another often involves several important considerations that go beyond a simple tradeoff between prediction and causality. Each of these four methods (data-driven versus causal versus traditional versus endogeneity/heterogeneity models) involve different levels of data. In addition, the application of the model (modeling purpose) needs to be considered as well. For example, endogeneity models and heterogeneity models would seem to be superior to traditional models in both prediction and causality; however these models typically use highly detailed datasets, and the models are complex in their application. In contrast, traditional safety models have relatively modest data requirements that are easy to apply, but their utility comes at the expense of a loss of predictive capability and lack of insight into causal influences (with the added risk of biased inference).

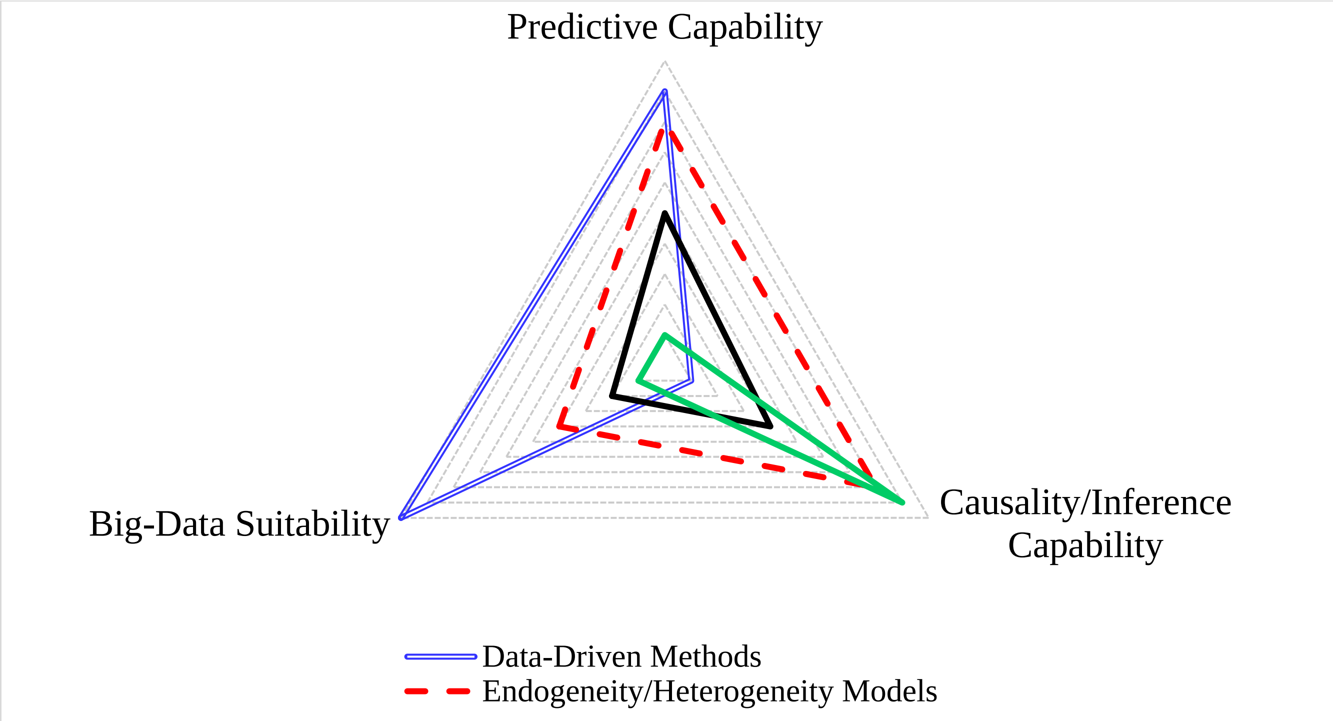


Figure 2.1. Current modeling trade-offs between big-data suitability, predictive capability, and causality/inference capability.

With extremely large datasets (big data) such as those that might be available in naturalistic driving studies, traditional models, advanced endogeneity/heterogeneity models, and causal effect models can be challenging to estimate, often making data-driven methods the preferred approach. In fact, data-driven methods can cover a wide range of data sizes, but, with smaller data sizes, the advantages of other methods to uncover causality tend to be preferred among analysts. Also, data driven methods may not be adequately complemented with domain expertise, resulting in inference driven primarily by statistical reasoning. The advent of artificial intelligence (AI) methods and the explosive growth of AI potentially opens the door for introducing some level of "automated" domain expertise to fine-tune data driven models that are developed strictly by statistical reasoning. But, at the end, human judgement and domain expertise are still likely to be needed in some form, especially in the context of driving the formulation of models of causal inference, since directional relationships between variables are formulated based on apriori knowledge of influential factors. As an example, in big data problems studying the impact of factors affecting fatality likelihoods, sample size is a significant issue. Fatalities on average occur at the rate of roughly 0.6 percent of all reported accidents. To extract meaningful policy, very large amounts of driving data are required to develop a sample size of non-traditional variables (for example, relating to impaired driving, access to taverns, breweries and pubs along commuter routes and proximity of these locations to drivers' residences). In a purely data-driven model, this insight will not be extracted because the database may not initially contain distances from breweries, taverns, and pubs to commuter routes. If one were to estimate a model of fatality likelihoods, domain expertise helps fine-tune a data driven model to include distances and therefore measures of "access" to undesirable effects, since the likelihood of alcohol-impaired driving has a well-known causal effect on fatalities. There is also anecdotal and published evidence in the literature that correlates higher fatality rates with robust economic outlook. The contextual awareness value of domain expertise is therefore lacking in models that are developed on pure statistical reasoning. Therefore, it can be reasoned that big data models (and data driven models in general) could potentially suffer from a model-based data-definition disconnect which can cause issues relating to the identification of relevant variables and potential "missing data" issues. While some of this disconnect may be addressed by automated and trained AI systems, human involvement by way of domain expertise and judgment will still remain a requirement.

The discussion above raises an important issue. If the goal of big data modeling is to provide added insight, then the burden of proof lies in the quality of statistical information extracted from those models. In this sense, big-data modeling is not merely an exercise in techniques that accommodate large amounts of data or simply draw associations among variables, but the predominant burden of proof lies in the ability of these models to provide higher-quality inference ("big" inference). The example of drunk-driving fatalities described above is one example of big inference that can be limited without a basic understanding of the sources of unobserved heterogeneity. Another example of limited inference from big data relates to not adequately making efforts to disentangle causation from correlation, leading to a comingling of the two that can lead to misinformed policy actions (more on this later). These issues can be described as limited big inference in the absence of model-based data definitions using domain expertise. On the other hand, big-data inference can bring in variables that can serve as a source of heterogeneity due to scale. For example, if one were to estimate driving risk models based on naturalistic driving datasets, several non-traditional fine-resolution variables can become available for modeling, such as lane offsetting variables or vehicle kinematic measures such as pitch, yaw, and roll.

Figure 2.1 suggests that the future of big data applications in traffic safety modeling lies at the intersection of strong domain knowledge and quality of extraction of statistical information, and this intersection is heavily influenced by methods that attempt to uncover, to the extent possible, causal effects (after controlling for sources of correlation) and unobserved heterogeneity. Therefore, as a baseline for further evaluation of big data and data driven models, endogeneity models and heterogeneity models can potentially serve as useful tools for both model selection and model definition purposes.

Given the above discussion, with data size and application limitations, what are the potential consequences of trading off predictive capability to understand causality and what factors will compromise our understanding of causality to get better predictive capabilities? Various aspects of this tradeoff are discussed in the following sections, after first discussing causality considerations in safety modeling. In the rest of this chapter, we do not discuss causal inference models because, as already indicated, these models have rarely been applied in the accident analysis literature and are not typically based on individual-accident level data.

2.2 Causality versus Other Explanations in Relationships

The difference between causality and other possible relationship structures involving variables will always be important from a policy action perspective and from the behavioral perspective of improving safety. This is an issue that has been long discussed and remains an important consideration as we enter a "big data" landscape. One possible reason for causality being incorrectly inferred may simply be the fact that the sample being used in safety analysis is itself not representative of the larger population, and thus a relationship estimated for a specific sample may not reflect "true" causal relationships in the larger population. For example, the use of observed accidents, and particularly data conditioned on an accident having occurred, can be potentially problematic for both accident occurrence likelihood and injury-severity statistical modeling because individuals involved in accidents may not be a random sample of the population.[3] That is, the fact that less-safe drivers will be over-represented could potentially present a transference problem of the relationship to the population at large. Further, less-safe drivers may be particularly over-represented in specific types of accidents. To see the problem more clearly, consider a statistical model of the resulting accident-injury severities on a mountain pass. A study of this problem may conclude that high snow accumulations increase the resulting injury-severities in crashes. Injury severity will be known only after a crash has occurred, so it is conditional on a crash having occurred. However, due to the substantial increase in risk involved in driving in snow-related conditions, some drivers may choose to take other modes of travel or avoid traveling adverse weather. Thus, it is possible that the individuals who continue to drive over the mountain pass in adverse conditions are self-selected drivers with risk profiles significantly different from the driving population as a whole. This makes the interpretation of the high-snow-accumulation variable challenging. The variable's estimated parameter could be picking up the actual effect of the snow or merely picking up the unique risk characteristics of the drivers who continue to drive in snowy conditions. It is also possible this effect could be much more subtle than this extreme weather case. For example, safe drivers may avoid dangerous roadway sections or dangerous intersections with specific types of traffic controls by choosing alternate routes than drivers with less of a concern for safety (see, for example, Bhat et al., 2014). In such situations, estimating models on observed crash data will tend to overstate the risk of dangerous roadway

[3] The authors gratefully acknowledge Clifford Winston of the Brookings Institution for identifying this potential issue in traditional safety modeling, and subsequent discussions. There are also the related issues of endogeneity (Bhat et al., 2014) and of under-reporting of accidents, particularly less severe accidents. That is, minor accidents are less likely to be reported to police, which in turn affects what the analyst sees as observed accidents. This has been shown to create model estimation problems as discussed in Mannering and Bhat (2014).

segments and intersections because these roadways tend to have drivers with higher risk than the overall driving population. Some studies have considered only severe accidents (such as fatalities) thereby potentially compounding the problem because the sample is further restricted making less-safe drivers even more over-represented in the sample.

Another possible, and broader, reason why causality is co-mingled with other associative effects is that many of the explanatory variables used in accident-likelihood and injury-severity models could be viewed as endogenous, causing inconsistent parameter estimates and compromising the interpretation of the statistically estimated parameters (Washington et al., 2011, Abay et al., 2013). For example, seat belt use may be endogenous to injury severity. In other words, individuals who do not wear seat belts may be overrepresented in severe injuries (conditional on an accident), but this may be because those who do not wear seat belts are intrinsically aggressive drivers and this aggressive driving itself may contribute to severe injuries. Thus, one may have to consider seat belt use as an endogenous variable to determine the true causal engineering benefit of seat belt use in preventing serious injuries conditional on an accident. Importantly, such considerations are not merely esoteric scholarly pursuits, but are very germane to assessing the potential effectiveness of various countermeasures and selecting priority measures. In the next few sections, we discuss the ability to investigate causality effects from different types of data/methods.

2.3. Causality and the Nature of Traditional Accident Data

Of all the many safety-related studies that have been undertaken over the years, those that are based on police-reported accident data have formed the primary basis for developing statistical models to help guide specific safety-related highway and traffic-control improvements. Over the years, the analyses of these data have become increasingly sophisticated, evolving from simplistic regression analyses to highly sophisticated endogeneity/heterogeneity methods. Although the front-line statistical methods used to analyze these data are the mainstay of academic journals, from an application perspective, the culmination of this research is embodied in the Highway Safety Manual (AASHTO, 2010). The Highway Safety Manual approach is based on police-reported vehicle accident data and has used that empirical basis to provide a practical and readily accessible way of quantifying the likelihood of safety-related impacts of specific highway improvements.

With regard to the likelihood of accidents, using police-reported accident data, studies commonly seek to model the number of accidents occurring on a highway entity, such as a segment of highway or intersection, over some specified time period using count-data or other statistical methods (Lord and Mannering, 2010). Explanatory variables may include roadway characteristics such as traffic volume, lane widths, pavement friction, highway grade and curvature, and so on. Regarding the injury severity of accidents (occupant injury levels such as no injury, possible injury, evident injury, disabling injury, and fatality), discrete-outcome statistical methods are typically applied (Savolainen et al., 2011). Information on injury severity is available only after an accident has occurred (thus conditioned on an accident having occurred). Using data conditioned on the fact that an accident has occurred, the explanatory variables can be expanded from the highway-segment data used in the accident-likelihood models to include accident-specific variables such as seat-belt use, blood-alcohol level of drivers, weather conditions at the time of the accident, and so on.

As discussed earlier, the use of observed accidents, and particularly data conditioned on an accident having occurred, can be potentially problematic. For sure, the possibility of such selectivity would make the interpretation of the parameters difficult, specifically for weather-related parameters and more so for some modes of highway travel (for example, motorcyclists are particularly likely to self-select in rain and snow as discussed in Mannering, 2018). More importantly, for forecasting with models estimated with traditional police data and even other real-time data, anything that would shift the self-selectivity of road users in adverse weather or on unsafe routes would result in inaccurate predictions. As examples of this self-selectivity shift, newer vehicles with advanced safety features may make drivers more confident in adverse weather conditions, thus changing the mix of drivers in such conditions. Regarding route choice, safe drivers may seek to avoid dangerous roadway segments and intersections, but as congestion increases, they may alter their travel routes as they trade off time and safety and this, in turn, could change the mix of drivers on specific roadway segments.

Methods to attempt to control for self-selectivity and related considerations are discussed in the next section. Data requirements and econometric complexities to implement these procedures for accident data analysis can be formidable obstacles. To circumvent data barriers, many economists have sought more simplistic causal-inference approaches to address identification issues and uncover causality, particularly with the application of ordinary least-squares regressions to choice applications (Dale and Krueger, 2002). This is generally done by using control variables such as indicator variables and fixed effects, with the intent of achieving the equivalent of a randomized trial where self-selectivity and endogeneity can be strictly eliminated (Angrist and Pischke, 2009; 2015;

2017). However, the generalizability of the fixed-effects results can be questionable, and even in a truly randomized trial likely temporal shifts in observed behavior can make prediction problematic with ongoing temporal variations inducing unknown errors in fixed effects (Mannering, 2018). In the relatively complex non-linear models of the likelihood and severity of highway crashes that include many explanatory variables relating to roadway characteristics, traffic conditions, weather conditions, and vehicle and driver characteristics, identification of control variables and their incorporation into the model is much more challenging than the more aggregated-data methods applied by economists to address this problem. In addition, predictive application can be quite limited because the variables used as controls may also be of interest for predictive purposes. It is important to note that even analyses that consider the likelihood of an accident, such as accident frequency models, that typically include roadway characteristics and do not include any driver characteristics, are still potentially affected by selectivity. For example, safe drivers may choose to avoid roads with certain characteristics so the observed accidents on specific roads may not be drawn for a random sample of the driving population. Thus, an estimated parameter for a dangerous curve could theoretically be overstated since high-risk (more accident-likely) drivers may be overrepresented on that curve.[4]

The potential bias that selectivity introduces and the effect it may have on prediction is not fully understood, though evidence of the potential biases due to ignoring self-selection has been presented in Shin and Shankar (2013) in an analysis of accident severity likelihoods. But, as pointed out in Mannering (2018), the issue is likely to be very context dependent. For instance, because everyone has a chance of being involved in an

[4] While such road selectivity among safe drivers may exist, the authors are unaware of any studies that have quantified this effect.

accident (even the safest drivers), it may be that an accident data sample collected just so happens to include the full spectrum of individuals from the safest to the least safe. In addition, when considering the injury severities in an accident, it is not clear whether the drivers observed in accidents will have more severe injuries, less severe injuries, or about the same injuries relative to drivers not observed in accidents. For example, drivers frequently appearing in accident data bases may get involved in more accidents of lower-injury severity than those less frequently involved in accidents. It is also important to note that in the cases just mentioned, the resulting injury severities are fundamentally different from traditional endogeneity applications that often have an outcome determined by a choice. In the case of vehicle accidents, once various driver actions are taken, the resulting injury severity is determined by physics where forces are transferred through the vehicle to its occupants (though even the physics involved in the crash are influenced by underlying risk profiles of the driver including vehicle choice and other factors). However, endogeneity of variables in accident data, where the self-selection is based on a choice (such as wearing seat belts or not, or whether a motorist decides to drive at all or not in severe weather, or where traffic engineers choose to place specific types of traffic control devices, or where engineers decide to place additional lighting), is likely to be a more serious issue, as has been demonstrated by Eluru and Bhat (2007), Oh and Shankar (2011), and Bhat et al. (2014).

What is clear, is that selectivity of any form (based on human choice or otherwise) should certainly be considered in the interpretation of any model results that use traditional accident data (data that only includes accident-involved individuals), and even naturalistic driving and observed traffic data since selectivity on safe and less-safe routes could be a factor.

2.4. Endogeneity, Unobserved Heterogeneity and Causality

As just discussed, traditional statistical approaches to the analysis of highway safety data (based on observed accident data) have struggled with a variety of statistical issues, most notably endogeneity bias and omitted variables bias, because traditional statistical methods are often estimated with limited data for practical reasons. Despite these limitations, traditional models have the advantage of being accessible and easily applicable, and they have had a measurable real-world impact on highway safety practice. Nonetheless, traditional methods can be substantially enhanced in their value by recognizing elements of endogeneity and unobserved heterogeneity.

Endogeneity considerations (including those involving self-selectivity as discussed earlier) may be handled in one of two broad ways (for more details, please see Bhat and Eluru, 2009). One approach is based off Heckman's seminal work in the 1970s (Heckman, 1979), and has been extended to numerous transportation applications that have been undertaken over the years (for reviews see Mannering and Hensher, 1987; Washington et al., 2011). In particular, using variations of Heckman-style methods, transportation applications have considered a number of issues in this regard, such as selectivity bias corrections for vehicle usage models (Mannering and Winston, 1985; Mannering, 1986a, 1986b; Oh and Shankar 2011; Shin and Shankar 2013), which are needed because, for example, individuals that own newer vehicles (which are capable of being driven more with fewer repairs) are a non-random, self-selected sample of higher-use vehicle owners.

There has also been work with selectivity-bias corrections for average speed by route (Mannering et al., 1990), with the idea being that drivers attracted to specific routes are a non-random sample (for example, faster drivers may be more likely to take freeways and slower drivers may be more likely to take arterials). The basis of the Heckman-style approach is to start with a probabilistic model that captures the selectivity process and then to incorporate the probability of the outcomes under consideration to correct the bias in the model that estimates the magnitude of the outcome. In the case of safety research, this would presumably start with a model that considered individuals' overall probability of being accident involved (or, following the weather-related example earlier, the probability of a motorist driving in adverse weather), and then use this to correct statistical models to the overall frequency and severity of crashes as gathered from observed accidents. However, classic Heckman-style selectivity corrections are manageable because the equation being corrected is a simple linear model with a continuous variable (vehicle usage in miles driven per year, average speed in miles per hour, etc.). In the analysis of accident data, the likelihood of an accident and its resulting injury severities are typically modeled using non-linear count-data and discrete outcome models (Lord and Mannering, 2010; Savolainen et al., 2011), which makes a Heckman-style selectivity correction (using control function approaches) an econometric challenge, particularly if unobserved heterogeneity and other more advanced econometrics are involved in the model as well (Mannering and Bhat, 2014; Mannering et al., 2016).

Another approach to handling endogeneity is inspired by the work of Heckman (1974) and Lee (1983). Rather than use a two-step Heckman type approach, this second approach models the potential endogenous variable jointly with the outcome of interest.

While this second approach has been used in a general transportation context for a long time (see Hamed and Mannering, 1993, Bhat, 1996, and Bhat, 1998), the approach has only been relatively more recently applied to models in the safety literature (Eluru and Bhat, 2007, Bhat and Eluru, 2009, Oh and Shankar 2011, Spissu et al., 2009, Pinjari et al., 2009, Abay et al., 2013, and Bhat et al., 2014). Thus, for example, by modeling seat belt use as well as injury severity in a joint model system (allowing for correlation in the error terms of the underlying equations determining these discrete outcomes, say because of aggressive/risky driving behavior), one can estimate the remaining "true" causal effect of seat belt use on injury severity (addressing also the situation that aggressive drivers are likely to be over-represented in accident-only data). Importantly, through the use of copula methods employed in some of the more recent applications listed earlier of the joint approach, a variety of parametric distributions may be used to characterize the nature of the joint distribution of the errors in the joint system. While the Heckman-type control function approach is generally considered to be more robust to miss-specification of the error distributions, this issue is at least assuaged in the joint model system by testing different distributions forms through copulas and selecting the best fit copula (see Mannering and Bhat, 2014). Further, the joint model system is estimated in a "one-shot deal" and does not incorporate corrections for the standard errors as needed in the second step of Heckman-type methods. The joint model approach also technically does not need the a priori identification of an instrument variable that affects the selection equation (seat belt use in the example above) but not the outcome equation (injury severity) because identification is facilitated through the assumed parametric distribution of the error terms. However, for stability purposes, having at least one variable affecting the selection equation but not the outcome equation is helpful even in the joint model approach, and such exclusion restrictions can be determined through empirical estimations.[5]

While endogeneity models attempt to account for self-selectivity and related broader jointness issues, heterogeneity models (including random parameters models, latent class models and others) recognize the presence of countless factors that are unlikely to be observed by the data analyst (unobserved heterogeneity) and that influence accident likelihoods and resulting injury severities, despite the presence of a large number of potential explanatory variables. Because heterogeneity models have been the focus of an entire paper recently in the safety field (see Mannering et al., 2016), we do not expend too much space discussing the motivation and methods for such models here. But, using the random parameters application as an example, these heterogeneity models allow the effect of explanatory variables to vary from one accident to the next and from one roadway to the next in (or other units of observation for accident analysis, such as drivers, counties, vehicles, etc.). This can account for a vast variety of unobserved factors and can also potentially mitigate the selectivity issue (that riskier drivers will be over-represented) by giving different parameter values to different observations. However, restrictive distributional assumptions are often made, and prediction can be challenging due to the complexity of the models and the observation-specific estimated parameters. In the process of incorporating unobserved heterogeneity through random-parameter type specifications, it is important that observed heterogeneity not be given less attention. From a causality and

[5] Identification ensures the parameters of interest are uniquely estimable (see for example, Manski 1995; Manski 2009). Lavieri et al. (2016), based on the Generalized heterogeneous data model (GHDM) of Bhat (2015), extend this joint modeling approach by using a small set of common latent stochastic constructs affecting multiple outcomes to generate a parsimonious covariance matrix across the multiple outcomes.

policy insight perspective, it is critical that all sources of observed heterogeneity (through observed exogenous variables) be tested and specified first, and unobserved heterogeneity, as referred to in our label of "heterogeneity models", be included to recognize the inevitable presence of the moderating effect of unobserved factors *after* accommodating for the presence of observed heterogeneity, rather than *in-lieu* of observed heterogeneity.

2.5. Data Driven Methods, Big Data and Causality

Due to the structure of the models and estimation procedures, traditional statistical models and endogeneity/heterogeneity models have difficulties in processing very large amounts of data (big data). There are a number of data driven methods that have been applied to the analysis of accident data with the intent of uncovering correlations and developing accurate predictive models. Still, the field of accident analysis is ripe for additional applications of non-regression data-driven methods (which are often free from standard parametric assumptions used in traditional regressions). The class of non-regression methods is fairly broad, inclusive of: instance-based algorithms (such as K-nearest neighbor, or support vector machines, etc.); regularization algorithms (such as the least absolute shrinkage and selection operator); decision tree algorithms (such as classification and regression trees); Bayesian networks (such as naïve Bayes and Bayesian networks among others); clustering (K-means, expectations-maximization, etc.); association rule algorithms; artificial neural networks (such as back-propagation and stochastic gradient descent); deep learning algorithms (such as the convolutional neural network, deep belief network, etc.); dimensionality reduction algorithms (such as principal component analysis and its variants); ensembling algorithms (such as boosting and bagging, random forests); feature selection algorithms, reinforcement learners, natural language processing, and so on. In accident analysis, for example, outside of the numerous studies on regression applications, support vector machines have been employed (Li et al., 2008) for the estimation of both frequency and severity outcomes (Li et al, 2008; Li et al, 2011). In addition, artificial neural networks (Abdelwahab and Abdel-Aty, 2001; Abdel-Aty and Pande, 2005; Chang 2005; Delen et al., 2006; Abdel-Aty et al., 2008;), support vector machines (Li et al., 2008; Yu and Abdel-Aty, 2013), Bayesian networks (Hossain and Muromachi, 2012; Sun and Sun, 2015), classification and regression trees and hierarchical tree-based regression (Karlaftis and Goulias, 2002), Bayesian neural networks (Riviere et al., 2006; Xie et al., 2007), deep belief networks (Pan et al., 2017) and classification trees (Pande and Abdel-Aty, 2006a,b) have been applied to evaluate real-time crash risk.

While the "universe" of data-driven methods is rich for application to accident analysis, several limitations exist relative to traditional econometric and statistical methods. Better prediction is a potential benefit; and the field of statistical reasoning has provided excellent tools for improved "curve fitting" of observations in a fundamental sense. However, questions still remain regarding the appropriate measures of the inferential quality of the data-driven algorithms. First and foremost, among the measures, is the measure of "why?" Are the variables extracted from the data-driven methods able to provide insight into cause and effect that are robust over time, and transferable to other domains? The current answer is that to date no data-driven method has been shown to provide true cause and effect and true cause-and-effect transportability to another domain of search. That is, for example, even if the training dataset was exhaustively analyzed to reveal purported cause and effect, the algorithm would more than likely fail in a different learning scenario with a very different set of hidden causal relationships. Transfer learning, domain adaptation and intelligent causal rule generation are still well beyond the reach of the big-data/AI claims that are published in the literature.

Most existing applications of data-driven methods in the accident-analysis field have also not really dealt with big data (where data-driven methods become the dominant approach), but instead have dealt with data sizes that place them in direct competition with other traditional statistical techniques. With traditional data sets (in terms of size), sophisticated forms of these data-driven methods have been shown to predict accident data with comparatively high accuracy (earning high predictability marks in Figure 2.1). However, the inability to uncover causality and provide substantive inferences has been a historical weakness of these approaches, often earning them a "black-box" designation because of the difficulty of unraveling how specific elements might influence predictions with these approaches (giving it low marks for causality in Figure 2.1). While data-driven methods are likely to become increasingly popular with the emergence of truly high dimensional big-data in transportation safety (National Academies, 2013), the fundamental limitations relating to causality must still be given consideration in the interpretation and application of results. The bottom line is that, while data-driven methods may do well in capturing associations between one variable and another (that is, how variation in a variable influences another variable), they do not intrinsically study the issue of what exactly is the root cause of why variation in one variable influences another variable. While one could claim that this is the same even with traditional "structure-based" econometric analyses, there is some level of domain theory and knowledge that underlies structure-based analyses that facilitates drawing more causal inferences (especially when endogeneity and heterogeneity issues are recognized). In particular, traditional structure-based methods are driven by well-informed causal frameworks based on domain knowledge. While the relationships implicit in these frameworks may be characterized as assumptions by some, it is important to note that assumptions need to be made in all kinds of analyses, including data-driven analyses (for example, regardless of the methods used, one has to define what are the outcome variables and what are the explanatory variables, and not every variable can be associated with each other variable).

2.6. Discussion and Conclusions

Safety analysts often face challenges in trading off the predictive capability of the methodological approach with its ability to uncover underlying causality. The trade-offs must consider available data in terms of the number of variables and number of observations as well as the intended use of the results. In some practical applications, highway safety engineers may need to know highly specific information. For example, what impact would increasing the shoulder width from 4 to 6 feet on a two-lane rural road with specific traffic characteristics and geographic location have on the likelihood and resulting injury severity of crashes. Getting to this level of detail necessitates specific data requirements and advanced methodologies, and likely some compromise between predictive accuracy and underlying causality. Those who strongly support causality as the only correct approach are often highly critical of methods that do not fully address causality, sometimes arguing that no prediction is better than a prediction based on a flawed causal model (although they rarely if ever provide empirical evidence to support this argument). But this argument does not fully appreciate the potential benefits of having some level of predictive capability. In contrast, those who consider purely data-driven analyses neglect potential insights into underlying causality. Without an understanding of underlying causality, changes in vehicle technology, roadway features, and human behavior may fundamentally shift model parameters that would ultimately impact predictions and safety-policies.

An ideal model would be one that uncovers causality, has excellent predictive capabilities, and is scalable to very large data. However, with currently available methods, safety analysts are often forced into a causality/prediction tradeoff that can entail serious compromises. Thus there is a clear need in the safety field to ground intrinsically predictive models within causal frameworks, while also taking insights from intrinsically predictive models (especially from big data) to improve upon causal structures through insights from associations involving variables not typically available in traditional safety data. One promising direction for future research would be a hybrid modeling approach of data-driven and statistical methods (with strong consideration to causal elements). Such a hybrid approach is likely to be perfected over time as integrative techniques are perfected and access to more and more big data becomes available. However, during this development period it is important that strong domain knowledge remain at the front and center of all analytic approaches and their subsequent interpretations for predictions and policy actions.

Chapter 3

An exploratory analysis of the role of socio-demographic and health-related factors in ridesourcing behavior

Natalia Barbour, Yu Zhang, Fred Mannering

3.1. Introduction

Technological advancements have allowed for the growth and development of the sharing economy, which is a phenomenon based on renting and borrowing goods and services instead of owning them. Regarding transportation, shared mobility is still in its early stages and has not yet fully matured, however the concept (which includes the sharing vehicles, bicycles, electric scooters, among others) has been found to have multiple monetary, social, and environmental benefits (Shaheen et al., 2016; Xue et al. 2018).

Ridehailing has become an element of the new economy in the transportation context that has enjoyed rapid growth and transformation in recent years. With smart phone applications, ridehailing services link personal vehicle drivers with passengers who need rides right away or at a future time. One important feature of ridehailing applications is to track and display the real-time locations of drivers and passengers, giving the ability to estimate waiting times. By incorporating multiple technological advancements, ridehailing companies, or called transportation network companies (TNCs) offer a mode of transportation that has a similar flexibility of a traditional taxi but at a lower cost (Rayle et al. 2016). TNCs such as Uber and Lyft have flourished in the United States (Dias et al., 2017). Ridehailing platforms also allow for sharing a ride with another customer traveling in similar direction. Ridesharing belongs under the umbrella of ridehailing and was found to be an important alternative transportation mode because it permits to target individuals who lack access to transit and who are willing to share rides (Erdogan et al., 2014). The need to assess ridehailing does not only arise from its impact on travel behavior but how it is changing urban transportation and economic efficiency relative to the status quo. A key concern is to identify how ridehailing is disrupting existing modes of travel (Jin et al., 2018). Researchers that have compared taxis and ridehailing services seem to confirm this significant modal shift away from taxis (Anderson, 2014; Glöss et al., 2016). Other researchers have found significant modal shift from public transit to ridehailing (Rayle et al., 2016). Although, there are situations where ridehailing complements public transit and serves as a first and last mile solution, the net effect of ridehailing is likely detrimental to public transit use. A substantial body of literature has acknowledged that younger, better-educated, and more affluent individuals are more likely to be ridehailing users (Clewlow and Mishra, 2017; McGrath, 2015; Rayle et al., 2016). Since ridehailing has been associated with a particular user profile, it raises the question of transportation equity, especially in cases when ridehailing substitutes public transit. Evaluating usage rates and recognizing groups that are more or less likely to use ridehailing services contributes to the conversation about transportation equity and transportation poverty.

Although past work, such as that by Dias et al. (2017), has provided some insight into the complex interactions that contribute to the use of mobility-on-demand such as ridehailing and carsharing services, the fact that people’s preferences and experiences with such ridehailing services are rapidly evolving has made understanding the factors that influence ridehailing usage rates, and the temporal evolution of these factors, a challenge. The current chapter seeks to provide additional insights into ridehailing usage rates by gathering detailed data relating to potential ridehailing users' socio-demographics, travel behavior, and health-related characteristics, and then using these data to estimate a random parameters logit model to assess peoples' probability of selecting one of four ridehailing usage-rate categories. Finding what groups tend to use these services more often and what factors contribute to their usage rates is essential to creating an equitable transportation system.

The chapter begins with a literature review that focuses on the brief review of ridehailing services. This is followed by a discussion of survey design, methodology, and the presentation and discussion of model estimation results. Lastly, the chapter concludes with a policy implications, a summary and discussion.

3.2. Overview of Ridehailing Users and Usage Rates

Ridehailing is a subset of the more general shared-mobility concept, and a more extensive body of literature can be found relating to shared mobility options such as car-sharing services (Dias et al., 2017). In fact, the literature on the usage and impacts of carsharing have provided numerous insights. For example, research on the impact of carsharing on vehicle ownership found that households understandably tend to decrease the number of owned vehicles after becoming carshare members (Cervero et al., 2007, Martin et al., 2010, Menon et al., 2019). In other work, Clewlow (2016) found a link between standard carsharing and sustainable travel, including higher modal shares of transit, walking and biking, lower household vehicle ownership, and higher rates of electric vehicle ownership. Regarding socio-demographic factors, education, household characteristics and propensity for non-motorized transportation were found to be significant predictors of carsharing usage (Costain et al., 2012; Coll et al., 2014). Similar variables were taken in consideration in literature aiming to analyze ridehailing usage.

Although literature specifically related to ridehailing has been continuously evolving, previous work has provided some insight into the characteristics and demand for this type of service. For example, Rayle et al. (2016) found both similarities and differences between taxis and ridehailing services in San Francisco area, where market demand and trip lengths were found to be similar for taxis and ridehailing services. However, relative to taxis, they also found that ridehailing services had shorter wait times and operated more consistently across the day, with regard to time and location.

Regarding socio-demographics, Rayle et al. (2016) found that ridehailing customers were generally younger and more highly educated compared to the overall population in San Francisco. Similarly, Dias et al. (2017) found that users of ridehailing services tended to be younger, had higher education, higher income, and lived in more densely populated neighborhoods. Another study done by Deka and Fei (2019) examined both frequency and propensity to ridehailing and their finding also indicated that young people, people with higher income and education, workers, and people with fewer cars in household tend to use ridehailing more extensively than others. They found that women and non-Hispanic white people may have somewhat lower frequency of using ridehailing, but that non-Hispanic whites have greater propensity to use such services. When it comes to the adoption of on-demand services Alemi et al. (2018) also confirmed that younger, better-educated individuals and individuals of non-Hispanic origin are more likely to adopt them. These same authors also concluded that millennials tend to use on-demand ride services more often compared to their older counterparts and that those living in densely populated areas also had higher usage rates. Intuitively, the fact that ridehailing frequency is higher for people living in areas with higher population and employment density was also confirmed by Alemi et al. (2018) as well as others, including Hughes and MacKenzie (2016) and Wang and Mu (2018) who found greater availability of ridehailing services in high-density urban areas.

In general, prior literature has shown that there are multiple factors determining individual's propensity to participate in shared mobility, and in some cases ridehailing in particular. However, in addition to the traditional socio-demographic factors (such as age, and income), health-related factors can potentially play a role. This is perhaps more obvious with shared active transportation modes (such as bikesharing) because of its potential to improve health (Saelens et al., 2003; Frank et al, 2004; Maizlish et al., 2017). This has been supported in the literature with research such as that conducted by Barbour et al. (2019), where they found that a high body mass index (BMI greater than 25) was statistically significant in determining bikesharing usage. However, it could be argued that various health-related variables capture a wide variety of life-style and cultural characteristics that could easily have their influence extend beyond traditional active transportation modes. This possibility will be explored in the current chapter in addition to considering how traditional socio-demographics might affect ridehailing usage rates.

3.3. Survey and Research Design

To collect data for this research, a survey questionnaire was developed to focus on socio-demographic questions, travel behavior and travel patterns, traffic crash history, as well as other health-related questions and preferences regarding shared mobility. The answers to survey questionnaire were collected between February and April of 2018, and the survey was disseminated using online survey software. To ensure that variety of socio-demographic groups who exhibit different transportation behaviors were reached, several different outlets were used to distribute the survey. The CycleHop Bike Share Company, which operates bikesharing programs in Tampa, St. Petersburg, Orlando, and the University of South Florida (Tampa campus) made the survey available to its registered users as well as promoted it via social media. A mailing list from the University of South Florida Tampa campus was also used to distribute the survey.

A total of 675 responses were gathered with the emphasis of the survey being the usage of ridehailing services such as Uber or Lyft. Respondents were asked how frequently they used such ridehailing services and it was found that 118 respondents specified that they have never used either of these services, 280 indicated using them less than once a month, 180 a few times a month followed, and 97 who used them at least once a week.

Regarding variables that may affect these ridehailing usage rates, data were collected on health-related questions such as weight, height and self-assessed health, and the body mass index (BMI) for all 675 respondents was computed. Out of 675 respondents, 410 were found to have a normal body mass index (BMI equals 25 or less) 174 were found to have high body mass index (BMI between 25 and 30) whereas 91 respondents were classified as obese (BMI greater than 30). For self-assessed health only 2 respondents indicated extremely bad health whereas 19 assessed their health as slightly bad, 43 said it was neither good nor bad, followed by 411 and 200 who classified their health as good or extremely good, respectively. Because of the experimental nature of this study, and possible concerns with confidentiality relating to private health data, the type of illness or health condition was not specified. Data were also collected on bikesharing registration and out of 675 respondents, 301 were also registered users of a bikesharing system. In addition to analyzing the ridehailing usage rates, the intent of the study was to explore potential links with behavioral patterns and preferences for shared mobility systems.

Because the research was based on independently collected data that was tied to a particular region within the United States, the study has certain limitations. While the sample is not representative of general population in the United States, it is interesting to note that the usage frequency of ridehailing services obtained in this study follows a similar distribution to the data independently collected by Alemi et al. (2018). Although the datasets were collected approximately three years apart, both of them peak in the ‘less than once a month’ usage category.

3.4 Methodological Approach

As mentioned above, the variable of interest is how often ridehailing services are used with the following choices provided to survey respondents; have never used them, less than once a month, a few times a month, at least once a week. Because of the discrete nature of these four choices, a discrete outcome approach is appropriate. To implement such an approach, a function that determines a probability of how often ridehailing services are used is defined for the four ride-sourcing usage categories mentioned above as,

$$F_{kn} = \boldsymbol{\beta}_k \mathbf{X}_{kn} + \varepsilon_{kn} \tag{3.1}$$

where $\mathbf{X}_{kn}$ is a vector of explanatory variables that affect the probability of observation n selecting ridehailing-usage category k, $\boldsymbol{\beta}_k$ is a vector of estimable parameters, and ε_{kn} is a disturbance term which is assumed to be generalized extreme value distributed. In the above equation, possible unobserved heterogeneity can be accounted for by introducing random parameters, which allow the parameters to potentially vary across observations, giving $\boldsymbol{\beta}_{kn}$ (Washington et al., 2011; Mannering et al., 2016). Further, the possibility of variations in the means and variances of random parameters is accounted for by allowing $\boldsymbol{\beta}_{kn}$ be a vector of estimable parameters that varies across observations defined as (Seraneeprakarn et al., 2017; Behnood and Mannering, 2017):

$$\boldsymbol{\beta}_{kn} = \beta_k + \Theta_{kn}\mathbf{Z}_{kn} + \sigma_{kn} EXP\left(\omega_{kn}\mathbf{W}_{kn}\right) \nu_{kn} \tag{3.2}$$

where β_k is the mean parameter estimate across all ridehailing usage categories, $\mathbf{Z}_{kn}$ is a vector of explanatory variables that captures heterogeneity in the mean that affect ride-sourcing usage category k, $\boldsymbol{\Theta}_{kn}$ is a corresponding vector of estimable parameters, $\mathbf{W}_{kn}$ is a vector of explanatory variables that captures heterogeneity in the standard deviation σ_{kn} with corresponding parameter vector ω_{kn}, and v_{kn} is a disturbance term.

With the above, the resulting random parameters multinomial logit ride-sourcing usage category probabilities are (McFadden and Train, 2000; Washington et al, 2011),

$$P_n\left(k\right) = \int \frac{EXP\left(\boldsymbol{\beta}_k \mathbf{X}_{kn}\right)}{\sum_{\forall K} EXP\left(\boldsymbol{\beta}_k \mathbf{X}_{kn}\right)} f\left(\boldsymbol{\beta} \mid \boldsymbol{\varphi}\right) d\boldsymbol{\beta} \tag{3.3}$$

where $P_n(k)$ is the probability of observation n selecting ride-sourcing usage category k, $f(\boldsymbol{\beta} \mid \boldsymbol{\varphi})$ is the density function of $\boldsymbol{\beta}$ with $\boldsymbol{\varphi}$ referring to a vector of parameters of the density function (mean and variance), and all other terms are as previously defined. This model is estimated by simulated maximum likelihood with 1,000 Halton draws, which has been shown to be an efficient estimation approach (McFadden and Train, 2000; Bhat, 2003; Milton et al., 2008; Anastasopoulos and Mannering, 2009; Behnood and Mannering, 2016).

A wide range of distributional assumptions will be considered in the empirical portion of this chapter including the normal, lognormal, triangular, and uniform distributions.

In addition to the model estimation results, marginal effects are calculated to assess the effects that each explanatory variable has on response probabilities. The marginal effect gives the effect that a one-unit increase in an explanatory variable has on the response probabilities. Average marginal effects over all respondents will be reported (Washington et al., 2011).

Finally, it is important to mention that another methodological alternative to these data would be to use an ordered probability model (such as the ordered probit) because the usage categories are roughly ordered from low to high (from never to at least once a week). While such an ordered approach was considered, the fact that ordered models place a strong restriction on how variables can affect outcome probabilities, the more flexible form of the mixed logit model was chosen (estimation results that show significant variables in the intermediate categories of less than once a month and a few times a month support this choice). Please see Mannering and Bhat (2014) for a more extensive discussion of this point.

3.5 Model Estimation Results

The summary statistics for variables included in the final model estimation are presented in Table 3.1. Table 3.2 presents the results of random parameters logit model for the usage of ridehailing services. A total of 675 observations were used in the mixed logit model estimation and 18 variables were found significant in the four ridehailing usage functions (never use them, use less than once a month, use a few times a month, use at least

Table 3.1. Summary Statistics for Variables Included in Final Model Estimations.

Variable Description	Mean	Standard Deviation
Male indicator (1 if respondent is a male, 0 otherwise)	0.42	0.49
Older age indicator (1 if respondent is at least 50 years old, 0 otherwise)	0.21	0.40
Millennial age indicator (1 if respondent is less than 35 years old, 0 otherwise)	0.54	0.50
Low annual household income indicator (1 if annual household income is less than $75k, 0 otherwise)	0.54	0.50
High annual household income indicator (1 if annual household income more than $175k, 0 otherwise)	0.14	0.35
Black/African American ethnicity indicator (1 if respondent is Black/African American, 0 otherwise)	0.05	0.22
Children under 6-year-old present in household (1 if respondents indicated children present, 0 otherwise)	0.11	0.31
One vehicle household indicator 1 if a household owns (or leases) 1 vehicle, 0 otherwise)	0.30	0.45
Small household size indicator (1 if a household size is at most 2 people, 0 otherwise)	0.64	0.48
Lack of commute indicator (1 if respondent does not commute, 0 otherwise)	0.05	0.23
Bikesharing system registration indicator (1 if respondent is registered for a bikesharing system, 0 otherwise)	0.47	0.50
Short parking time for the most regular trip (1 if parking time is less than 5 minutes, 0 otherwise)	0.61	0.49
Medium parking time for the most regular trip (1 if parking time is between 5 and 10 minutes, 0 otherwise)	0.18	0.39
Commute by driving alone indicator (1 if respondent drives alone for their most common trip, 0 otherwise)	0.73	0.44
Short one-way distance to the grocery store indicator (1 if the distance to grocery store is 1 mile or less, 0 otherwise)	0.16	0.37
Crash involvement indicator (1 if respondent was involved in a vehicle crash, 0 otherwise)	0.54	0.50
Self-assessed health indicator (1 if respondent assessed their health as good or extremely good, 0 otherwise)	0.91	0.29
Obese BMI indicator (1 if respondent has a BMI (body mass index) greater than 30, 0 otherwise)	0.13	0.24
High BMI indicator (1 if respondent has a BMI (body mass index) greater than 25, 0 otherwise)	0.39	0.48

Table 3.2. Random Parameters Logit Model for Frequency Use of Ridehailing Services (All Random Parameters are Normally Distributed)

Variable description*	Estimated Parameter	t-Statistic	Marginal Effects			
			Never [N]	Less than once a month [L]	A few times a month [F]	At least once a week [W]
Constant [S]	0.19	0.40				
Socio-demographic factors						
Male indicator (1 if respondent is a male, 0 otherwise) [N] (*Standard deviation of parameter distribution*)	-2.32 (*4.02*)	-1.44 (*2.35*)	-0.006	-0.013	-0.008	0.028
Older age indicator (1 if respondent is at least 50 years old, 0 otherwise) [N]	1.35	4.24	0.048	-0.027	-0.018	-0.003
Older age indicator (1 if respondent is at least 50 years old, 0 otherwise) [L]	0.80	2.67	-0.016	0.030	-0.012	-0.002
Millennial age indicator (1 if respondent is less than 35 years old, 0 otherwise) [W]	1.87	4.73	-0.020	-0.044	-0.046	0.11
Low annual household income indicator (1 if annual household income is less than $75k, 0 otherwise) [N]	0.79	2.93	0.051	-0.025	-0.019	-0.008
Low annual household income indicator (1 if annual household income is less than $75k, 0 otherwise) [L]	0.97	4.45	-0.030	0.094	-0.044	-0.020
High annual household income indicator (1 if annual household income more than $175k, 0 otherwise) [F]	-0.53	-1.73	0.004	0.006	-0.011	0.002
Black/African American ethnicity indicator (1 if respondent is Black/African American, 0 otherwise) [L]	-1.12	-2.00	0.003	-0.007	0.003	0.001
Household characteristics						
Children under 6-year-old present in household (1 if respondents indicated children present, 0 otherwise) [N]	0.60	1.83	0.062	-0.029	-0.025	-0.008
One vehicle household indicator 1 if a household owns (or leases) 1 vehicle, 0 otherwise) [L]	-0.38	-1.59	0.005	-0.018	0.009	0.003

Small household size indicator (1 if a household size is at most 2 people, 0 otherwise) [F]	0.44	2.13	-0.014	-0.025	0.049	-0.010
Travel behavior						
Lack of commute indicator (1 if respondent does not commute, 0 otherwise) [N]	1.15	2.67	0.012	-0.006	-0.004	-0.002
Bikesharing system registration indicator (1 if respondent is registered for a bikesharing system, 0 otherwise) [N]	-0.96	-3.64	-0.043	0.016	0.021	0.007
Bikesharing system registration indicator (1 if respondent is registered for a bikesharing system, 0 otherwise) [L]	-0.79	-3.19	0.013	-0.058	0.032	0.013
Bikesharing system registration indicator (1 if respondent is registered for a bikesharing system, 0 otherwise) [W]	0.69	2.11	-0.005	-0.011	-0.015	0.032
Short parking time for the most regular trip (1 if parking time is less than 5 minutes, 0 otherwise) [F]	-0.49	-2.49	0.015	0.023	-0.047	0.009
Medium parking time for the most regular trip (1 if parking time is between 5 and 10 minutes, 0 otherwise) [W]	-0.75	-1.74	0.002	0.003	0.004	-0.009
Commute by driving alone indicator (1 if respondent drives alone for their most common trip, 0 otherwise) [W]	-1.11	-3.55	0.011	0.022	0.024	-0.056
Short one-way distance to the grocery store indicator (1 if the distance to grocery store is 1 mile or less, 0 otherwise) [W]	0.68	2.63	-0.006	-0.011	-0.004	0.021
Crash involvement						
Crash involvement indicator (1 if respondent was involved in a vehicle crash, 0 otherwise) [L] (*Standard deviation of parameter distribution*)	0.41 (*2.02*)	1.80 (*1.89*)	-0.011	0.045	-0.022	-0.012

Health indicators						
Self-assessed health indicator (1 if respondent assessed their health as good or extremely good, 0 otherwise) [N]	-0.66	-1.78	-0.071	0.034	0.029	0.008
Self-assessed health indicator (1 if respondent assessed their health as good or extremely good, 0 otherwise) [W]	-1.45	-3.67	0.018	0.039	0.042	-0.099
High BMI indicator (1 if respondent has a BMI (body mass index) greater than 25, 0 otherwise) [W]	-0.57	-1.63	0.003	0.005	0.005	-0.013
Obese BMI indicator (1 if respondent has a BMI (body mass index) greater than 30, 0 otherwise) [N]	0.65	2.10	0.015	-0.007	-0.006	-0.002
Number of observations	675					
Log likelihood at zero	-935.75					
Log likelihood at convergence	-797.63					

* Parameter defined for: [N] Have Never Used; [L] Less Than Once a Month; [F] Few Times a Month; [W] At Least Once a Week

once a week). Only variables that produced statistically significant model parameters (at least the 90% confidence level on a two-tailed t-test) were included in the model. Table 3.2 shows that only two of these variables produced parameters with a statistically significant standard deviation (random parameter), both were normally distributed since other tested distributions did not produce significantly better results. In addition, no variables were found to have statistically significant heterogeneity in the mean or variance. Thus, the random parameters in the estimated model reduce to $\boldsymbol{\beta}_{kn} = \beta_k + \nu_{kn}$ (see Equation 3.2).

In Table 3.2, explanatory variables were grouped into five main categories: socio-demographic factors, household characteristics, travel behavior, crash involvement, and health indicators. Regarding the socio-demographic factors, the male indicator was found to produce a normally distributed parameter with a mean -2.32 and standard deviation equal to 4.02. This results in roughly 28% of male respondents being more likely to have never used ridehailing services and 72% males less likely to do so. This finding captures additional unobserved factors determining gender related travel behavior and reflects the non-homogenous behavior among males. Marginal effects show the overall effect of this variable is that males have a higher probability of at least once a week usage relative to females, and lower probabilities for all other usage categories.

Respondents who were at least 50 years old, produced statistically significant parameters in two of the usage-category functions. The net effect (see marginal effects) was overall a higher probability of never using ridehailing relative to this age group's younger counterparts. This finding is consistent with prior literature addressing adoption of new technologies and services among older adults (Lee and Coughlin, 2015). In contrast, respondents less than 35 years old (primarily millennials) were found to be much more likely to use ridehailing services at least once a week relative to their older counterparts (a 0.11 higher probability as indicated by the marginal effects in Table 3.2), which is also consistent with prior research that found increased use of ridehailing services among millennials (Alemi et al., 2018; Deka and Fei, 2019; Rayle et al., 2016).

Regarding income level, respondents with lower income (annual household income less than $75,000 per year) were found to produce statistically significant parameters in two of the usage-category functions with the net effect indicating that this income group was more likely to be in the two lower-usage categories (see marginal effects) than their higher-income counterparts. Equality and equity in transportation have been widely addressed in recent studies (Teunissen et al., 2015; Pereira et al., 2016), with low-income groups often being denied opportunities due to the lack of flexible mobility options or due affordability issues. Prior studies conducted by Alemi et al. (2018) and Deka and Fei (2019), among others, also confirmed that the use of ridehailing services is more prevalent among individuals with higher income. While findings in Table 3.2 support this literature, at the other end of the income spectrum, households making more than $175,000 per year were found to have a lower probability of using ridehailing a few times a month, with higher probabilities of using less or more often (see marginal effects). This reflects the rather complex effect income can have on ridehailing-usage rates.

Race was also found to be a statistically significant factor in ridehailing behavior with respondents who indicated being African-American being found, on average, to have a lower probability of using these services less than once a month. Although statistically significant, the small size of the marginal effects reflects the relatively small impact of race on ridehailing usage. The authors would like to emphasize and point out the issue of race being used as an explanatory variable in the study. Although the biological concept of race has long been controversial in psychology and peer reviewed research and some psychologists challenged it in the past, many researchers have adopted it as a reasonable foundation and used it as an inferential research factor and variable (Yee et al, 1993). From a statistical perspective, the African-American indicator is likely capturing unobserved characters associated with race such as those related to cumulative experiences and opportunities, media exposure, societal conditioning, and so on.

In addition to socio-demographic variables, multiple household characteristics and travel behavior variables were found to play a significant role in ridehailing usage. Respondents from households with children under the age of 6 present had, on average, 0.062 higher probability to have never used ridehailing services. The presence of children was also found to be a significant factor preventing respondents from taking bikesharing in previous work (Barbour et al., 2019). Dias et al. (2017) also emphasized the presence of children in the household to be a significant predictor of travel behavior. The effect of children on travel behavior is critical since caregivers' travel-related decisions can be strongly influenced by their presence.

Vehicle ownership and household size were also found to play a role in ridehailing usage. Respondents from one-vehicle households were less likely to use ridehailing a few times a month compared to their counterparts whose households own or lease more than one vehicle. In contrast, respondents from households with at most 2 people living in them were more likely to use ridehailing a few times a month. These results indicate that household size and vehicle availability have interesting and statistically significant effects on ridehailing usage.

For travel-behavior effects, respondents who did not commute were found to be more likely to have never used ridehailing. This suggests that the presence of a commute was found conducive to engaging in using Uber/Lyft like services more frequently.

The impact of willingness to use other modes of shared mobility (in this case bikesharing) on ridehailing usage was also found to produce statistically significant results. An indicator variable for respondents who reported that they were registered users of a bikesharing system was found to be statistically significant in three of the usage functions, with the net effect of the bikesharing registration indicator (see marginal effects in Table 3.2) being that such respondents were more likely to have higher ridehailing usage rates. This finding aligns with some of the prior studies that try to identify groups of people who are more likely to use shared mobility modes (Dias et al., 2017).

Estimation findings shown in Table 3.2 also show that parking time is an important factor in ridehailing usage. Respondents with short parking times for their most regular trip (less than 5 minutes) were found to be less likely to use ridehailing services a few times a month, suggesting that people with longer parking times could plausibly find them convenient and timesaving. Respondents with parking times for their most regular trip between 5-10 minutes were less likely to use Uber/Lyft type services at least once a week.

Respondents who indicated commuting by driving alone had, on average, 0.056 lower probability of using ridehailing at least once a week. This result means that the respondents who commute by other means were more likely to use shared mobility modes more frequently. Once again, a common trait among people who are more willing to use shared and alternate modes of travel was identified.

Respondents who indicated living within 1 mile of a grocery store (which likely indicates urban setting) were found to be more likely to use ridehailing at least once a week. This variable likely reflects the fact that dense development is more conducive to ridehailing where such services can easily substitute for conventional transportation modes. This finding is consistent with prior studies that link urban development and shared mobility which also concluded that ridehailing is most popular and available in high- density urban areas (MacKenzie, 2016; Wang and Mu, 2018).

Regarding traffic accident history, the crash-involvement indicator produced a normally distributed parameter with a mean 0.41 and a standard deviation equal to 2.02. This suggests that 58% of respondents who were involved in a traffic crash were more likely to use ridehailing services less than once a month and 42% less likely. This variable could potentially reflect the level of trust in independent drivers and their privately-owned vehicles. Some people may exhibit more caution when engaging in ridehailing services and others may not.

The last group of variables were health related. An indicator variable for respondents who self-assessed their health as being good or extremely good was found to be statistically significant in two of the ridehailing usage functions. The net effect of this variable (see marginal effects in Table 3.2) were higher probabilities in the less than once a month and a few times a month usage categories, relative to respondents who did not rate their health in these categories.

Regarding other health measures, an indicator variable for the body mass index (BMI) was found to be statistically significant (in slightly different forms) in two of the ridehailing usage functions. In the never-used function, respondents whose body mass index (BMI) was classified as obese (BMI greater than 30) was significant and these respondents had, on average, 0.015 higher probability to have never used ridehailing services. This could be related to a variety of factors such as comfort, the need to travel, the trip purpose, or even some psychological factors given that taking ridehailing involves social networking and engagement. Prior studies have already started to address and analyze the type of social ties and their impact on travel behavior and activity type with the size and diversity of one's core network found to be positively correlated with the variety and frequency of travel-generating activities (Maness, 2017). In addition to this obese BMI indicator, a high BMI indicator (BMI greater than 25) was found to be statistically significant in the at least once a week usage category. Respondents with high body mass index had, on average, 0.013 lower probability to use ridehailing services compared with their counterparts who had a normal body mass index (keep in mind that all respondents in the obese-BMI category will also be in the high-BMI category). These two variables underscore the potential importance of health-related issues in ridehailing behavior and align with previous findings relating to BMI and the use of bikesharing systems (Barbour et al., 2019). BMI indicators have been found to be statistically significant factors in behaviors relating to transportation and it is plausible that they capture some of the life-style choices that determine the propensity to use shared mobility in general.

3.6 Policy Implications and Directions for Further Research

Transportation network companies and their ridehailing services clearly play an important role in the emerging era of shared mobility. Still, the impacts of such companies and their services has been polarizing and because of the relative uncertainty of the long-term impacts on the transportation system.

This chapter sought to provide a clearer understanding on how the customers approach these services and what factors are predictors of their frequency of use. Considering current efforts that are being undertaken in the area of transportation equity, a finding deserve additional scrutiny. Older respondents (older than 50 years old), poorer respondents (those with annual household income less than $75,000), and those with young children (the presence of children under 6 years of age in their household) all had a higher probability of never having used ridehailing services. As new modes of transportation are being deployed and technology advances, it is essential to create policies that will not leave out the most vulnerable members of society but instead contribute to meeting their transportation needs. Social equity has been often overlooked in planning and policy strategies resulting in an historical marginalization of the most vulnerable members of society by policies that did not necessarily provide access to the same social and economic opportunities (Sanchez et al., 2003). Additionally, low-income populations tend to have lower vehicle ownership and vehicle access and are subsequently more dependent on alternative transportation modes (Fletcher et al., 2005; Mackett, 2014). The findings herein suggest that lower income, older age, and the presence of small children in the household are characteristics that should be targeted as a means of reducing transportation inequity. For example, to support parents of young children and assisting them in getting around, having ridehailing drivers offer car seats or at the very minimum booster seats would be a good first step. Approaches such as these could allow emerging transportation options such as ridehailing to address transportation-equity issues.

It is also interesting to note that respondents who were registered for a bikesharing system were less likely to never have used ridehailing services. This creates an opportunity to further cultivate the ‘sharing mindset’; possibly through policy and pricing help to guide the customers who currently make a single-rider trip to consider sharing a ride with another user. Sharing rides does not only increase the efficiency of a vehicle but could also help to break single occupancy trip mindset. Policy efforts that encourage sharing ridehailing trips could be beneficial on multiple fronts and lead the transportation system into the future by altering behavioral economics and system efficiency.

Finally, the fact that many health-related variables were significant in this study suggests that exploring these effects further could be a fruitful research direction. Continuing to enhance the transportation field’s understanding of how transportation and public health relate to each other has potential to assist in building better, healthier, and more equitable transportation systems.

Chapter 4

Individuals' willingness to rent their personal vehicle to others: An exploratory assessment of the peer-to-peer carsharing

Natalia Barbour, Yu Zhang, Fred Mannering

4.1. Introduction

The concept of sharing economy aims to maximize and optimize the use of resources and decrease cost while increasing accessibility of products, goods, or services. It has been widely adopted in the area of transportation under the term shared mobility. In recent years, the sharing economy and shared mobility have been considered an important element of a sustainable transportation system.

The breakdown of shared mobility modes includes bikesharing, carsharing, ridesharing, and the sharing-related potential of private and public transportation network companies (Kodransky and Lewenstein, 2014). Each of them operates on different paradigms and each has a different business model. Regarding carsharing, there currently are four distinctive business models: round-trip carsharing, one-way (or point-to-point) carsharing, peer-to-peer carsharing, and fractional ownership (Shaheen et al., 2018). Most round-trip and one-way carsharing companies allow their users to rent a vehicle on hourly or daily bases by paying a monthly or annual fee.

In this chapter the focus falls mainly on the peer-to-peer carsharing model, which encompasses personal vehicle sharing and allows renting a personally owned vehicle and collecting a monetary compensation for it. Sharing practices and economies are hybrid collections of collaborative ways of consuming for profit-seeking consumption and production (Dowling et al., 2018). While cars have long been shared among family members and close friends, the recent concept of peer-to-peer carsharing has just recently begun to gain popularity. The available vehicle fleet in peer-to-peer carsharing is composed only of personally owned cars. Although the vehicle owners do receive a monetary compensation for renting out their cars, operations of the whole network are usually managed by a third-party company (nonprofit or for-profit), and this company keeps a percentage of that compensation as the commission. One of the upsides of this model is that the operator (rental agency) does not have to buy the vehicle fleet and thus the initial start-up costs are negligible compared to a traditional carsharing model, in which the provider owns the vehicle fleet. This concept supports the sharing and use of already owned and underutilized goods, which in effect allows rental agencies to use the vehicle fleet without the need for maintenance.

Over the years, there have been a few studies that address peer-to-peer carsharing. For example, Shaheen et al. (2012) found that the peer-to-peer carsharing model can also considerably decrease overall operating costs mostly because vehicle capital comprises almost 70% of total operating expenses for traditional carsharing companies. Past work has also noted that some of the desirable consequences of carsharing include reduced car ownership, lower greenhouse gas emissions coupled with its potential to relieve congestion and provide convenient mobility solutions especially in the areas with limited parking (Shaheen et al., 2008; Correia and Viegas, 2011).

The wide application and desirable consequences of peer-to-peer carsharing make this emerging mode worthy of further investigation. Because the presence of a sufficiently large vehicle fleet is a key component for peer-to-peer carsharing to be successful, the focus of this chapter is to explore the vehicle supply side of peer-to-peer carsharing. Specifically, the current study seeks to examine individuals' willingness to rent their personal vehicles by collecting detailed data related to potential vehicle providers' socio-demographics, travel history, travel behavior, and health-related information, and then use the data to estimate a statistical model of their willingness to rent their vehicles in a peer-to-peer system. The increasing interest in the sharing economy and the exponential growth of this business model pose numerous questions rooted in behavioral economics and travel behavior. Transportation-related decisions are influenced by variety of factors, financial being one of them. However, the current study seeks to explore a wider variety of factors and their relationship to the willingness to rent a personal vehicle to a peer-to-peer fleet. The model estimation results will provide insights into the likelihood of individuals allowing someone to use their personal vehicle in exchange for financial compensation.

Peer-to-peer carsharing reflects an important consumer behavior shift. Over the last few decades the consumer behavior has shifted from “you are what you own” to “you are what you can access” (Wilhelms et al., 2017; Belk, 1988, Belk 2014). Such shift in behavior coupled with the introduction of new transportation models merits the efforts to understand this new paradigm. Because the empirical work seeking to understand the role of socio-demographic factors relating to the adoption intention of peer-to-peer carsharing models is limited (Prieto et al., 2017) current efforts have been undertaken.

The chapter begins with a literature review, which focuses on the phenomenon of sharing economy and human factors in peer-to-peer carsharing. The literature review section is followed by survey and research design, methodological approach, then results, discussion and finally the chapter concludes with a summary and discussion of findings.

4.2. The Sharing Economy Phenomenon

Although the term “sharing economy” has become widely used in recent years, some researchers have argued that this is not a novel concept. As pointed out in Frenken and Schor (2017), the sharing economy has been practiced in working class and lower income communities as a way of surviving with limited resources. Nevertheless, the same authors indicated an interesting newness in what is viewed as the modern sharing economy is the notion of sharing one's assets and goods with people outside their social networks. Rapid growth of technology and digital platforms allow for easy and convenient communication with strangers, and thus have made the sharing of assets less complex than it was in the past years.

The practical implication of sharing economy is that it gives others access to an individual's personal, under-utilized assets, with a possibility of monetary benefit. Benkler (2014) defined shareable goods as goods with excess capacity where the owner does not consume the whole product all the time. Some examples of goods potentially having excess capacity include vehicles, houses, boats, and even books and tools. Motivations for participating in sharing economy include economic, environmental, and social or lifestyle related (Bocker and Meelen, 2017).

There is also a separate body of literature addressing the psychological motivation for participating in sharing economy based on common psychological theories. Bellotti et al. (2015) summarized some of the theories of motivation. The theory of the hierarchy of needs suggests sharing economy participation is an outgrowth of psychological and self-fulfillment needs (Maslow, 1943). Self-determination theory is another behavioral theory that is often used to explain sharing economy related behaviors (Hamari et al., 2015; Bellotti et al., 2015; Tussyadiah, 2016). Self-determination theory postulates that motivations can be distinguished as intrinsic (enjoyment, sustainability) or extrinsic (economic benefits, reputation) (Deci and Ryan, 1985, Hamari et al., 2015). Somewhat different, social exchange theory focuses on the formation of motivations for economic relationships through exchanges in a network (Bellotti et al., 2015). Some studies (Hawlitschek et al., 2018) designed to understand consumer motives for peer-to-peer sharing (driver-related, prerequisites, and impediments) were grounded in the Theory of Planned

Behavior (Ajzen, 1991, 1985). This theory implies that a motive to either engage or not engage in certain activities is dependent on beliefs about that activity, with individuals tending to engage in behaviors based on their evaluation of the information available to them.

These theories formed the basis for empirical studies that have sought to identify measurable factors that determine sharing economy participation probabilities. Empirical studies found that older age, education, gender, income, marital status, and the state of being environmentally friendly are significant factors in sharing economy participation (Diamantopoulos et al., 2003; Li et al., 2005; Cornwell et al., 2008; Shen and Saijo, 2008; Hellwig et al., 2015).

4.3. Human Factors in Peer-to-Peer Carsharing

Regarding shared mobility related behaviors, and the process of renting a personal vehicle to receive a monetary compensation, Wilhelms et al. (2017) and Shaheen et al. (2018) found that participants are mainly driven by financial reasons. Nevertheless, sharing economy advocates argued that it could also be a result of more altruistic and environmental-sustainability motives (Belk, 2010). In contrast, some researchers pointed out that enhanced access to vehicles could increase automobile usage and distance traveled and thus lead to more congestion and pollution. They argue that potential increases in vehicle access and usage could eventually impact the accessibility and cost of other modes of transportation such as public transit and taxis (Benjaafar et al., 2018).

Shaheen et al. (2018) found that income, education, race, gender, and age were important factors in peer-to-peer carsharing behaviors. Peer-to-peer carsharing users had slightly higher incomes compared to the US population as a whole. They were also found to be more likely to be males and white and tended to be younger and more educated than the general population. In their work, over half (55%) of the survey respondents who participated in peer-to-peer carsharing used the system one or more times a month. The most active users (8% of the collected sample) used the system five or more times per month. The users primarily used the carsharing system to meet their basic transportation needs including running errands as well as some long-distance recreational travel. Prieto et al. (2017) examined the demand side of peer-to-peer carsharing and found that living in the city center, being male, and being single increase the probability of using peer-to-peer carsharing option. Shaheen et al. (2018) also studied the challenges to the operators of peer-to-peer carsharing. They identified key barriers of peer-to-peer carsharing, which are lack of predictability, reliability, and a fleet exhibiting major differences among the vehicles such as age, maintenance or wear and tear. Due to the lack of state and national policies, vehicle insurance was another reported challenge. Operators were forced to find ways within industry frameworks (many times flawed) to successfully run their peer-to-peer companies. Insurance also reemerged as a key issue with peer-to-peer vehicle services as most state insurance laws have not kept pace with the introduction of peer-to-peer models (Shaheen et al., 2016).

Ballús-Armet et al. (2014) asked respondents in their survey whether they would be willing to rent out their personal vehicle. About 50% of respondents (53% and 47% in San Francisco and Oakland, respectively) indicated having some concerns regarding insurance and liability. Some other concerns mentioned were fear of damage and fear of renting out their automobile to an unreliable customer. Despite the above concerns, more than 25% of surveyed vehicle owners indicated willingness to rent out their vehicles through a peer-to-peer carsharing service. In addition to being motivated by the financial compensation, the respondents were also willing to make use of an otherwise underused asset.

Several different pricing strategies are currently being tested and practiced. Some operators allow the owners to determine the price while others (DriveMycar) do not. There is also a hybrid model that allows the owner to name the price but permits the platform to adjust it higher (Turo). Another pricing approach lets the platform suggest a price (JustShareIt) but allows owners to modify it. Because we are still in the early stages of peer-to-peer carsharing, there is likely to be an experimental and transitional phase in pricing strategies. However, there appears to be a trend toward operators determining the price through developing of sophisticated pricing engines (Benjaafar et al., 2018).

To date, research efforts addressing the socio-demographic factors relating to the adoption patterns of peer-to-peer carsharing have been relatively limited. There has been some work that suggests that socio-demographics in peer-to-peer carsharing significantly influences travel patterns (Dill et al., 2019), but few if any studies have addressed the role of socio-demographic factors in the willingness to participate in the peer-to-peer carsharing. As stated by Martin (2016) and Wilhelms et al. (2017) the literature focusing on the motives for of peer-providers granting others access to their cars remains limited. Despite gaining media attention, academia has mostly focused on studying implications of standard carsharing. The intent of the current study is to develop a deeper understanding of factors determining the propensity to participate in peer-to-peer carsharing and, in particular, to rent out a personal vehicle for financial compensation. To achieve this goal, survey data containing detailed socio-demographic and travel information will be used to estimate a statistical model of individuals’ participation likelihood.

4.4 Survey and Research Design

To gather data for the current study, a web-based survey that focused on preferences and behaviors related to shared mobility was designed. This chapter is part of a wider study and the survey was intended to examine a variety of shared mobility related behaviors. The survey was disseminated between February and April 2018, and the data-collection process was performed using online survey software with the help of CycleHop Bike Share Company and the University of South Florida. The survey was shared with the registered CycleHop Bike Share Company members of bikesharing programs located in Florida (Orlando, Tampa, and St. Petersburg) and the students and faculty of the Tampa campus of University of South Florida as well as advertised online. Because the topics of interest to the general study included shared mobility related behaviors, it was important to collect sufficient sample of both: shared mobility users and the ones who do not use shared mobility modes. To expand the reach of the survey and gather information about the non-users, the link to the survey was also advertised on social media through multiple platforms. This approach allowed the survey to reach a wide variety of respondents and resulted in obtaining information about a diverse group in terms of gender, age, income, and travel behavior. In the final dataset for this work, less than half respondents were bikesharing users (46.7%).

Although, a sufficient number of observations was collected, the survey has some limitations. Because of the multiple channels of survey distribution, the exact location of respondents' residence is not known. Another shortcoming, which is common to stated preference surveys, is the difficulty of stating a preference without prior experience of the studied phenomenon. However, in this case, respondents were asked about their perception towards their currently owned vehicle and how likely they were to rent it to another person, hence it could be argued that they were more aware of how they would behave if presented with such an option.

The survey collected information on socio-demographic factors, household composition, travel behavior, travel history, commute type and length, shared mobility related behaviors as well as some health-related questions such as body mass index and overall health status. Respondents were also asked about their willingness to rent out a personal vehicle to receive a monetary compensation (which will serve as the dependent variable in this study). The respondents were given five possible responses to this question: extremely unlikely, unlikely, unsure, likely and extremely likely. To assure the consistency and validity of the data, the responses of users who do not own a vehicle were not used in the analysis because this would involve having them make a hypothetical choice which would be distinctively different from those respondents currently owning vehicles. Out of the 644 observations that were used to estimate the model 285 (44%) respondents indicated that they were extremely unlikely to rent out their personal vehicle, 208 (32%) respondents implied they were unlikely to do so, 62 (10%) answered they unsure while 69 (11%) and 20 (3%) were likely and extremely likely, respectively.

In addition to the typical socio-demographic and travel related variables that are often used in shared mobility behaviors studies, the dataset was expanded with health-related questions and bikesharing registration status. In the sample used for this research 301 respondents indicated to be registered users of a bikesharing system. Interestingly, some studies have found an association between certain lifestyle choices and positive attitudes toward shared mobility modes in general (Lavieri et al., 2017). Barbour et al. (2019a) concluded that being registered for a bikesharing system had an impact on the likelihood of frequency usage of ridesourcing services such as Uber or Lyft. Selected socio-demographic variables from the collected sample are shown in Figure 4.1. It should be emphasized that a much wider set of variables was used for the analysis.

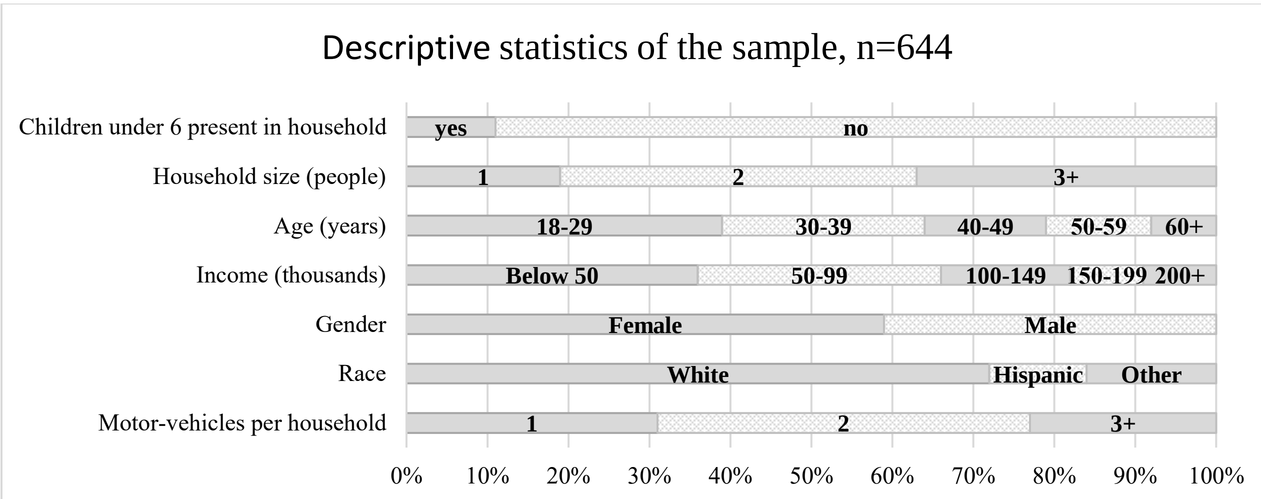


Figure 4.1. Descriptive statistics of selected variables in the collected sample.

4.5 Methodological Approach

The objective of the study is to investigate factors playing key roles in the willingness to rent out a personal vehicle in order to receive a financial incentive. Because the dependent variable is an ordered response to the survey question: how likely would you be to rent out your personal vehicle for monetary compensation (with response alternatives extremely unlikely, unlikely, unsure, likely, and extremely likely), standard ordered-response modeling approaches are appropriate methods of analyzing the survey data (Greene, 1997; Washington et al., 2011). Such approaches begin with defining an unobserved variable, z, as a linear function of explanatory variables,

$$z_i = \boldsymbol{\beta}\mathbf{X}_i + \varepsilon_i, \tag{4.1}$$

where $\mathbf{X}$ is a vector of explanatory variables that determines the discrete ordering of observation i, $\boldsymbol{\beta}$ is a vector of estimable parameters, and ε_i is a disturbance term. Equation 1 is further used to define observed ordinal data, y_i :

$$\begin{aligned} y_i &= 1 \quad \text{if } z_i \le \mu_0 \\ &= 2 \quad \text{if } \mu_0 < z_i \le \mu_1 \\ &= 3 \quad \text{if } \mu_1 < z_i \le \mu_2 \\ &= 4 \quad \text{if } \mu_2 < z_i \le \mu_3 \\ &= 5 \quad \text{if } z_i \ge \mu_3, \end{aligned} \tag{4.2}$$

where 1 = extremely unlikely, 2 = unlikely, 3 = unsure, 4 = likely, and 5 = extremely likely, and μ's are estimable parameters (thresholds) that define y_i and are estimated jointly with the model parameters $\boldsymbol{\beta}$. To determine the probability of the five specific ordered responses for each observation i, an assumption on the distribution of ε_i in Equation 1 must be made. An ordered probit model results if ε_i is assumed to be normally distributed across observations. With this assumption, the ordered category selection probabilities can be written as (Washington et al., 2011),

$$\begin{aligned} P(y = 1) &= \Phi(-\boldsymbol{\beta}\mathbf{X}_i) \\ P(y = 2) &= \Phi(\mu_1 - \boldsymbol{\beta}\mathbf{X}_i) - \Phi(-\boldsymbol{\beta}\mathbf{X}_i) \\ P(y = 3) &= \Phi(\mu_2 - \boldsymbol{\beta}\mathbf{X}_i) - \Phi(\mu_1 - \boldsymbol{\beta}\mathbf{X}_i) \\ P(y = 4) &= \Phi(\mu_3 - \boldsymbol{\beta}\mathbf{X}_i) - \Phi(\mu_2 - \boldsymbol{\beta}\mathbf{X}_i) \\ P(y = 5) &= 1 - \Phi(\mu_3 - \boldsymbol{\beta}\mathbf{X}_i), \end{aligned} \tag{4.3}$$

where $\Phi(.)$ is the cumulative normal distribution.

A positive value of $\boldsymbol{\beta}$ indicates that an increase in $\mathbf{X}_i$ will increase the probability of getting the highest response (extremely likely) and will decrease the probability of getting the lowest response (extremely unlikely) thus the model allows for the clear interpretation of the extreme categories. However, the direction of the effect that the explanatory variables has on the dependent variable is difficult to interpret in the intermediate categories (unlikely, do not know/ cannot say, likely) without computing marginal effects as (Washington et al., 2011),

$$\frac{P_i\left(y = n\right)}{\partial \mathbf{X}_i} = \left[\phi\left(\mu_{n-1} - \boldsymbol{\beta}\mathbf{X}_i\right) - \phi\left(\mu_n - \boldsymbol{\beta}\mathbf{X}_i\right)\right]\boldsymbol{\beta}, \tag{4.4}$$

where $P(y = n)$ is the probability of outcome response n, μ represents the thresholds, and $\phi(.)$ is the standard normal density. Marginal effects indicate the magnitude of effect that a one-unit change in an independent variable has on outcome category n's selection probability, and average marginal effects (averaged over all observations) will be reported.

It is also important to consider the possibility of unobserved heterogeneity in model estimation, which accounts for the fact that unobserved factors may be present in the data that make the effect of explanatory variables vary across individual observations or groups of observations. To account for this possibility, several statistical approaches are available including random parameters models, latent class (finite mixture) models, Markov switching models, or combinations of the above (Mannering et al., 2016). In this study unobserved heterogeneity is considered by estimating a random parameters model with,

$$\boldsymbol{\beta}_i = \boldsymbol{\beta} + \varphi_i , \tag{4.5}$$

where $\boldsymbol{\beta}_i$ is a vector of observation parameters and φ_i is a randomly distributed term (for example, normally distributed term with mean zero and variance σ^2). Maximum likelihood estimation is used to estimate random parameters ordered probit model and 1,000 Halton will be used to arrive at the final model estimation. 1,000 Halton draws have been shown to sufficiently allow accurate parameter estimates (Bhat, 2003; Milton et al., 2008; Anastasopoulos and Mannering, 2009; Behnood and Mannering, 2016).

4.6 Model Estimation Results

Summary statistics of variables found to be statistically significant in the model are presented in Table 4.1. Random parameters ordered probit model estimation results are presented in Table 4.2 with corresponding marginal effects in Table 4.3.

Table 4.1. Summary statistics for variables included in final model estimations.

Variable Description	Mean	Standard Deviation
Female indicator (1 if respondent is female, 0 otherwise)	0.59	0.49
Age indicator (1 if respondent is 40 years old or greater, 0 otherwise)	0.36	0.48
Caucasian indicator (1 if respondent is Caucasian, 0 otherwise)	0.72	0.45
High annual household income indicator (1 if respondents annual household income is more than $200,000/year, 0 otherwise)	0.15	0.35
One-person household indicator (1 if respondent lives alone, 0 otherwise)	0.19	0.39
One motor-vehicle indicator (1 if respondent's household owns or leases 1 motor-vehicle, 0 otherwise)	0.31	0.46
Grocery store time-distance indicator (1 if respondent has less than 5 minutes but more than 1 mile to a grocery store, 0 otherwise)	0.20	0.41
Bikesharing registration indicator (1 if respondent is a registered user of a bikesharing system, 0 otherwise)	0.47	0.50

Turning to the estimation results in Tables 4.2 and 4.3, it is interesting that the body mass index and other health-related variables were not found to be statistically significant in the model. Because health-related variables were previously found to be statistically significant in determining bikesharing and ridesourcing behaviors it was suspected that such variables may capture life-style choices that could affect rental likelihoods, but the model estimation results do not support this (Barbour et al., 2019a, Barbour et al., 2019c). However, Table 4.2 shows that gender is a statistically significant variable in determining the willingness to rent out a personal vehicle to receive a monetary compensation. As indicated by the average marginal effects (Table 4.3) female respondents were found to have a higher probability of being extremely unlikely to rent their personal vehicle relative to their male counterparts. Female respondents had, on average, 0.072 higher probability to be extremely unlikely to rent their personal vehicle. This finding is consistent with Shaheen et al. (2018) who found that males were more inclined to use peer-to-peer carsharing services relative to females.

Table 4.2. Random parameters ordered probit model estimation of willingness to rent a personal vehicle to receive a financial reward (extremely unlikely, unlikely, unsure, likely, extremely likely) (all random parameters are normally distributed).

Variable Description	**Estimated Parameter**	***t*-Statistic**
Constant	0.53	4.32
Female indicator (1 if respondent is female, 0 otherwise)	-0.18	-1.99
Age indicator (1 if respondent is 40 years old or greater, 0 otherwise) (*Standard deviation of parameter distribution*)	-0.36 (*0.49*)	-2.59 (*5.96*)
Caucasian indicator (1 if respondent is Caucasian, 0 otherwise)	-0.19	-1.87
High annual household income indicator (1 if respondent's annual household income is more than $200k, 0 otherwise) (*Standard deviation of parameter distribution*)	-0.39 (*0.90*)	-2.59 (*5.85*)
One-person household indicator (1 if respondent lives alone, 0 otherwise)	-0.22	-1.45
One motor-vehicle indicator (1 if respondent's household owns or leases 1 motor-vehicle, 0 otherwise) *(Standard deviation of parameter distribution)*	0.21 (*0.61*)	1.53 (*7.22*)
Grocery store time-distance indicator (1 if respondent has less than 5 minutes but more than 1 mile to a grocery store, 0 otherwise)	-0.20	-1.69
Bikesharing registration indicator (1 if respondent is a registered user of a bikesharing system, 0 otherwise)	0.16	1.72
Threshold, μ_1	1.00	16.61
Threshold, μ_2	1.42	19.43
Threshold, μ_3	2.32	19.31
Number of observations	644	
Log-likelihood (constant only)	-836.08	
Log-likelihood at convergence	-817.18	

Table 4.3. Marginal effects of the random parameters ordered probit model estimation of respondent's willingness to rent a personal vehicle in order to receive a financial reward.

	Average Marginal Effects				
Variable Description	**Extremely Unlikely**	**Unlikely**	**Unsure**	**Likely**	**Extremely Likely**
Female indicator (1 if respondent is female, 0 otherwise)	0.072	-0.021	-0.018	-0.026	-0.007
Age indicator (1 if respondent is 40 years old or greater, 0 otherwise)	0.141	-0.047	-0.035	-0.047	-0.012
Caucasian indicator (1 if respondent is Caucasian, 0 otherwise)	0.076	-0.020	-0.019	-0.028	-0.008
High annual household income indicator (1 if respondent's annual household income is more than $200k, 0 otherwise)	0.156	-0.060	-0.037	-0.047	-0.011
One-person household indicator (1 if respondent lives alone, 0 otherwise)	0.091	-0.032	-0.023	-0.030	-0.007
One motor-vehicle indicator (1 if respondent's household owns or leases 1 motor-vehicle, 0 otherwise)	-0.081	0.022	0.021	0.030	0.008
Grocery store time-distance indicator (1 if respondent has less than 5 minutes but more than 1 mile to a grocery store, 0 otherwise)	0.078	-0.027	-0.019	-0.026	-0.007
Bikesharing registration indicator (1 if respondent is a registered user of a bikesharing system, 0 otherwise)	-0.062	0.018	0.016	0.022	0.006

Respondents who are 40 years old or greater, produced a normally distributed random parameter with the mean equal -0.36 and standard deviation 0.49. Although the effect of this variable suggests heterogeneous behavior among this group of users (as reflected by the presence of a statistically significant random parameter), this age group, on average, had a higher probability to be extremely unlikely to rent their personal vehicles to others as indicated by marginal effects in Table 4.3. An older age indicator has previously been identified by many researchers to be an important factor in determining the adoption of sharing economy systems (Diamantopoulos et al., 2003; Li et al., 2005; Cornwell et al., 2008; Shen and Saijo, 2008; Hellwig et al., 2015).

Respondents who identified themselves as Caucasians had a higher probability to be extremely unlikely to rent their autos relative to others. Previous studies addressed, to some degree, the behavior differences between Caucasians and other races and found that Caucasians do not only have higher accessibility to automobiles (Berube et al., 2006) but were also more skeptical to engage in shared mobility behaviors and modes (McKenzie, 2015). The authors would like to point out and exercise caution with regards to including the race variable in the study. Although, biological concept of race has long been controversial in psychology and many psychologists have challenged it, other researchers have adopted it as a deductive premise and applied it as an inferential research factor and variable (Yee et al, 1993). From a statistical perspective, the Caucasian indicator is likely capturing unobserved characters associated with race such as those related to cumulative experiences and opportunities, media exposure, societal conditioning, and so on.

Respondents from households with a high annual household income (above $200,000/year) produced a normally distributed random parameter with mean -0.39 and standard deviation equal to 0.90. Although the effect of this variable was found to vary significantly across the respondents, the average marginal effect equal to 0.156 (in the extremely unlikely category) suggests that the respondents from this group had a higher probability to be extremely unlikely to rent their personal vehicle to others. Other studies also found that income was a significant factor in determining the likelihood of engaging in shared mobility options (Shaheen et al., 2014; Woodcock et al., 2014; Dias et al., 2017). Psychological studies that touched upon the motivations to participate in shared mobility have argued that, although financial need may be a driving factor in shared mobility engagement, altruistic and environmentally sustainable motives could be equally important (Belk, 2010). This could explain the non-homogenous behavior in this group as reflected by the statistically significant random parameter. This heterogeneous behavior among respondents from higher income households could indicate an opportunity for peer-to-peer carsharing model to be adopted among a wide variety of socio-demographic groups as well as highlights the fact that the reasons to participate in sharing economy are not driven strictly by financial needs (discussed in detail by Wilhelms et al., 2017).

Regarding household composition, respondents from one-person households had a higher probability to be extremely unlikely to rent out their personal vehicle (as indicated by marginal effects presented in Table 4.3). Although marginally significant, this finding likely captures higher auto reliance among respondents from one-person households.

Respondents from households with only one vehicle produced a normally distributed random parameter with a mean 0.21 and standard deviation equal to 0.61 (recall that responses from respondents whose households did not own any vehicles were excluded from the analysis). The effect of this variable suggests there is a significant difference in behaviors among respondents from one-vehicle households. It could be speculated that these differences are a result of lifestyle choices and other aspects of decision making that are not directly addressed in the survey. That is, there is likely considerable variance in one-vehicle households ranging from an income-constrained household with heavy use of their single vehicle to households choosing to own one vehicle as a lifestyle choice to minimize their carbon footprint, and these households may also be highly motivated to participate in sharing economy initiatives.

Regarding travel patterns, the explanatory variable indicating one-way travel time to a grocery story of five minutes but one-way distance to a grocery store being more than one mile was statistically significant in the model. Other travel-related variables such as total daily travel time or commute distance and time were considered but they did not produce statistically significant parameters. Respondents from such households (whose one-way travel time to a grocery store is less than five minutes but the distance is more than one mile) were found to have, on average, higher probability to be extremely unlikely to rent their personal vehicle. Presumably, this variable captures a mix of factors relating to respondents' lifestyle choices and preferences, as well as residential/commercial forms and levels of traffic congestion since having quick access over longer distances provides information on the residential and commercial spatial distribution, and the status of transportation network. Grocery store location with respect to respondent's residence was also statistically significant in previous papers addressing adoption of new technologies such as shared automated vehicles (Barbour et al., 2019b).

Lastly, bikesharing registration indicator was found to be statistically significant in the model estimation results (Table 4.2). As indicated by the average marginal effects (Table 4.3), respondents who noted that they are registered for a bikesharing system had, on average, 0.062 lower probability to be extremely unlikely to rent out their personal automobile and higher probability to belong to other willingness categories. This result offers an important insight relating to travel behavior and it aligns with prior studies that link shared mobility behaviors across multiple modes and transportation options (Barbour et al., 2019c).

4.7 Discussion

Jacoby et al. (1977) suggest that a consumer contemplating disposition of their belongings has three choices: keep the belonging; get rid of the belonging completely; or temporarily get rid of the belonging, either by loaning or renting it out. The literature that examines the motives of the option of temporary disposition is limited (Philip et al., 2015). Some authors (Philip et al., 2015) found that the providers in peer-to-peer sharing were primarily motivated by economic considerations, which is consistent with prior literature (Wilhelms et al., 2017) and to some degree consistent with the results presented herein. Current research found considerable heterogeneity among people from households with relatively high annual income (more than $200,000). This finding suggests that renting enables usage of their idle possessions, which many providers find gratifying even in the absence of financial need. Other research, such as that of Bardhi and Eckhardt (2012), found that temporary disposition of one's belongings also provides a way to protect the environment and reduce waste.

Another interesting finding is the fact that female respondents were found to be more likely to be extremely unlikely to rent their personal vehicle. This finding is also reflected in prior literature (Shaheen et al., 2018) and it could reveal higher reliance on personal vehicles by female users or, in other words, a lack of other transportation options that could substitute a personal vehicle. Women's mobility has been extensively studied and a large body of research has confirmed that women do not have the same mobility patterns, mobility needs, or face the same dangers compared to their male counterparts (Dunckel-Graglia, 2013; Mazumder and Pokhare, 2019). Gender differences that have been found across the transportation literature reflect a much wider issue. The novelty is in the fact that gender inequity in transportation can be now studied simultaneously with the introduction of new modes and business models, which provide an opportunity for immediate improvement.

4.8 Summary and Conclusions

This research offers some initial perspectives on factors affecting people's willingness to share their personal vehicles. It explores the role of socio-demographic characteristics coupled with travel patterns and travel behavior factors to identify the key variables that play a role in likelihood of renting a personal vehicle to others. Caucasian female respondents who live in one-person households and have less than five minutes one-way travel time but more than one-mile distance to a grocery store had a higher probability to be extremely unlikely to rent their personal vehicle to receive a monetary compensation in exchange. In contrast, respondents who indicated to be registered users of a bikesharing system, were found to have a lower probability to be extremely unlikely to rent their personal automobiles. The variables that found to have heterogeneous effects across respondents (as reflected by the presence of statistically significant random parameters) were age (at least 40 years old), high annual household income (above $200,000), and households that owned only one motor-vehicle.

Based on the above findings and the fact that there is no national consensus as to how to approach and regulate the peer-to-peer carsharing model, and it is clear that a few policy-related matters must be addressed for the peer-to-peer business model to be successful. The fact remains that changes in consumer behavior and disruptions in the transportation system pose considerable challenges for the car industry and policy makers. As Prieto et al. (2017) point out, peer-to-peer carsharing reveals a potential market that is far less segmented than standard carsharing services. This suggests that peer-to-peer carsharing services are potentially more promising than traditional carsharing because they can potentially appeal to a much broader demographic. Still, studies have shown that policies that support carsharing are needed to ensure success (Kim, 2015).

From the insurance perspective, some insurance companies currently exclude coverage if insured vehicles are used for peer-to-peer carsharing, which forces vehicle suppliers to take on alternate insurance or purchase additional coverage. The lack of affordable insurance policies may clearly be a detriment to the expansion of peer-to-peer carsharing. Additionally, peer-to-peer carsharing companies are in direct competition with traditional rental-car companies which operate at airports and other locations where they need to rent out office space and enough area to store their vehicles. Peer-to-peer carsharing operators also deliver vehicles to customers at the airports, but they are not burdened with such overhead costs and therefore may have an inherent competitive advantage. While there have been attempts to regulate this issue, a national policy was not yet been established. Finally, there is clearly environmental, economic, and social value in utilizing individuals' personal resources and the findings of this study suggest that the motivation to share such resources is not driven wholly by financial reasons, which opens up additional policy dimensions. For example, the findings indicate that female respondents are more likely to be extremely unlikely to rent their personal vehicle, and thus potentially miss out on opportunities for additional income. Prieto et al. (2017) also concluded that being male increases the probability of using peer-to-peer sharing. Same authors also found indication that peer-to-peer users have fewer safety concerns and find more flexibility in arranging car trips through peer-to-peer platforms. Being aware of the unfairness of current transportation system when it comes to accessibility and safety across genders, policies that would equalize the opportunities are suggested. This implies the need for developing appropriate policies that would address potential gender-related equality issues related to peer-to-peer carsharing.

Overall, the findings must be used with caution because people's opinions, perceptions, and preferences regarding peer-to-peer carsharing will likely be changing as they become more familiar with this type of carsharing and additional policies are formed. Sheela and Mannering (2020) provide statistical evidence of such changing opinions, perceptions, and preferences in their study of autonomous vehicle adoption, and peer-to-peer carsharing could follow the same pattern.

Finally, it is important to consider the limitations of the study and to point out that, although this research did not consider the price point at which an individual becomes willing to rent their vehicle (as this price point will vary for each user), it made an assumption that the individual willing to rent their vehicle would be satisfied with the compensation (which is consistent with some of the business models that allow for flexible pricing). Another limitation is the overall lack of popularity and consequently understanding of this business model by general public. Because of the very fluid and uncertain regulations as well as the vagueness about the state of practice, there were likely varied levels of understanding of how peer-to-peer carsharing works and that lack of understanding could be playing a role in some of the responses.

Chapter 5

A temporal analysis of driver-injury severities in crashes involving aggressive and non-aggressive driving

Mouyid Islam, Fred Mannering

5.1. Introduction

Aggressive driving is identified by a wide range of unsafe driving decisions that endanger the safety of drivers and other road users. The Fatality Analysis Reporting Systems indicates that about 53% of fatal crashes are due to aggressive driving (AAA Foundation for Traffic Safety, 2016). Moreover, a driver survey conducted by the American Automobile Association (AAA) Foundation for Traffic Safety in 2011 indicated that about 90% of the drivers viewed aggressive driving as a very serious or somewhat serious threat to their own safety (AAA Foundation for Traffic Safety, 2011). The National Highway Traffic Safety Administration (NHTSA) defines aggressive driving as a combination of moving traffic offenses to endanger other persons or property. In the state of Florida, aggressive driving is identified if at least two of the following occurs: speeding, unsafe or improper lane change, following too closely, failure to the yield right of way, improper passing, and failure to obey traffic control devices (FDOT, 2019). Likewise, New York State defines an aggressive driver as one who operates a motor vehicle in a selfish, bold, or pushy manner, without regard for the rights or safety of the other users of the streets and highways (NYSDOT, 2019). AAA Foundation for Traffic Safety (2009) defined aggressive driving as any unsafe driving behavior, performed deliberately and with ill-intention or disregard for safety. Tasca (2000) characterized aggressive driving as deliberate behavior, likely to increase the risk of collision that is motivated by impatience annoyance, hostility and/or attempt to save time. Other studies have suggested it is behavior that is intended to hurt others (Galovski and

Blanchard, 2002) or an act that disregards safety, irrespective of the deliberate intent of endangering others (AAA Foundation for Traffic Safety, 2009).

While definitions clearly vary, there is general agreement that aggressive driving encompasses a wide variety of driver actions including headlight flashing (Diekmann et al., 1996; Ellison-Potter et al., 2001; Turner et al. 1975), yelling at other drivers (Hennessy and Wiesenthal, 1999; Tasca, 2000), and profanity and obscene gestures (Ellison-Potter et al., 2001; Sarkar et al., 2000; Turner et al., 1975) which often lead to transgressive behavior that include speeding (James and Nahl, 2000), running stop signs and red lights (James and Nahl, 2000; Tasca, 2000) and tailgating (Diekmann et al., 1996; Ellison-Potter et al., 2001; Turner et al., 1975).[6] Officers at the scene of a crash often identify the presence of aggressive driving by recognizing its correlation to speeding and other forms of traffic violations, as well as the level of driver frustration with the driving environment (traffic congestion) and expressions of anger (Neuman et al., 2003). Susceptibility to aggressive driving has been linked to numerous factors including fatigue, personality traits (including disregard for others, habitual/clinical behavior, disregard for the law, etc.), and stress (Stuster, 2004; Oz et al., 2010; Taylor and Dorn, 2006; Jovanovic et al., 2011; Alfonso et al., 2016).[7]

To study aggressive driving behavior and its impacts, several research efforts have used moving vehicles or fixed cameras (Kaysi and Abbany, 2007; Paleti et al., 2010; Tarko et al., 2011), while other studies have explored identifying aggressive driving behavior patterns with the use of driving simulation and survey data (Al-Shihabi and Mourant, 2007; Harder et al., 2008; AAA

[6] Interestingly, some authors have argued and categorized ‘aggressive driving’ and ‘road rage’ are synonymous, but there is clear distinction between them with road rage being associated with specific unlawful criminal behaviors (Shinar, 1998).

[7] It should be noted that some have argued that aggressive driving can be related self-defense and aggression as part of the human survival mechanism (Lang and Bradley, 2013).

Foundation for Traffic Safety, 2009; Philippe et al., 2009; Rong et al., 2011; Calvi et al., 2012; Joanisse et al., 2013; Ouimet et al., 2013; Sarwar et al., 2017). Other studies have investigated aggressive driving behavior from traditional crash data (Paleti et al., 2019), driving simulation studies (Sarwar et al., 2017; Fountas et al., 2019), observational studies of traffic (Tarko et al., 2011), and assessments of study/survey participants (Alonso et al., 2019; Nesbit and Conger, 2012; Berdoulat et al., 2012).

Table 5.1 provides a summary of factors that past research efforts have shown to be associated with aggressive driving. This table indicates that a wide variety of variables relating to driver characteristics and actions, and vehicle, roadway and crash characteristics have been found to be associated with aggressive driving. The intent of the current study is to use the findings of past research as a basis for understanding how aggressive driving affects injury-severity outcomes. Because past research has shown aggressive driving is induced, at least in part, by factors that vary over time (changes in traffic congestion, driver frustration and other temporally varying factors[8]), we are particularly interested in studying how the effect of variables influencing injury-severity outcomes in aggressive-driving crashes change over time. These results will also be benchmarked against the injury-severity outcomes in crashes that do not involve aggressive driving. To do this, focus will be directed toward an assessment of driver-injury severities in single-vehicle crashes with a statistical comparison between resulting injury severities in crashes involving aggressive and non-aggressive driving over time.[9]

---

[8] There is also the possibility that aggressive driving could be significantly affected by changes in economic conditions which would vary over time. For example, Abay and Mannering (2016) provided empirical evidence showing a statistically significant relationship between risk taking in driving and finances.

[9] The emphasis on single-vehicle crashes is to focus specifically on driver error without complicating the analysis by introducing the potential responses of other involved drivers in multivehicle crashes.

Table 5.1. Variables found to be statistically significant indicators of aggressive driving in past studies.

Variables	Findings
Driver characteristics	
Age	Young drivers (16-20 years) were found to be more likely to drive aggressively under the influence of alcohol relative older drivers (Paleti et al., 2010); less than 45 years old drivers were found to be more likely to drive aggressively (Beck et al., 2006; Vanlaar et al., 2008); less than 26 years old drivers were found to be more likely to involved in aggressive driving, 18-24 years drivers were found to be likely to exhibit more aggressive driving (Kaiser et al., 2016)
Gender	Male drivers were found to be more likely to involved in aggressive driving (Paleti et al., 2010; Berdoulat et al., 2013),
Behaviors	Urgency, lack of premeditation, and lack of perseverance heave led to aggressive driving (Loo, 1979; Stanford et al., 1996; Berdoulat et al., 2013). High anger drivers or personality characteristics with predisposed to aggression were found to be more likely to engage in aggressive and risky behaviors (Burns and Wilde, 1995; Jonah, 1997; Vavrik, 1997; Deffenbacher et al., 2000, 2003; Iversen and Rundmo, 2002; Wells-Parker et al., 2002). Fatigue and distraction also contribute to aggressive driving (Fountas et al., 2019)
Experiences	Inexperienced drivers with low mileage were found to report more irritation for others' direct hostility than more experienced drivers (Bjorklund, 2008); experienced drivers licensed for 6 or more years were found to exhibit less aggressive driving behavior (Sarwar et al., 2017)
Driver actions	
Traffic violations	Tailgating, flashing lights, honking at drivers blocking the driveway, waving in and out of traffic, cutting in front of other traffic, running red light/ stop sign (Shinar, 1998; Bjorklund, 2008; Tarko et al., 2011)
Temporal characteristics	
Peak hours	Time between 6 to 9 AM has found to have possible correlation with aggressive driving behavior. (Paleti et al., 2010), the frequency of aggressive driving is higher when the value of time is higher (rush hours) (Shinar and Compton, 2004), rushing to destination showed aggressive driving behavior (Sarwar et al., 2017)
Vehicle characteristics	
Vehicle type	Drivers of sport-utility vehicles and pickup trucks were found to be more likely to be involved in aggressive behaviors (Paleti et al., 2010)
Roadway attributes	
Speed limits	Low- (less than 50 km/hr) and high-speed limits (more than 90 km/hr) were found to be positively related to aggressive driving behaviors (Paleti et al., 2010)
Crash characteristics	
Fixed object crash	Due to aggressive driving, fixed object crashes have been found to be more likely to occur (Paleti et al., 2010)

The chapter begins with a description of the available crash data. This is followed by a presentation of the methodological approach and estimation results, which include statistical tests for differences in aggressive and non-aggressive driver-injury outcomes, and tests for temporal instability in aggressive driving injury severities. Model estimation results are then presented and discussed, with a comparison and discussion of marginal effects, and the chapter concludes with a summary of findings and their implications.

5.2 Data Description

Data available for this study were crashes reported in the Florida Crash Analysis Reporting system (these are all police-reported crashes). For the purposes of this study, crash data were gathered over the three-year period from January 1, 2015, to December 31, 2017. Crash data filtered from the Florida Crash Analysis Reporting data system were linked with a vehicle and person dataset based on crash identification numbers. A police-officer defined variable in the crash data indicating the driver's actions at the time of crash being "operated the motor vehicle in erratic/reckless and aggressive manner" was considered aggressive driving.[10] The resulting combined dataset provided detailed information about the crash, including roadway characteristics, as well as vehicle and person characteristics. The combined dataset was filtered for single-vehicle aggressive driving related crashes, which resulted in a total of 2,531 observations over the studied three-year period. While the single-vehicle non-aggressive driving related crashes resulted in 64,082 observations in the same analysis period.

[10] The determination of whether the vehicle is being operated in erratic/reckless and aggressive manner is made by police-officers, all of whom are trained to identify such behavior. The authors acknowledge that other definitions of aggressive driving may also be appropriate, depending on available data.

Information available in the data includes the resulting injury severity of the driver (no injury, possible injury, non-incapacitating injury, incapacitating injury and fatality), type of vehicle, driver actions (evasive maneuvers, etc.), driver information (age, gender, usage of safety equipment, influence of alcohol, drug use), information relating to the time and location of the crash, roadway class, road surface condition, weather and light conditions, type of vehicles, lane and shoulder widths, median width, location of harmful events, traffic volume, percent of trucks, and so on.

5.3 Methodology

To account for possible unobserved heterogeneity in the statistical analysis of injury-severity data, more recent research has focused on random parameter approaches (Milton et al., 2008; Eluru et al., 2008; Morgan and Mannering, 2011; Anastasopoulos and Mannering, 2011; Kim et al., 2013; Venkataraman et al., 2013; Islam and Hernandez, 2013; Behnood and Mannering, 2015), latent class models (Behnood et al., 2014; Cerwick et al., 2014; Shaheed and Gkritza, 2014; Yasmin et al., 2014), a combination of the two (Xiong and Mannering, 2013), random parameters with heterogenity in means and variances (Behnood and Mannering 2017a, 2017b; Seraneeprakarn et al., 2017; Behnood and Mannering, 2019), and Markov switching models (Malyshkina and Mannering, 2009; Xiong et al., 2014). Savolainen et al. (2011), Mannering and Bhat (2014) and Mannering et al. (2016) provide a review of the statistical appraoches used to study crash-injury severities.

In this study, a random parameters logit model that accounts for possible heterogeneity in the means and variances of the random parameters is used to address possible unobserved heterogeneity. The injury severity of drivers in single-vehicle aggressive and non-aggressive crashes is considered with possible injury outcomes of no injury, minor injury (possible injury and non-incapacitating injury) and severe injury (incapacitating injury and fatality). Following recent work, the modeling approach starts by defining a function that determines injury-severity,

$$S_{kn} = \boldsymbol{\beta}_k \mathbf{X}_{kn} + \varepsilon_{kn} \tag{5.1}$$

where S_{kn} is an injury-severity function determining the probability of injury-severity outcome k in crash n, $\mathbf{X}_{kn}$ is a vector of explanatory variables that affect aggressive and non-aggressive drivers' injury-severity level k, $\boldsymbol{\beta}_k$ is a vector of estimable parameters, and ε_{kn} is the error term. If this error term is assumed to be generalized extreme value distributed, a standard multinomial logit model results as (McFadden, 1981),

$$P_n(k) = \frac{EXP\left[\boldsymbol{\beta}_k \mathbf{X}_{in}\right]}{\sum_{\forall K} EXP\left(\boldsymbol{\beta}_k \mathbf{X}_{In}\right)} \tag{5.2}$$

where $P_n(k)$ is the probability that aggressive and non-aggressive driving related crash n that will result in driver-injury severity outcome k and K is the set of the three possible injury-severity outcomes. To allow the possibility of one or more parameter estimates in the vector $\boldsymbol{\beta}_k$ vary across crash observations Equation 2 can be rewritten as (Train, 2009; Washington et al., 2020)

$$P_n(k) = \int \frac{EXP\left(\boldsymbol{\beta}_k \mathbf{X}_{kn}\right)}{\sum_{\forall K} EXP\left(\boldsymbol{\beta}_k \mathbf{X}_{kn}\right)} f\left(\boldsymbol{\beta}_k \mid \boldsymbol{\varphi}_k\right) d\boldsymbol{\beta}_k \tag{5.3}$$

where $f(\boldsymbol{\beta}_k|\boldsymbol{\varphi}_k)$ is the density function of $\boldsymbol{\beta}_k$ and $\boldsymbol{\varphi}_k$ is a vector of parameters describing the density function (mean and variance), and all other terms are as previously defined.

To account for the possibility of unobserved heterogeneity in the means and variances of parameters, let $\boldsymbol{\beta}_{kn}$ be a vector of estimable parameters that varies across crashes defined as (Mannering et al., 2016; Seraneeprakarn et al., 2017; Behnood and Mannering, 2017b; Waseem et al., 2019; Alnawmasi and Mannering, 2019; Behnood and Mannering, 2019; Washington et al., 2020):

$$\boldsymbol{\beta}_{kn} = \beta + \Theta_{kn}\mathbf{Z}_{kn} + \sigma_{kn}EXP\left(\Psi_{kn}\mathbf{W}_{kn}\right)\,\nu_{kn} \tag{5.4}$$

where β is the mean parameter estimate across all crashes, $\mathbf{Z}_{kn}$ is a vector of crash-specific explanatory variables that captures heterogeneity in the mean that affects aggressive and non-aggressive drivers' injury-severity level k, $\boldsymbol{\Theta}_{kn}$ is a corresponding vector of estimable parameters, $\mathbf{W}_{kn}$ is a vector of crash-specific explanatory variables that captures heterogeneity in the standard deviation σ_{kn} with corresponding parameter vector Ψ_{kn}, and ν_{kn} is a disturbance term.

During model estimation, numerous density functions were empirically evaluated for the term $f(\boldsymbol{\beta}_k|\boldsymbol{\varphi}_k)$. None were found to be statistically superior to the normal distribution, so this was used in all model estimations (this finding is consistent with past work including Milton et al., 2008; Alnawmasi et al., 2019 and others). All model estimations used simulated maximum likelihood with 1,000 Halton draws (McFadden and Train, 2000; Bhat, 2001; Train, 2009). Marginal effects were computed to determine the effect of explanatory variables on injury-severity probabilities. The marginal effect provides the effect that a one-unit increase in an explanatory variable has on the injury-outcome probabilities. The average marginal effect over all crash observations will be reported.

5.4 Likelihood Ratio Tests

There is an extensive body of literature that suggests that the effect of factors determining injury severity may change over time (Mannering, 2018). For example, Behnood and Mannering (2015) found that the effect that roadway characteristics, vehicle characteristics, and driver characteristics had on resulting driver-injury severities in Chicago varied significantly from one year to the next from 2004 to 2012. Subsequent work from these authors (Behnood and Mannering, 2016), showed similar temporal instability in pedestrian injuries resulting from vehicle accidents in Chicago. In other work, Alnawmasi and Mannering (2019) found similar results for motorcyclist injuries in Florida, with temporal instability observed in data from 2012 to 2016; Behnood and Mannering (2019) found temporal instability among injuries induced by crashes involving truck in Los Angeles from 2010-17; and Islam et al. (2019) found temporal instability in Florida work-zone crashes from 2012-17.

Given this, tests are not only conducted for differences between injury outcomes in crashes involving aggressive and non-aggressive driving, but also for temporal instability. This is done by running a series of likelihood ratio tests. To begin, for each year in the data (2015, 2016, and 2017) tests were conducted comparing aggressive and non-aggressive injury-severity outcomes. The test statistic is,

$$X_t^2 = -2\left[-LL\left(\boldsymbol{\beta}_{combined,t}\right) - LL\left(\boldsymbol{\beta}_{non-aggressive,t}\right) - LL\left(\boldsymbol{\beta}_{aggressive,t}\right)\right] \tag{5.5}$$

where, $LL(\boldsymbol{\beta}_{combined,t})$ is the log-likelihood at the convergence of the model that used all of the available aggressive and non-aggressive driving data in year t (either years 2015, 2016 or 2017), $LL(\boldsymbol{\beta}_{non\text{-}aggressive,t})$ is the log-likelihood at convergence of the model based on non-aggressive driving data only in year t, and $LL(\boldsymbol{\beta}_{aggressive,t})$ is the log-likelihood at the convergence of a model based on aggressive driving data only in year t. For the years 2015, 2016 and 2017 the model estimates gave an X^2 values of 171.256, 51.31, and 152.75, respectively. These values are all χ^2 distributed with 20 degrees of freedom give us 99.99% confidence that the null hypothesis that the non-aggressive and aggressive driving parameters are equal, can be rejected.

Next, the temporal stability of aggressive and non-aggressive injury-severity outcomes is tested with the likelihood ratio test in this case,

$$X_g^2 = -2\left[LL(\boldsymbol{\beta}_{2015-17,g}) - LL(\boldsymbol{\beta}_{2015,g}) - LL(\boldsymbol{\beta}_{2016,g}) - LL(\boldsymbol{\beta}_{2017,g})\right] \quad (5.6)$$

where, $LL(\boldsymbol{\beta}_{2015\text{-}17,g})$ is the log-likelihood at the convergence of the driver injury-severity model that used all data 2015 to 2017 for driver group g (either aggressive or non-aggressive driving data), $LL(\boldsymbol{\beta}_{2015,g})$ is the log-likelihood at convergence of the model using only 2015 data for driver group g, $LL(\boldsymbol{\beta}_{2016,g})$ is the log-likelihood at the convergence of the model using only 2016 data for driver group g, and $LL(\boldsymbol{\beta}_{2017,g})$ is the log-likelihood at the convergence of the model using only 2017 data for driver group g. For crashes involving aggressive drivers, model estimates gave an X^2 of 42.02 which is χ^2 distributed with 23 degrees of freedom (the number of parameters found to be statistically significant in the model using all data years, 2015-17). This χ^2 value gives 99% confidence that the null hypothesis that the parameters are equal parameters over these three years (2015, 2016 and 2017) can be rejected. For crashes involving non-aggressive drivers, model estimates gave an X^2 of 26.73 which is χ^2 distributed with 19 degrees of freedom (the number of parameters found to be statistically significant in the model using all data years, 2015-17). This χ^2 value gives 89% confidence that the null hypothesis that the parameters are equal parameters over these three years (2015, 2016 and 2017) can be rejected. These tests suggest temporal instability is clearly present in aggressive driving crashes, and to a lesser extent in non-aggressive driving crashes.[11]

5.5 Model Estimation Results

Figure 5.1 illustrates the percent distribution of severe, minor and injury crashes for aggressive and non-aggressive drivers over the three-year analysis period (2015-17). This figure

[11] Multiple combinations of years were also tested as part of the temporal instability tests, but likelihood ratio tests indicate that the separate-year models where statistically justified.

shows that, while there is not much variation in the aggregate injury-severity totals over time, there is a noticeable difference between aggressive and non-aggressive driver injuries with aggressive drivers being much more involved in crashes resulting in injury, particularly severe injury. Table 5.2 provides the summary statistics of all explanatory variables found to be statistically significant in one or more of the six models estimated (aggressive and non-aggressive models for each of the three years analyzed).

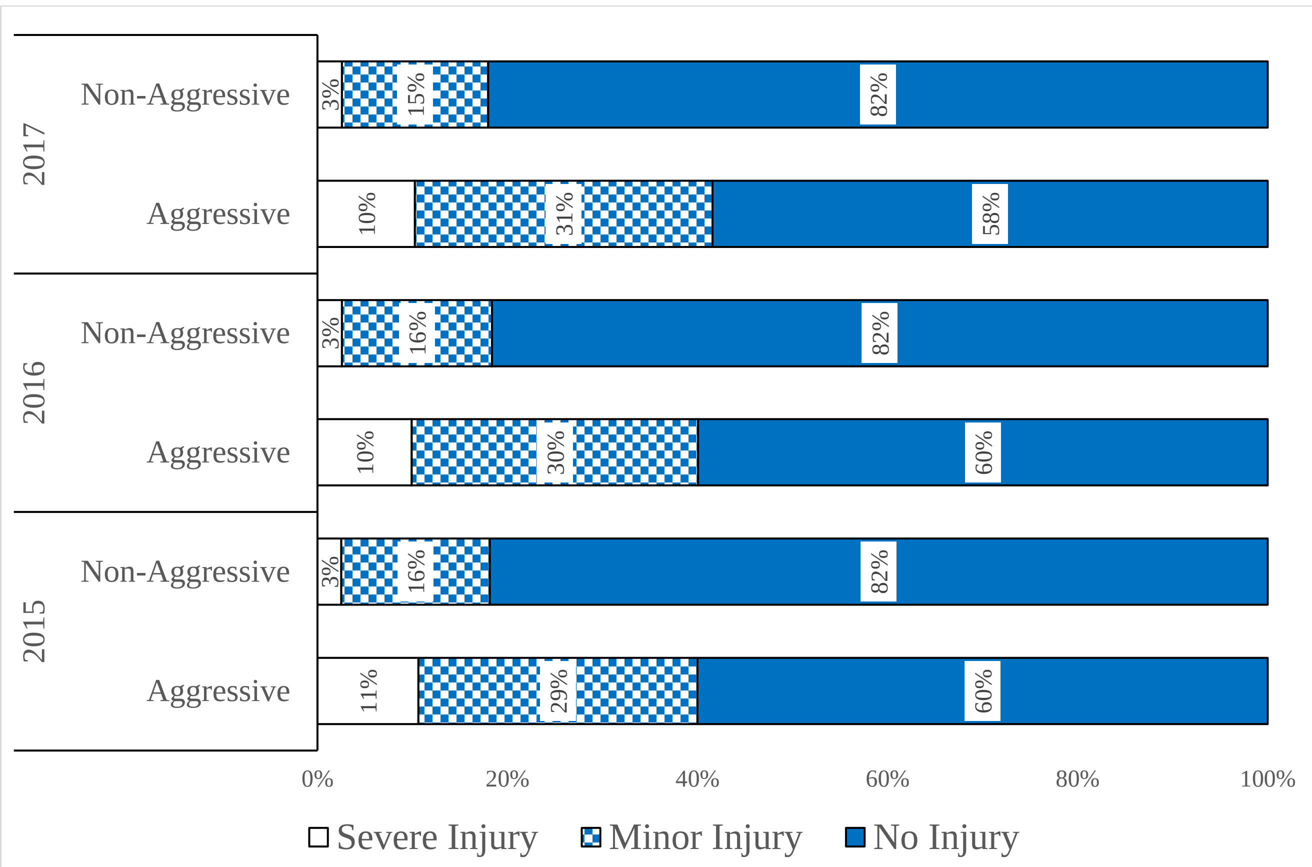


Figure 5.1. Driver Injury Severity for Aggressive and Non-Aggressive Driving over the Years: 2015 – 2017.

Table 5.2. Aggressive-driver descriptive statistics (non-aggressive driver statistics in parentheses) of variables found to significantly influence severity-outcome probabilities.

Variable	2015		2016		2017	
	Mean	**Std. Dev.**	**Mean**	**Std. Dev.**	**Mean**	**Std. Dev.**
Spatial characteristics						
District 2 indicator (1 if crash occurred in District 2, 0 otherwise)	0.129 (0.164)	0.336 (0.370)	0.144 (0.175)	0.351 (0.380)	0.128 (0.172)	0.334 (0.377)
District 5 indicator (1 if crash occurred in District 5, 0 otherwise)	0.192 (0.154)	0.394 (0.361)	0.176 (0.155)	0.381 (0.362)	0.178 (0.163)	0.382 (0.369)
District 6 indicator (1 if crash occurred in District 6, 0 otherwise)	0.092 (0.154)	0.290 (0.361)	0.110 (0.142)	0.313 (0.349)	0.106 (0.150)	0.308 (0.357)
District 7 indicator (1 if crash occurred in District 7, 0 otherwise)	0.154 (0.134)	0.361 (0.341)	0.123 (0.131)	0.329 (0.338)	0.140 (0.127)	0.347 (0.333)
Roadway characteristics						
Urban principal arterial indicator (1 if crash occurred on urban principal arterials, 0 otherwise)	0.129 (0.181)	0.336 (0.385)	0.129 (0.180)	0.335 (0.384)	0.147 (0.134)	0.354 (0.341)
Dry surface indicator (1 if road surface condition was dry, 0 otherwise)	0.836 (0.720)	0.370 (0.448)	0.831 (0.750)	0.374 (0.433)	0.842 (0.747)	0.364 (0.434)
Wet surface indicator (1 if road surface condition was wet, 0 otherwise)	0.111 (0.157)	0.315 (0.364)	0.117 (0.131)	0.322 (0.337)	0.102 (0.130)	0.303 (0.337)
Straight roadway section indicator (1 if straight section of the roadway, 0 otherwise)	0.796 (0.907)	0.402 (0.289)	0.825 (0.912)	0.379 (0.283)	0.812 (0.911)	0.390 (0.283)
Curved segment indicator (1 if roadway curves to the right or left of travel direction, 0 otherwise)	0.193 (0.089)	0.394 (0.285)	0.170 (0.085)	0.376 (0.279)	0.182 (0.085)	0.386 (0.279)
Narrow shoulder width indicator (1 if shoulder width is below 4 ft, 0 otherwise)	0.149 (0.225)	(0.356) (0.418)	0.165 (0.228)	0.372 (0.419)	0.128 (0.107)	0.334 (0.309)
Wider shoulder width indicator (1 if shoulder width is between 8 to 12 feet, 0 otherwise)	0.047 (0.169)	0.212 (0.375)	0.049 (0.162)	0.217 (0.369)	0.031 (0.085)	0.175 (0.279)
Crash characteristics						
Collision with roadside fixed object indicator (1 if collided with roadside fixed object as the first harmful event, 0 otherwise)	0.600 (0.263)	0.489 (0.440)	0.562 (0.265)	0.496 (0.441)	0.547 (0.260)	0.497 (0.438)
Collision with non-fixed object indicator (1 if collided with non-fixed object as the first harmful event, 0 otherwise)	0.231 (0.595)	0.421 (0.490)	0.269 (0.596)	0.444 (0.490)	0.279 (0.607)	0.448 (0.488)

Non-colliding object indicator (1 if first harmful event was identified as non-colliding object such as a rollover, etc., 0 otherwise)	0.111 (0.095)	0.315 (0.293)	0.107 (0.093)	0.310 (0.291)	0.126 (0.088)	0.332 (0.284)
On-road as harmful event location indicator (1 if harmful event occurred inside the roadway, 0 otherwise)	0.268 (0.680)	0.443 (0.466)	0.248 (0.684)	0.432 (0.464)	0.278 (0.683)	0.448 (0.465)
Off-road as harmful event location indicator (1 if harmful event occurred outside the roadway, 0 otherwise)	0.464 (0.160)	0.498 (0.366)	0.474 (0.164)	0.499 (0.370)	0.489 (0.167)	0.499 (0.373)
Driver characteristics						
Exceeding speed limit by more than 10 mi/h indicator (1 if travel exceeded the speed limit by more than 10 mi/h, 0 otherwise)	0.448 (0.016)	0.497 (0.127)	0.458 (0.015)	0.498 (0.125)	0.413 (0.015)	0.492 (0.123)
Suspected Alcohol use indicator (1 if alcohol use is suspected in the crash involved driver, 0 otherwise)	0.318 (0.010)	0.466 (0.100)	0.268 (0.008)	0.443 (0.093)	0.275 (0.010)	0.446 (0.100)
Younger aged driver indicator (1 if driver age below 30, 0 otherwise)	0.655 (0.337)	0.475 (0.472)	0.618 (0.338)	0.485 (0.473)	0.608 (0.330)	0.488 (0.470)
Middle aged driver indicator (1 if driver age between 30 to 49 years, 0 otherwise)	0.271 (0.354)	0.444 (0.478)	0.271 (0.350)	0.444 (0.477)	0.286 (0.355)	0.452 (0.478)
Restraint usage indicator (1 if shoulder and lap belt used, 0 otherwise)	0.512 (0.893)	0.499 (0.308)	0.566 (0.902)	0.495 (0.296)	0.567 (0.906)	0.495 (0.291)
Zero blood alcohol content (BAC) indicator (1 if BAC is zero in the crash involved driver, 0 otherwise)	0.872 (0.997)	0.333 (0.054)	0.899 (0.997)	0.300 (0.051)	0.904 (0.996)	0.293 (0.060)
Vehicle characteristics						
Passenger car involvement indicator (1 if passenger car being involved in the crash, 0 otherwise)	0.620 (0.549)	0.485 (0.497)	0.642 (0.544)	0.479 (0.498)	0.610 (0.532)	0.487 (0.498)
Sport Utility Vehicle (SUV) involvement indicator (1 if SUV being involved in the crash, 0 otherwise)	0.144 (0.139)	0.351 (0.346)	0.123 (0.148)	0.329 (0.356)	0.137 (0.155)	0.344 (0.362)
Traffic characteristics						
Low traffic condition indicator (1 if AADT is below 4,000 veh/day, 0 otherwise)	0.192 (0.280)	0.394 (0.449)	0.211 (0.288)	0.408 (0.453)	0.137 (0.125)	0.344 (0.331)
Temporal characteristics						
Weekday indicator (1 if crash occurred during the weekdays, 0 otherwise)	0.640 (0.721)	0.479 (0.448)	0.598 (0.727)	0.490 (0.445)	0.643 (0.726)	0.478 (0.445)
Weekend indicator (1 if crash occurred during the weekend, 0 otherwise)	*0.359 (0.278)*	*0.479 (0.448)*	*0.401 (0.272)*	*0.490 (0.445)*	*0.356 (0.273)*	*0.478 (0.445)*
Early morning indicator (1 if crash occurred between midnight to 6AM, 0 otherwise)	*0.310 (0.141)*	*0.462 (0.348)*	*0.329 (0.144)*	*0.470 (0.351)*	*0.340 (0.142)*	*0.473 (0.349)*

Model estimation results for aggressive drivers are presented in Tables 5.3, 5.4 and 5.5 for the years 2015, 2016 and 2017, respectively. The models have reasonably good overall statistical fit with ρ^2 values of 0.229, 0.246 and 0.242 for 2015, 2016 and 2017, respectively. Note that for the 2015, 2016 and 2017 models, the constant term specific to minor injury was found to be the only statistically significant random parameter, and in all cases this parameter had statistically significant heterogeneity in the mean and variance. For all three years, it was found that the mean varied by whether the first harmful event was identified as non-colliding object (such as a rollover, etc.). In all three years, having the first harmful event being non-colliding object increased the mean of the random parameter making minor injury more likely and other injury levels less likely. In 2015 (Table 5.3) the variance of the constant for minor injury was a function of a low-traffic volume indicator, with traffic volumes less than 4000 veh/day increasing the minor-injury constant variance and reflecting high variability among low-volume roads for aggressive drivers. In 2016 (Table 5.4) the variance was a function of the dry-surface indicator, with crashes occurring on dry roads increasing the variance of the minor-injury constant. Finally, in 2017 (Table 5.5) the minor-injury constant variance was a function of the weekend indicator, with greater variance in the minor-injury constant for crashes that occurred on weekends for aggressive drivers.

Model estimation results for non-aggressive drivers are presented in Tables 5.6, 5.7 and 5.8 for the years 2015, 2016 and 2017, respectively. The models have noticeably better fit than their aggressive driving counterparts with ρ^2 values of 0.559, 0.549 and 0.557 for 2015, 2016 and 2017, respectively. As was the case for the aggressive drivers, for the 2015, 2016 and 2017 models, the constant term specific to minor injury was found to be the only statistically significant random parameter and this parameter again had statistically significant heterogeneity in the mean and variance. For all three years, it was found that the mean varied by the driver's blood alcohol

Table 5.3. Model results of mixed logit with heterogeneity in means and variance for aggressive driving in single-vehicle crashes in Florida 2015.

Variable*	Parameter Estimates	t-stat	Marginal Effects		
			No Injury	Minor Injury	Severe Injury
Constant [SI]	-1.277	-4.29			
Random parameter (normally distributed)					
Constant [MI] (*Standard deviation of parameter distribution*)	-3.603 (*3.949*)	-2.23 (*2.24*)			
Heterogeneity in the mean of random parameter					
Constant [MI]: Non-colliding object indicator (1 if first harmful event was identified as non-colliding object such as a rollover, etc., 0 otherwise)	3.417	2.31			
Heterogeneity in the variance of random parameter					
Constant [MI]: Low traffic volume (1 if traffic volume less than 4000 veh/day, 0 otherwise)	0.670	2.17			
Spatial characteristics					
District 7 indicator (1 if crash occurred in District 7, 0 otherwise) [NI]	-0.518	-1.81	-0.0117	0.0046	0.0071
Roadway characteristics					
Urban principal arterial indicator (1 if crash occurred on urban principal arterials, 0 otherwise) [SI]	0.674	1.99	-0.0081	-0.0010	0.0092
Straight roadway section indicator (1 if straight section of the roadway, 0 otherwise) [SI]	-0.558	-1.94	0.0306	0.0043	-0.0349
Wet road surface indicator (1 if crash occurred on wet surface, 0 otherwise) [NI]	-0.539	-1.67	-0.0086	0.0033	0.0053
Crash characteristics					
Collision with roadside fixed object indicator (1 if collided with roadside fixed object as the first harmful event, 0 otherwise) [MI]	1.018	1.68	-0.0356	0.0426	-0.0070
Driver characteristics					
Exceeding speed limit by more than 10 mi/h indicator (1 if travel exceeded the speed limit by more than 10 mi/h, 0 otherwise) [MI]	1.535	2.12	-0.0431	0.0521	-0.0089
Suspected Alcohol use indicator (1 if alcohol use is suspected in the crash involved driver, 0 otherwise) [SI]	0.932	3.76	-0.0289	-0.0045	0.0334
Restraint usage indicator (1 if shoulder and lap belt used, 0 otherwise) [NI]	1.401	5.48	0.0781	-0.0411	-0.0370
Number of observations			885		

Number of estimated parameters	14
Log-likelihood at zero	-972.27
Log-likelihood at convergence	-749.08
$\rho^2 = 1 - LL(\boldsymbol{\beta})/LL(0)$	0.229

*SI = Severe Injury; MI = Minor Injury; NI = No Injury

Table 5.4. Model results of mixed logit with heterogeneity in means and variance for aggressive driving in single-vehicle crashes in Florida 2016.

Variable*	**Parameter Estimates**	**t-stat**	**Marginal Effects**		
			No Injury	**Minor Injury**	**Severe Injury**
Constant [SI]	-1.631	-6.25			
Random parameter (normally distributed)					
Constant [MI] (*Standard deviation of parameter distribution*)	-2.347 (*2.535*)	-2.43 (*1.84*)			
Heterogeneity in the mean of random parameter					
Constant [MI]: Non-colliding object indicator (1 if first harmful event was identified as non-colliding object such as a rollover, etc., 0 otherwise)	2.446	2.20			
Heterogeneity in the variance of random parameter					
Constant [MI]: Dry surface indicator (1 if crash occurred on dry road surface, 0 otherwise)	0.752	2.14			
Spatial characteristics					
District 7 indicator (1 if crash occurred in District 7, 0 otherwise) [MI]	-1.822	-1.88	0.0098	-0.0124	0.0025
Roadway characteristics					
Harmful event location indicator (1 if harmful event occurred inside the roadway, 0 otherwise) [SI]	0.715	2.27	-0.0133	-0.0024	0.0158
Curved segment indicator (1 if roadway curves to the right or left of travel direction, 0 otherwise) [NI]	-0.640	-2.14	-0.0157	0.0058	0.0099
Wet road surface indicator (1 if crash occurred on wet surface, 0 otherwise) [MI]	1.374	1.86	-0.0144	0.0172	-0.0028
Crash characteristics					
Collision with non-fixed object indicator (1 if collided with non-fixed object as the first harmful event, 0 otherwise) [SI]	-1.571	-3.73	0.0122	0.0016	-0.0139
Driver characteristics					
Exceeding speed limit by more than 10 mi/h indicator (1 if travel exceeded the speed limit by more than 10 mi/h, 0 otherwise) [NI]	-0.765	-3.11	-0.0482	0.0188	0.0294
Suspected Alcohol use indicator (1 if alcohol use is suspected in the crash involved driver, 0 otherwise) [SI]	0.571	1.94	-0.0115	-0.0020	0.0135
Restraint usage indicator (1 if shoulder and lap belt used, 0 otherwise) [NI]	1.587	5.77	0.0893	-0.0518	-0.0376

Number of observations	826
Number of estimated parameters	14
Log-likelihood at zero	-907.45
Log-likelihood at convergence	-684.33
$\rho^2 = 1 - LL(\boldsymbol{\beta})/LL(0)$	0.246

*SI = Severe Injury; MI = Minor Injury; NI = No Injury

Table 5.5. Model results of mixed logit with heterogeneity in means and variance for aggressive driving in single-vehicle crashes in Florida 2017.

Variable*	Parameter Estimates	t-stat	Marginal Effects		
			No Injury	Minor Injury	Severe Injury
Constant [SI]	-2.063	-7.86			
Random parameter (normally distributed)					
Constant [MI] (*Standard deviation of parameter distribution*)	-3.979 (*3.425*)	-2.61 (*2.41*)			
Heterogeneity in the mean of random parameter					
Constant [MI]: Non-colliding object indicator (1 if first harmful event was identified as non-colliding object such as a rollover, etc., 0 otherwise)	2.325	2.36			
Heterogeneity in the variance of random parameter					
Constant [MI]: Weekend indicator (1 if crash occurred in the weekend, 0 otherwise)	0.408	1.82			
Spatial characteristics					
District 6 indicator (1 if crash occurred in District 6, 0 otherwise) [MI]	-1.708	-1.81	0.0092	-0.0107	0.0015
Roadway characteristics					
Harmful event location indicator (1 if harmful event occurred inside the roadway, 0 otherwise) [SI]	0.942	2.95	-0.0202	-0.0046	0.0248
Curved segment indicator (1 if roadway curves to the right or left of travel direction, 0 otherwise) [NI]	-0.887	-2.91	-0.0244	0.0094	0.0151
Crash characteristics					
Collision with non-fixed object indicator (1 if collided with non-fixed object as the first harmful event, 0 otherwise) [SI]	-1.683	-3.78	0.0128	0.0020	-0.0148
Driver characteristics					
Middle aged driver indicator (1 if driver age between 30 to 49 years, 0 otherwise) [SI]	0.529	1.90	-0.0121	-0.0026	0.0146
Exceeding speed limit by more than 10 mi/h indicator (1 if travel exceeded the speed limit by more than 10 mi/h, 0 otherwise) [NI]	-1.053	-4.14	-0.0661	0.0272	0.0389
Blood Alcohol Content indicator (1 if BAC is zero in the crash involved driver, 0 otherwise) [MI]	1.988	2.01	-0.1183	0.1433	-0.0250
Restraint usage indicator (1 if shoulder and lap belt used, 0 otherwise) [NI]	0.961	3.86	0.0641	-0.0354	-0.0287
Number of observations			820		

Number of estimated parameters	14
Log-likelihood at zero	-900.86
Log-likelihood at convergence	-682.56
$\rho^2 = 1 - LL(\boldsymbol{\beta})/LL(0)$	0.242

*SI = Severe Injury; MI = Minor Injury; NI = No Injury

Table 5.6. Model results of mixed logit with heterogeneity in means and variance for non-aggressive driving in single-vehicle crashes in Florida 2015.

Variable*	Parameter Estimates	t-stat	Marginal Effects		
			No Injury	Minor Injury	Severe Injury
Constant [SI]	-0.902	-9.10			
Random parameter (normally distributed)					
Constant [MI] (*Standard deviation of parameter distribution*)	-0.112 (*1.828*)	-0.19 (*10.13*)			
Heterogeneity in the mean of random parameter					
Constant [MI]: Zero blood alcohol content (BAC) indicator (1 if BAC is zero in the crash involved driver, 0 otherwise)	1.027	1.78			
Heterogeneity in the variance of random parameter					
Constant [MI]: Early morning indicator (1 if crash occurred between midnight to 6AM, 0 otherwise)	0.285	5.25			
Spatial characteristics					
District 2 indicator (1 if crash occurred in District 2, 0 otherwise) [SI]	-0.385	-2.88	0.0010	0.0003	-0.0012
District 6 indicator (1 if crash occurred in District 6, 0 otherwise) [MI]	-0.619	-6.59	0.0055	-0.0058	0.0003
Roadway characteristics					
Harmful event location indicator (1 if harmful event occurred outside the roadway, 0 otherwise) [SI]	0.522	6.09	-0.0089	0.0095	-0.0006
Harmful event location indicator (1 if harmful event occurred inside the roadway, 0 otherwise) [NI]	0.726	10.85	0.0407	-0.0334	-0.0073
Straight segment indicator (1 if roadway straight of travel direction, 0 otherwise) [MI]	-0.873	-8.91	0.0590	-0.0624	0.0034
Wide shoulder width indicator (1 if shoulder width is 8 to 12 ft, 0 otherwise) [NI]	-0.788	-11.47	-0.0163	0.0129	0.0034
Wet surface indicator (1 if road surface condition was wet, 0 otherwise) [NI]	-0.361	-5.29	-0.0062	0.0050	0.0013
Crash characteristics					
Collision with fixed object indicator (1 if collided with fixed object as the first harmful event, 0 otherwise) [SI]	0.588	5.81	-0.0041	-0.0010	0.0051
Driver characteristics					
Middle aged driver indicator (1 if driver age between 30 to 49 years, 0 otherwise) [NI]	0.218	3.94	0.0069	-0.0056	-0.0013
Restraint usage indicator (1 if shoulder and lap belt used, 0 otherwise) [NI]	2.687	24.53	0.2052	-0.1751	-0.0301
Vehicle characteristics					

Passenger car involvement indicator (1 if passenger car being involved in the crash, 0 otherwise) [SI]	-1.001	-9.23	0.0053	0.0011	-0.0063
Sport Utility Vehicle (SUV) involvement indicator (1 if SUV being involved in the crash, 0 otherwise) [MI]	0.301	3.66	-0.0034	0.0036	-0.0002
Traffic characteristics					
Low traffic condition indicator (1 if AADT is below 4,000 veh/day, 0 otherwise) [SI]	0.213	2.09	-0.0011	-0.0003	0.0014
Temporal characteristics					
Weekday indicator (1 if crash occurred during the weekdays, 0 otherwise) [MI]	- 0.218	-3.40	0.0118	-0.0124	0.0007
Number of observations			21,100		
Number of estimated parameters			20		
Log-likelihood at zero			-23180.72		
Log-likelihood at convergence			-10221.75		
$\rho^2 = 1 - LL(\boldsymbol{\beta})/LL(0)$			0.559		

*SI = Severe Injury; MI = Minor Injury; NI = No Injury

Table 5.7. Model results of mixed logit with heterogeneity in means and variance for non-aggressive driving in single-vehicle crashes in Florida 2016.

Variable*	**Parameter Estimates**	**t-stat**	**Marginal Effects**		
			No Injury	**Minor Injury**	**Severe Injury**
Constant [SI]	-0.756	-7.05			
Random parameter (normally distributed)					
Constant [MI] (*Standard deviation of parameter distribution*)	-0.062 (*1.940*)	-0.11 (*10.36*)			
Heterogeneity in the mean of random parameter					
Constant [MI]: Zero blood alcohol content (BAC) indicator (1 if BAC is zero in the crash involved driver, 0 otherwise)	0.974	1.70			
Heterogeneity in the variance of random parameter					
Constant [MI]: early morning (1 if crash occurred between midnight to 6AM, 0 otherwise)	0.217	4.27			
Spatial characteristics					
District 2 indicator (1 if crash occurred in District 2, 0 otherwise) [SI]	-0.354	-2.78	0.0010	0.0003	-0.0013
District 6 indicator (1 if crash occurred in District 6, 0 otherwise) [MI]	-0.691	-6.90	0.0056	-0.0059	0.0003
Roadway characteristics					
Harmful event location indicator (1 if harmful event occurred outside the roadway, 0 otherwise) [SI]	0.559	6.38	-0.0098	0.0105	-0.0008
Harmful event location indicator (1 if harmful event occurred inside the roadway, 0 otherwise) [NI]	0.814	11.67	0.0456	-0.0374	-0.0082
Straight segment indicator (1 if roadway straight of travel direction, 0 otherwise) [MI]	-0.701	-6.99	0.0476	-0.0502	0.0026
Narrow shoulder width indicator (1 if shoulder width is below 4 ft, 0 otherwise) [NI]	0.119	1.77	0.0024	-0.0019	-0.0005
Wet surface indicator (1 if road surface condition was wet, 0 otherwise) [NI]	-0.408	-5.64	-0.0060	0.0047	0.0013
Crash characteristics					
Collision with fixed object indicator (1 if collided with fixed object as the first harmful event, 0 otherwise) [SI]	0.773	7.72	-0.0060	-0.0014	0.0074
Driver characteristics					
Younger aged driver indicator (1 if driver age below 30 years, 0 otherwise) [SI]	-0.291	-2.89	0.0017	0.0004	-0.0021
Middle aged driver indicator (1 if driver age between 30 to 49 years, 0 otherwise) [NI]	0.226	4.02	0.0072	-0.0057	-0.0015
Restraint usage indicator (1 if shoulder and lap belt used, 0 otherwise) [NI]	2.685	24.32	0.2104	- 0.1773	- 0.0331

Vehicle characteristics					
Passenger car involvement indicator (1 if passenger car being involved in the crash, 0 otherwise) [SI]	-0.593	-5.56	0.0040	0.0007	-0.0047
Sport Utility Vehicle (SUV) involvement indicator (1 if SUV being involved in the crash, 0 otherwise) [MI]	0.320	4.02	-0.0040	0.0042	-0.0002
Traffic characteristics					
Low traffic condition indicator (1 if AADT is below 4,000 veh/day, 0 otherwise) [SI]	0.396	3.84	-0.0024	-0.0005	0.0030
Temporal characteristics					
Weekday indicator (1 if crash occurred during the weekdays, 0 otherwise) [MI]	-0.137	-2.09	0.0075	-0.0079	0.0004
Number of observations			21,144		
Number of estimated parameters			21		
Log-likelihood at zero			-23442.19		
Log-likelihood at convergence			-10551.99		
$\rho^2 = 1 - LL(\boldsymbol{\beta})/LL(0)$			0.549		

*SI = Severe Injury; MI = Minor Injury; NI = No Injury

Table 5.8. Model results of mixed logit with heterogeneity in means and variance for non-aggressive driving in single-vehicle crashes in Florida 2017.

Variable*	**Parameter Estimates**	**t-stat**	**Marginal Effects**		
			No Injury	**Minor Injury**	**Severe Injury**
Constant [SI]	-1.017	-9.17			
Random parameter (normally distributed)					
Constant [MI]	-0.474	-0.85			
(*Standard deviation of parameter distribution*)	(*1.814*)	(*9.93*)			
Heterogeneity in the mean of random parameter					
Constant [MI]: Zero blood alcohol content (BAC) indicator (1 if BAC is zero in the crash involved driver, 0 otherwise)	1.361	2.48			
Heterogeneity in the variance of random parameter					
Constant [MI]: early morning (1 if crash occurred between midnight to 6AM, 0 otherwise)	0.241	4.61			
Spatial characteristics					
District 5 indicator (1 if crash occurred in District 5, 0 otherwise) [SI]	0.266	2.41	-0.0010	-0.0003	0.0013
District 6 indicator (1 if crash occurred in District 6, 0 otherwise) [MI]	-0.730	-7.70	0.0064	-0.0067	0.0003
Roadway characteristics					
Harmful event location indicator (1 if harmful event occurred outside the roadway, 0 otherwise) [SI]	0.502	5.88	0.0446	-0.0368	-0.0078
Harmful event location indicator (1 if harmful event occurred inside the roadway, 0 otherwise) [NI]	0.790	11.80	-0.0091	0.0098	-.0007
Straight segment indicator (1 if roadway straight of travel direction, 0 otherwise) [MI]	-0.651	-6.82	0.0451	-0.0475	0.0024
Narrow shoulder width indicator (1 if shoulder width is below 4 ft, 0 otherwise) [NI]	0.213	2.50	0.0023	-0.0015	-0.0007
Wet surface indicator (1 if road surface condition was wet, 0 otherwise) [NI]	-0.352	-4.98	-0.0052	0.0041	0.0011
Crash characteristics					
Collision with fixed object indicator (1 if collided with fixed object as the first harmful event, 0 otherwise) [SI]	0.779	7.91	-0.0060	-0.0014	0.0074
Driver characteristics					
Younger aged driver indicator (1 if driver age below 30 years, 0 otherwise) [SI]	-0.419	-4.06	0.0022	0.0006	-0.0027
Middle aged driver indicator (1 if driver age between 30 to 49 years, 0 otherwise) [NI]	0.256	4.66	0.0083	-0.0066	-0.0017
Restraint usage indicator (1 if shoulder and lap belt used, 0 otherwise) [NI]	2.574	24.03	0.2066	-0.1737	-0.0329

Vehicle characteristics					
Passenger car involvement indicator (1 if passenger car being involved in the crash, 0 otherwise) [SI]	-0.745	-7.05	0.0044	0.0008	-0.0052
Sport Utility Vehicle (SUV) involvement indicator (1 if SUV being involved in the crash, 0 otherwise) [MI]	0.231	2.99	-0.0029	0.0031	-0.0001
Traffic characteristics					
Low traffic condition indicator (1 if AADT is below 4,000 veh/day, 0 otherwise) [SI]	1.429	13.43	-0.0080	-0.0017	0.0097
Temporal characteristics					
Weekday indicator (1 if crash occurred during the weekdays, 0 otherwise) [MI]	-0.169	-2.68	0.0094	-0.0099	0.0005
Number of observations			21,644		
Number of estimated parameters			21		
Log-likelihood at zero			-23778.36		
Log-likelihood at convergence			-10521.93		
$\rho^2 = 1 - LL(\boldsymbol{\beta})/LL(0)$			0.557		

*SI = Severe Injury; MI = Minor Injury; NI = No Injury

content, with zero blood alcohol content increasing the minor-injury constant mean and thus likelihood of minor injury (and subsequently decreasing the likelihood of no injury and severe injury). In contrast to the aggressive driving results, in all three years the minor-injury constant variance was influenced by the same variable, an early-morning indicator (crash occurring between midnight and 6AM), with crashes in this early morning period having a higher minor-injury constant variance for non-aggressive drivers.

All six models shown in Tables 5.3, 5.4, 5.5, 5.6, 5.7 and 5.8 show a wide variety of spatial, roadway, crash, driver, vehicle, traffic, and temporal characteristics influencing resulting driver injury severities. To compare the findings of all 6 models, Table 5.9 presents the marginal effects of all statistically significant variables for aggressive and non-aggressive driving crashes by year and injury levels. For the spatial variables, the Florida Department of Transportation District 2 indicator (Jacksonville across the state to the gulf coast of Florida) shows that District 2 had a lower probability of severe injuries for non-aggressive drivers in 2015 and 2016 but not in 2017 and not for aggressive drivers. The District 5 indicator (Orlando and surrounding area) shows that District 5 had a higher probability of severe injury for the non-aggressive driving model in 2017 but was insignificant in all other models. The District 6 indicator (Miami area) was found to be significant in all years for non-aggressive drivers but only in 2017 for aggressive drivers. In all cases where this indicator variable was significant, the probability of a minor injury was lower and the probability of no injury or severe injury was higher. Finally, District 7 (Tampa Bay area) was only statistically significant for aggressive drivers in 2015 and 2016 (and insignificant in all other models). In both 2015 and 2016 aggressive drivers were more likely to be severely injured in District 7 relative to other districts.

Table 5.9. Comparison of marginal effects between aggressive and non-aggressive driving in single-vehicle crashes over the years (marginal effects for non-aggressive drivers in parentheses)

Variables	No Injury			Minor Injury			Severe Injury		
	2015	2016	2017	2015	2016	2017	2015	2016	2017
Spatial characteristics									
District 2 indicator	–	–	–	–	–	–	–	–	–
(1 if crash occurred in District 2, 0 otherwise)	(0.0010)	(0.0010)	–	(0.0003)	(0.0003)	–	(-0.0012)	(-0.0013)	–
District 5 indicator	–	–	–	–	–	–	–	–	–
(1 if crash occurred in District 5, 0 otherwise)	–	–	(-0.0010)	–	–	(-0.0003)	–	–	(0.0013)
District 6 indicator	–	–	0.0092	–	–	-0.0107	–	–	0.0015
(1 if crash occurred in District 6, 0 otherwise)	(0.0055)	(0.0056)	(0.0064)	(-0.0058)	(-0.0059)	(-0.0067)	(0.0003)	(0.0003)	(0.0003)
District 7 indicator	-0.0117	0.0098	–	0.0046	-0.0124	–	0.0071	0.0025	–
(1 if crash occurred in District 7, 0 otherwise)	–	–	–	–	–	–	–	–	–
Roadway characteristics									
Urban principal arterial indicator (1 if crash occurred	-0.0081	–	–	-0.0010	–	–	0.0092	–	–
on urban principal arterials, 0 otherwise)	–	–	–	–	–	–	–	–	–
Wet surface indicator	-0.0086	-0.0144	–	0.0033	0.0172	–	0.0053	-0.0028	–
(1 if road surface condition was wet, 0 otherwise)	(-0.0062)	(-0.0060)	(-0.0052)	(0.0050)	(0.0047)	(0.0041)	(0.0013)	(0.0013)	(0.0011)
Straight roadway section indicator	0.0306	–	–	0.0043	–	–	-0.0349	–	–
(1 if straight section of the roadway, 0 otherwise)	(0.0590)	(0.0476)	(0.0451)	(-0.0624)	(-0.0502)	(-0.0475)	(0.0034)	(0.0026)	(0.0024)
Curved segment indicator (1 if roadway curves to the	–	-0.0157	-0.0244	–	0.0058	0.0094	–	0.0099	0.0151
right or left of travel direction, 0 otherwise)	–	–	–	–	–	–	–	–	–
Narrow shoulder width indicator	–	–	–	–	–	–	–	–	–
(1 if shoulder width is below 4 ft, 0 otherwise)	–	(0.0024)	(0.0023)	–	(-0.0019)	(-0.0015)	–	(-0.0005)	(-0.0007)
Wide shoulder width indicator (1 if shoulder width is	–	–	–	–	–	–	–	–	–
between 8 to 12 feet, 0 otherwise)	(-0.0163)	–	–	(0.0129)	–	–	(0.0034)	–	–
Crash characteristics									
Collision with roadside fixed object indicator	-0.0356	–	–	0.0426	–	–	-0.0070	–	–
(1 if collided with roadside fixed object as the first harmful event, 0 otherwise)	(-0.0041)	(-0.0060)	(-0.0060)	(-0.0010)	(-0.0014)	(-0.0014)	(0.0051)	(0.0074)	(0.0074)
Collision with non-fixed object indicator	–	0.0122	0.0128	–	0.0016	0.0020	–	-0.0139	-0.0148
(1 if collided with non-fixed object as the first harmful event, 0 otherwise)	–	–	–	–	–	–	–	–	–

On-road as harmful event location indicator (1 if harmful event occurred inside the roadway, 0 otherwise)	– (0.0407)	-0.0133 (0.0456)	-0.0202 (-0.0091)	– (-0.0334)	-0.0024 (-0.0374)	-0.0046 (0.0098)	– (-0.0073)	0.0158 (-0.0082)	0.0248 (-0.0007)
Off-road as harmful event location indicator (1 if harmful event occurred outside the roadway, 0 otherwise)	– (-0.0089)	– (-0.0097)	– (0.0446)	– (0.0095)	– (0.0105)	– (-0.0368)	– (-0.0006)	– (-0.0008)	– (-0.0078)
Driver characteristics									
Exceeding speed limit by more than 10 mi/h indicator (1 if travel exceeded the speed limit by more than 10 mi/h, 0 otherwise)	-0.0431	-0.0482	-0.0661	0.0521	0.0188	0.0272	-0.0090	0.0294	0.0389
Suspected Alcohol use indicator (1 if alcohol use is suspected in the crash involved driver, 0 otherwise)	-0.0289	-0.0115	–	-0.0045	-0.0020	–	0.0334	0.0135	–
Younger aged driver indicator (1 if driver age below 30, 0 otherwise)	– –	– (0.0017)	– (0.0022)	– –	– (0.0004)	– (0.0006)	– –	– (-0.0021)	– (-0.0027)
Middle aged driver indicator (1 if driver age between 30 to 49 years, 0 otherwise)	– (0.0069)	– (0.0072)	-0.0121 (0.0083)	– (-0.0056)	– (-0.0057)	-0.0026 (-0.0066)	– (-0.0013)	– (-0.0015)	0.0146 (-0.0017)
Restraint usage indicator (1 if shoulder and lap belt used, 0 otherwise)	0.0781 (0.2052)	0.0893 (0.2104)	0.0641 (0.2066)	-0.0411 (-0.1751)	-0.0518 (-0.1773)	-0.0354 (-0.1737)	-0.0370 (-0.0301)	-0.0376 (-0.0311)	-0.0287 (-0.0329)
Blood Alcohol Content indicator (1 if BAC is zero in the crash involved driver, 0 otherwise)	– –	– –	-0.1183 –	– –	– –	0.1433 –	– –	– –	-0.0250 –
Vehicle characteristics									
Passenger car involvement indicator (1 if passenger car being involved in the crash, 0 otherwise)	– (0.0053)	– (0.0040)	– (0.0044)	– (0.0011)	– (0.0007)	– (0.0008)	– (-0.0063)	– (-0.0047)	– (-0.0052)
Sport Utility Vehicle (SUV) involvement indicator (1 if SUV being involved in the crash, 0 otherwise)	– (-0.0034)	– (-0.0040)	– (-0.0029)	– (0.0036)	– (0.0042)	– (0.0031)	– (-0.0002)	– (-0.0002)	– (-0.0001)
Traffic characteristics									
Low traffic condition indicator (1 if AADT is below 4,000 veh/day, 0 otherwise)	– (-0.0011)	– (-0.0024)	– (-0.0080)	– (-0.0003)	– (-0.0005)	– (-0.0017)	– (0.0014)	– (0.0030)	– (0.0097)
Temporal characteristics									
Weekday indicator (1 if crash occurred during the weekdays, 0 otherwise)	– (0.0118)	– (0.0075)	– (0.0094)	– (-0.0124)	– (-0.0079)	– (-0.0099)	– (0.0007)	– (0.0004)	– (0.0005)

For all explanatory variables, the temporal stability of the marginal effects shown in Table 5.9 are of particular interest. While previous likelihood ratio tests indicate that both aggressive and non-aggressive driver crashes are temporally unstable, it is noteworthy that for non-aggressive driving crashes 12 variables were found to be statistically significant across all time periods. These include the District 6 indicator (1 if crash occurred in District 6, 0 otherwise), wet surface indicator (1 if road surface condition was wet, 0 otherwise), straight roadway section indicator (1 if straight section of the roadway, 0 otherwise), collision with roadside fixed object indicator (1 if collided with roadside fixed object as the first harmful event, 0 otherwise), on-road as harmful event location indicator (1 if harmful event occurred inside the roadway, 0 otherwise), off-road as harmful event location indicator (1 if harmful event occurred outside the roadway, 0 otherwise), middle aged driver indicator (1 if driver age between 30 to 49 years, 0 otherwise), restraint usage indicator (1 if shoulder and lap belt used, 0 otherwise), passenger car involvement indicator (1 if passenger car being involved in the crash, 0 otherwise), sport utility vehicle (SUV) involvement indicator (1 if SUV being involved in the crash, 0 otherwise), low traffic condition indicator (1 if AADT is below 4,000 veh/day, 0 otherwise), and the weekday indicator (1 if crash occurred during the weekdays, 0 otherwise). While the marginal effects of these variables tend to vary from year to year, some of the variations are quite modest, suggesting temporal stability of some effects. For example, the restrain usage indicator shows remarkably stable marginal effects over time, with the use of shoulder and lap restraints increasing the probability of no injury by around 0.21 relative to those not using shoulder and lap restraints (see Table 5.9).

In contrast, for crashes involving aggressive driving, just two variables were found to be statistically significant across all three time periods; exceeding speed limit by more than 10 mi/h indicator (1 if travel exceeded the speed limit by more than 10 mi/h, 0 otherwise), and again restraint usage. While exceeding the speed limit by more than 10 mi/h is statistically significant across all time periods, Table 5.9 shows considerable variation in marginal effects from on year to the next with, for example, the effect on severe-injury probabilities ranging from -0.0090 in 2015 to a positive 0.0389 in 2017. The marginal effect for restraint usage is more consistent over time, but much less consistent than it was for non-aggressive drivers (see Table 5.9). Interestingly, the effectiveness of safety belt usage in increasing the probability of no injury in aggressive driving crashes is less than half that in non-aggressive driving crashes. Some caution should be used in interpreting these results because safety belt use among aggressive drivers is much lower (mid-50% range) than for non-aggressive drivers (90% range), as shown in Table 5.2. For aggressive drivers, the parameter estimates may be capturing both the effectiveness of safety belts and the self-selectivity of risker drivers choosing to wear safety belts (see Eluru and Bhat (2007), Bhat et al. (2014) and Mannering et al. (2020) for a discussion of this point).

5.6 Discussion of Temporal Findings and Directions for Future Work

The findings of temporally consistent marginal effects for many explanatory variables in crashes involving non-aggressive drivers, and temporally inconsistent marginal effects in crashes involving aggressive drivers has important implications. It suggests that the temporal instability found in many studies may be the result of a subset of observations, and that temporal stability may exist in some if not most of the observations. Identification of temporally stable and temporally unstable observations could be made with a latent class structure where one class identifies crashes where parameter estimates vary over time and a second class identifies crashes where parameter estimates are fixed over time. Class-splitting functions could give critical insights into the factors that make some crashes temporally stable and others not (Fountas et al., 2018).

However, the structure of the model would not be trivial because allowing for temporally shifting parameters in one of the classes would be complicated, particularly if random parameters are considered (see Xiong and Mannering (2013) for a discussion of the estimation complexities in such a model). Still, the potential of such a model to unravel the complexities of temporal instability in safety data would be worth the effort if a computationally feasible estimation approach could be developed.

5.7 Summary and Conclusions

Using single-vehicle crash data related to aggressive and non-aggressive driving in Florida from 2015 to 2017, this study used a random parameters logit model (with heterogeneity in mean and variance) to explore the temporal stability of factors determining driver-injury severities over time. Three driver-injury levels were considered: no injury, minor injury (possible injury and non-incapacitating injury), and severe injury incapacitating injury and fatal injury). The estimated models showed a wide variety of factors significantly influencing driver-injury outcomes including vehicle type, crash type, roadway attributes, spatial and temporal characteristics, overall traffic volume, and driver factors.

Statistically significant differences were found between crashes involving aggressive and non-aggressive drivers, and both non-aggressive and aggressive crash injury-severity models exhibited statistically significant temporal instability over the three years considered. This is an important finding, and model estimation results that split aggressive and non-aggressive driving crashes could potentially be used to help guide injury-severity mitigation policies.

While both aggressive and non-aggressive driver models exhibit statistically significant temporal instability overall, the marginal effects of many of the explanatory variables in crashes involving non-aggressive drivers were relatively stable over time, whereas crashes involving non-aggressive drivers showed that only the restraint usage indicator had marginal effects that were relatively stable over time (see Table 5.9). Past research has tended to argue that this temporal instability is largely the result of global, fundamental changes in driving behavior and other associated factors (Mannering, 2018), but the findings in this chapter suggest that such changes may be largely driven not globally, but by a subset of the observation crash population. Future work that could identify temporally stable and unstable subsets of the crash population would provide a potentially valuable contribution in terms of guiding safety policies.

As with all studies, this study is not without its limitations. Although police officers are trained for consistency in interpretation, defining aggressive driving as having a vehicle operated in an erratic/reckless and aggressive manner is still open to the interpretation of the officer, and a potential source of error. There are other definitions of aggressive driving that may be worth exploring and these could potentially produce different results. These elements should also be given consideration in future work.

Chapter 6

Unobserved Heterogeneity and Temporal Instability in the Analysis of Work-Zone Crash-Injury Severities

Mouyid Islam, Nawaf Alnawmasi, Fred Mannering

6.1. Introduction

Work-zone safety has been increasingly recognized as a very serious problem. In the U.S., the number of fatal crashes in work-zones has increased substantially in recent years from 557 in 2012 to 710 in 2017 (a 27.5% increase). Single-vehicle crashes in work zones comprise about 20% of all crashes and are of particular interest because they reflect how the fundamentals of work zone design (transition areas, pavement markings, signage, etc.) impact potential driver errors that result in a crash. In Florida, the number of single-vehicle crashes occurring in work-zones (all injury severity levels) increased from 1,154 in 2012 to 1,500 in 2017 (a 23% increase). The numbers become more striking when considering the per-work-zone crash rate. As shown in Figure 6.1, although there are variations from year to year, there is a noticeable upward trend and, from 2012 to 2017, the number of single-vehicle crashes per work zone in Florida increased 50.7%. The distribution of driver-injury severities in these crashes has also varied from year to year, although there do not appear to be strong trends as illustrated in Figure 6.2. With the likely emphasis on renewing U.S. highway infrastructure in the coming years, understanding factors that influence the likelihood and severity of work-zone crashes is extremely important.

Work zone safety in general, and specifically injury severities in work zone crashes, has been the emphasis of a number of research studies over the years (Li and Bai, 2008, 2009; Harb et al., 2010; Meng et al., 2010; Tarko et al., 2011; Osman et al., 2018). While all of these studies

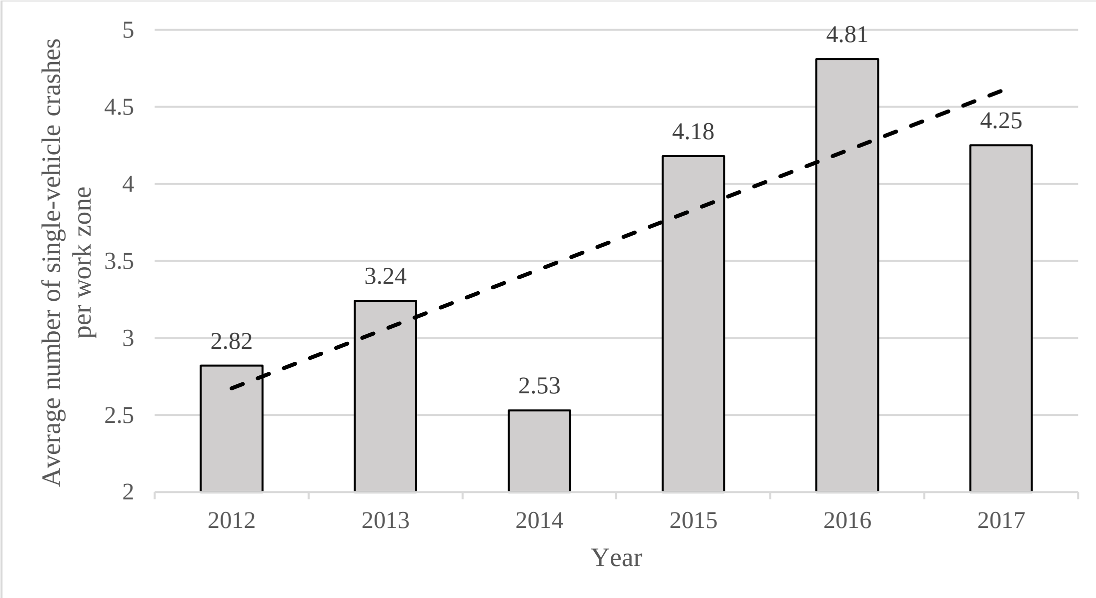


Figure 6.1. Average number of single-vehicle crashes per work-zone by year in Florida, 2012-17 (dashed line is the data trend line).

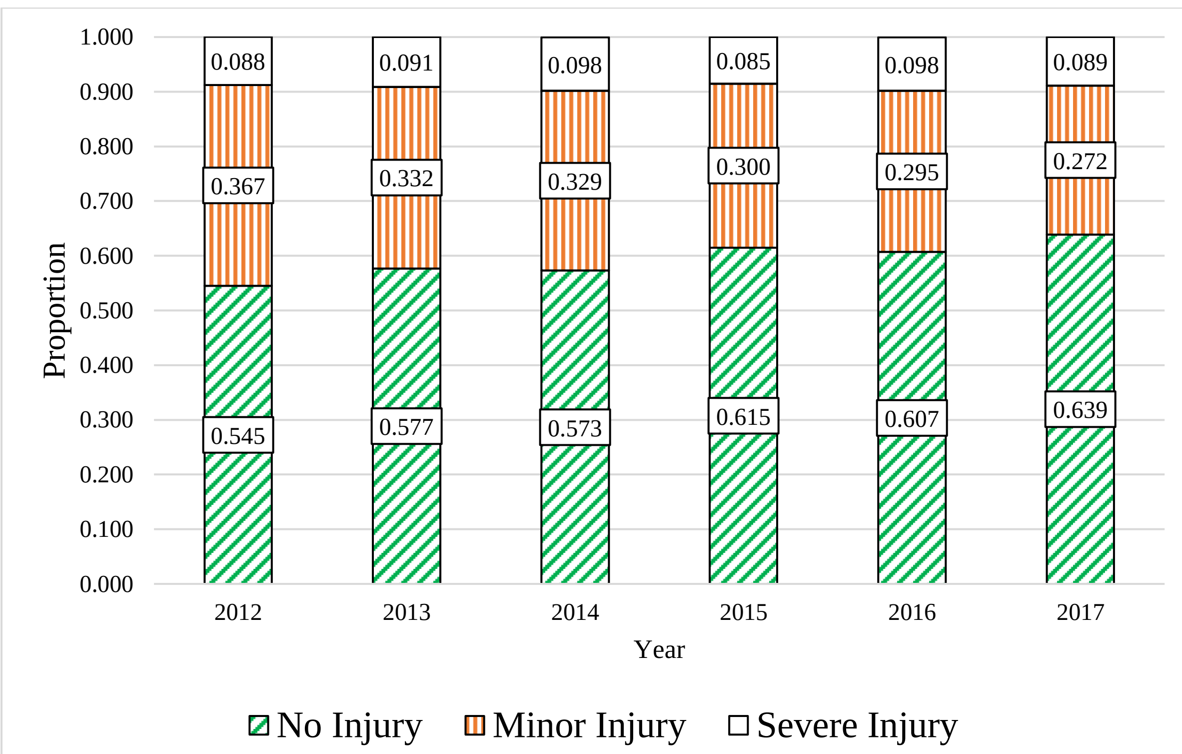


Figure 6.2. Proportion of single-vehicle work-zone crashes by severity by year in Florida, 2012-17.

have provided insights into the factors that determine work zone-injury severity, the issue of temporal stability in statistical models of work zone-injury severity (whether the influence of factors determining severity is stable over time) has not really been addressed to date. Mannering (2018) provides a detailed discussion of the potential causes of temporal instability in crash-data analyses and an abundance of recent work suggests that the influence of factors that determine injury severities, in general, may not be stable over time (Malyshkina and Mannering, 2009; Venkataraman et al, 2013; Behnood and Mannering, 2015; Behnood and Mannering, 2016; Alnawmasi and Mannering, 2019; Islam and Mannering, 2020). With such past empirical work finding evidence of temporal instability in the influence of factors affecting injury severity, temporal instability can also be expected to be a concern in work zone crashes. Identifying such instability and providing insight into how the influence of factors determining injury severities in work-zone crashes changes over time, could potentially be of long-term value in improving work zone safety. However, the assessment of temporal instability in work-zone crash severities is fundamentally different from the assessment of temporal instability in most past studies. This is because most past studies assess possible instability over a highway network (city or state) that is essentially unchanged from year to year. Thus, the unobserved highway-related factors that influence injury-severity outcomes are relatively stable over time. In the case of work zones, each work zone is unique with a unique set of unobservables and, from year to year, the mix of highway work zones changes as projects are started and completed. This means that any observed temporal instability in the factors influencing injury severities is likely to be a function of traditional sources as well as temporal variation in the mix of highways having active work zones. The combination of these two effects would make temporal instability much more likely than it would be on highway entities (highway segments) that have temporally constant attributes.

Using multi-year crash data, the intent of the current chapter is to estimate models of injury severity in work zone single-vehicle crashes and determine whether the influence of explanatory variables changes over time. The chapter begins with a description of the available work-zone data, followed by the presentation of the methodological approach. Statistical tests for temporal instability are then conducted and final model estimations are presented and discussed. Finally, the chapter concludes with a summary of findings and their implications.

6.2 Data Description

Data available for this study were the work-zone related crashes available in Florida's Crash Analysis Reporting data system (these are all police-reported crashes). For the purposes of this study, crash data were gathered over the six-year period from January 1, 2012, to December 31, 2017. Work-zone related crash data were filtered from the Florida's Crash Analysis Reporting data system and linked with a vehicle dataset based on the crash identification number. The resulting combined dataset provided detailed information about the crash, including roadway characteristics and conditions (including work zone characteristics), as well as vehicle and person characteristics. The linked dataset was filtered for single-vehicle work zone crashes, which resulted in 8,430 observations.

Information available in the data includes the resulting injury severity of the driver (no injury, possible injury, non-incapacitating injury, incapacitating injury and fatality), type of vehicle, driver actions (evasive maneuvers etc.), driver information (age, gender, usage of safety equipment, influence of alcohol, drug use), information relating to the time and location of the crash, roadway class, road surface condition, weather and light conditions, type of work zones, lane and shoulder widths, median width, location of harmful events, traffic volume, percent of trucks, presence of enforcement and workers in the work zones, and so on.

6.3 Methodology

Over the years, crash-related injury severities have been studied by a variety of ordered and unordered discrete outcome approaches including ordered logit/probit models, multinomial logit models, dual-state multinomial logit models, nested logit models, latent-class logit models, mixed (random parameters) logit models, Markov-switching models, and others (Savolainen et al., 2011; Islam and Hernandez, 2013; Mannering and Bhat, 2014; Mannering et al., 2016). To account for possible unobserved heterogeneity in the data, more recent research has focused on random parameter approaches (Milton et al., 2008; Eluru et al., 2008; Morgan and Mannering, 2011; Anastasopoulos and Mannering, 2011; Kim et al., 2013; Venkataraman et al., 2013; Behnood and Mannering, 2015), latent class models (Behnood et al., 2014; Cerwick et al., 2014; Shaheed and Gkritza, 2014; Yasmin et al., 2014) or combination of both (Xiong and Mannering, 2013) and heterogenity in means and variance (Venkataraman et. al, 2014; Behnood and Mannering 2017a, 2017b; Seraneeprakarn et al., 2017) to model the injury severities.

To study injury severity probabilities in the current chapter, a random parameters logit model that accounts for possible heterogeneity in the means and variances of the random parameters is used to address possible unobserved heterogeneity. The injury severity of drivers in single-vehicle work zone crashes are considered with possible injury outcomes of no injury, minor injury (possible injury and non-incapacitating injury) and severe injury (incapacitating injury and fatality). The modeling approach starts by defining a function that determines injury-severity (following the notation of Islam and Mannering, 2020),

$$S_{kn} = \boldsymbol{\beta}_k \mathbf{X}_{kn} + \varepsilon_{kn} \tag{6.1}$$

where S_{kn} is an injury-severity function determining the probability of driver-injury severity outcome k in work zone crash n, $\mathbf{X}_{kn}$ is a vector of explanatory variables that affect work zone driver-injury severity level k, $\boldsymbol{\beta}_k$ is a vector of estimable parameters, and ε_{kn} is an error term. Assuming the error term is assumed is generalized extreme value distributed, the standard multinomial logit model results as (McFadden, 1981),

$$P_n(k) = \frac{EXP\left[\boldsymbol{\beta}_k \mathbf{X}_{kn}\right]}{\sum_{\forall K} EXP\left(\boldsymbol{\beta}_k \mathbf{X}_{kn}\right)} \tag{6.2}$$

where $P_n(k)$ is the probability that work-zone crash n that will result in driver-injury severity outcome k with K being the set of the three possible injury-severity outcomes. If one or more parameter estimates in the vector $\boldsymbol{\beta}_k$ are allowed to vary across crash observations, Equation 6.2 can be rewritten as (Train, 2009; Washington et al., 2020),

$$P_n(k) = \int \frac{EXP\left(\boldsymbol{\beta}_k \mathbf{X}_{kn}\right)}{\sum_{\forall K} EXP\left(\boldsymbol{\beta}_k \mathbf{X}_{kn}\right)} f\left(\boldsymbol{\beta}_k \mid \boldsymbol{\varphi}_k\right) d\boldsymbol{\beta}_k \tag{6.3}$$

where $f(\boldsymbol{\beta}_k|\boldsymbol{\varphi}_k)$ is the density function of $\boldsymbol{\beta}_k$ and $\boldsymbol{\varphi}_k$ is a vector of parameters describing the density function (mean and variance), and all other terms are as previously defined.

The possibility of unobserved heterogeneity in the means and variances of parameters is also considered by letting $\boldsymbol{\beta}_{kn}$ be a vector of estimable parameters that varies across crashes defined as (Mannering et al., 2016; Seraneeprakarn et al., 2017; Behnood and Mannering, 2017; Waseem et al., 2019; Alnawmasi and Mannering, 2019; Behnood and Mannering, 2019; Islam and Mannering, 2020; Washington et al., 2020):

$$\boldsymbol{\beta}_{kn} = \beta_k + \Theta_{kn} \mathbf{Z}_{kn} + \sigma_{kn} EXP\left(\Psi_{kn} \mathbf{W}_{kn}\right) \nu_{kn} \tag{6.4}$$

where β_k is the mean parameter estimate across all crashes, $\mathbf{Z}_{kn}$ is a vector of crash-specific explanatory variables capturing heterogeneity in the mean that affects work zone injury-severity level k, $\mathbf{\Theta}_{kn}$ is the corresponding vector of estimable parameters, $\mathbf{W}_{kn}$ is the vector of crash-specific explanatory variables capturing heterogeneity in the standard deviation (σ_{kn}) with corresponding parameter vector Ψ_{kn}, and v_{kn} is a disturbance term.

The choice of density functions for the term $f(\boldsymbol{\beta}_k|\boldsymbol{\varphi}_k)$ in Equation 6.3 can be determined by numerical evaluation (Mannering et al., 2016) and in this chapter none of the various density function considered were found to be statistically superior to the normal distribution. This finding is consistent with other empirical studies such as Milton et al., 2008; Alnawmasi and Mannering, 2019; Islam and Mannering, 2020, among many others. Model estimations were conducted using a simulated maximum likelihood approach with 1,000 Halton draws (McFadden and Train, 2000; Bhat, 2001; Train, 2009). Marginal effects were computed to determine the effect of explanatory variables on injury-severity probabilities, with the marginal effect providing the effect that a one-unit increase in an explanatory variable has on the injury-outcome probabilities (for indicator variables this is the change in probability resulting from the indicator going from zero to one). The forthcoming marginal effect tables report the marginal effect averaged over all crash observations.

6.4 Tests for Temporal Stability

After extensively testing for temporal instability across years, it was determined that statistically significant differences existed in each year of the injury-severity data from 2012 to 2017. This was confirmed by a series of likelihood-ratio tests. The first test was based on a model estimated overall years of available data (2012-17) and each converged model representing each year. Using the converged models, the likelihood ratio test for this case is

$$X^2 = -2\begin{bmatrix} LL(\boldsymbol{\beta}_{2012-17}) - LL(\boldsymbol{\beta}_{2012}) - LL(\boldsymbol{\beta}_{2013}) - LL(\boldsymbol{\beta}_{2014}) \\ -LL(\boldsymbol{\beta}_{2015}) - LL(\boldsymbol{\beta}_{2016}) - LL(\boldsymbol{\beta}_{2017}) \end{bmatrix} \tag{5}$$

where, $LL(\boldsymbol{\beta}_{2012\text{-}17})$ is the log-likelihood at the convergence of the model that used all of the available data (2012-17), $LL(\boldsymbol{\beta}_{2012})$, $LL(\boldsymbol{\beta}_{2013})$, $LL(\boldsymbol{\beta}_{2014})$, $LL(\boldsymbol{\beta}_{2015})$, $LL(\boldsymbol{\beta}_{2016})$, and $LL(\boldsymbol{\beta}_{2017})$ is the log-likelihood at convergence of a model based on 2012, 2013, 2014, 2015, 2016, and 2017-data, respectively. Model estimates gave an X^2 of 848.31 which is χ^2 distributed with 103 degrees of freedom (the number of parameters found to be statistically significant in the model using all of the data year, 2012-17 excluding the numbers of parameters found statistically significant in each year). This χ^2 value gives 99.99% confidence that the null hypothesis that the parameters are equal parameters in all the years can be rejected.

To test for temporal instability further, additional likelihood ratio tests were run as (Washington et al., 2020),

$$X^2 = -2\left[LL(\boldsymbol{\beta}_{t_2 t_1}) - LL(\boldsymbol{\beta}_{t_1})\right] \tag{6}$$

where $LL(\boldsymbol{\beta}_{t_2 t_1})$ is the log-likelihood at convergence of a model containing converged parameters based on using time-period t_2's data, while using data from time-period t_1, and $LL(\boldsymbol{\beta}_{t_1})$ is the log-likelihood at convergence of the model using time-period t_1's data, with parameters no longer restricted to using time-period t_2's converged parameters as is the case for $LL(\boldsymbol{\beta}_{t_2 t_1})$. This test was also reversed such that time-period t_1 above becomes time period t_2 and time period t_2 above becomes subset t_1 (thus giving two test results for each model comparison). The resulting value X^2 is χ^2 distributed and can be used to determine if the null hypothesis that the parameters are equal in the two periods can be rejected. Using the converged parameters of the 2013 model (Table 6.1) as starting values and applying them to the 2012 data gave $X^2 = 76.61$ (from Equation 6). With 20 degrees of freedom, this gave a χ^2 confidence level of more than 99.9% that the null hypothesis that the two time periods are the same can be rejected. Using the converged parameters of the 2012 model (Table 6.1) as starting values and applying them to the 2013 data gave $X^2 = 68.45$ and, with 19 degrees of freedom, this also gave a χ^2 confidence level of more than 99.9% that the null hypothesis that the two time periods are the same can be rejected. Similarly, other two-year periods of interest were tested for temporal stability (see Table 6.1) and all tests indicated that the null hypothesis that the parameters are equal between years could be rejected.

Table 6.1. Likelihood ratio test results between different time periods based on random parameters approaches with heterogeneity in means and variances in Florida work zone crashes involving single vehicles (χ^2 values with degrees of freedom in parenthesis and confidence level in brackets).

t_1 (Eq. 6.6)	t_2 (Eq. 6.6)					
	2012	**2013**	**2014**	**2015**	**2016**	**2017**
2012	–	76.61 (20) [>99.99%]	79.04 (18) [>99.99%]	60.46 (21) [> 99.99%]	120.08 (20) [> 99.99%]	112.83 (27) [> 99.99%]
2013	68.45 (19) [> 99.99%]	–	80.26 (18) [>99.99%]	52.76 (21) [> 99.99%]	82.14 (20) [> 99.99%]	145.51 (27) [> 99.99%]
2014	110.81 (19) [> 99.99%]	52.41 (20) [>99.99%]	–	36.43 (21) [> 98.00%]	62.44 (20) [> 99.99%]	106.34 (27) [> 99.99%]
2015	84.58 (19) [> 99.99%]	63.23 (20) [>99.99%]	69.24 (18) [>99.99%]	–	97.00 (20) [> 99.99%]	95.01 (27) [> 99.99%]
2016	138.63 (19) [> 99.99%]	167.56 (20) [>99.99%]	252.64 (18) [>99.99%]	122.25 (21) [> 99.99%]	–	189.01 (27) [> 99.99%]
2017	100.82 (19) [> 99.99%]	121.39 (20) [>99.99%]	134.30 (18) [>99.99%]	66.77 (21) [> 99.99%]	52.36 (20) [> 99.99%]	–

6.5 Model Estimation Results

Table 6.2 presents the summary statistics for variables found to be statistically significant in the model estimations (this table provides values for each of year of the 2012-17 analysis period). Tables 6.3 to 6.8 present the model estimation results for work-zone crashes from 2012 to 2017 and Tables 6.9, 6.10 and 6.11 present the temporal comparison of marginal effects for no injury, minor, and severe injury over the analysis period, respectively. The models have reasonably good overall statistical fit with ρ^2 values of 0.329, 0.350, 0.355, 0.343, 0.319, and 0.403 for 2012, 2013, 2014, 2015, 2016, and 2017, respectively.

As shown in Tables 6.3 through 6.8, different parameters were found to be random (normally distributed) with statistically significant heterogeneity in means and variance in each of the six estimated models. For the 2012, 2014, and 2015 models, the constant term specific to minor injury was found to be a random parameter. For the 2013 model, the shoulder-median-work indicator was found to have a statistically significant random parameter in the severe injury function. For the 2016 model, the young driver indicator was found to have a statistically significant random parameter in the minor injury function, and for 2017 model, the young driver indicator (in minor injury) and shoulder-median-work indicator (severe injury) were found to have statistically significant random parameters. For the 2012 model (Table 6.3), the mean of the constant-term parameter increased if the harmful event location (the harmful event is the object struck that caused injury or property damage) occurred on the right shoulder and the variance of the parameter increased if the crash occurred in clear weather. For the 2013 (Table 6.4), the mean of the shoulder-median-work indicator specific to severe injury increased if an older driver was involved and the variance of the parameter increased if driver negligence was involved. For 2014

Table 6.2. Summary statistics for variables found to be statistically significant in model estimations.

Variables	2012		2013		2014		2015		2016		2017	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Environmental characteristics												
Rain indicator (1 if crash occurred at the time of raining, 0 otherwise)	0.184	0.388	0.229	0.420	0.234	0.424	0.199	0.399	0.184	0.388	0.176	0.381
Clear weather indicator (1 if crash occurred at the time of clear weather, 0 otherwise)	0.631	0.482	0.577	0.493	0.558	0.496	0.606	0.488	0.606	0.488	0.654	0.475
Dark indicator (1 if crash occurred at the time of darkness, 0 otherwise)	0.136	0.343	0.160	0.367	0.165	0.371	0.166	0.372	0.164	0.370	0.154	0.361
Traffic characteristics												
Low traffic volume indicator (1 if average annual daily traffic is below 40,000 vehicles/day, 0 otherwise)	0.294	0.455	0.362	0.480	0.373	0.483	0.338	0.473	0.069	0.254	0.126	0.332
Average percent of large trucks (1 if large truck volume between 7.5% and 12.5% of all traffic, 0 otherwise)	0.545	0.497	0.553	0.497	0.471	0.499	0.456	0.498	0.188	0.391	0.252	0.434
Temporal characteristics												
Late night indicator (1 if time of day is between 8 pm to 11:59 pm, 0 otherwise)	0.136	0.342	0.171	0.377	0.140	0.347	0.148	0.355	0.160	0.367	0.153	0.360
Afternoon indicator (1 if time of day is between 12 to 2:59 PM, 0 otherwise)	0.143	0.350	0.161	0.368	0.156	0.363	0.156	0.363	0.164	0.370	0.152	0.359
Earlier months indicator (1 if crash occurred January to April, 0 otherwise)	0.274	0.446	0.292	0.454	0.376	0.484	0.279	0.448	0.342	0.474	0.349	0.476

Variables	2012		2013		2014		2015		2016		2017	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Later months indicator (1 if crash occurred October to December, 0 otherwise)	0.249	0.432	0.221	0.415	0.194	0.395	0.265	0.441	0.205	0.403	0.22	0.414
Spatial characteristics												
District 1 indicator (1 if crash occurred in District 1, 0 otherwise)	0.123	0.329	0.082	0. 275	0.094	0.292	0.124	0.329	0.188	0.391	0.107	0.309
District 2 indicator (1 if crash occurred in District 2, 0 otherwise)	0.117	0.322	0.097	0.297	0.098	0.297	0.107	0.309	0.142	0.349	0.174	0.379
District 3 indicator (1 if crash occurred in District 3, 0 otherwise)	0.045	0.207	0.040	0.196	0.047	0.213	0.055	0.228	0.069	0.254	0.066	0.248
District 4 indicator (1 if crash occurred in District 4, 0 otherwise)	0.272	0.445	0.299	0.458	0.204	0.403	0.158	0.365	0.162	0.368	0.196	0.397
District 5 indicator (1 if crash occurred in District 5, 0 otherwise)	0.121	0.326	0.159	0.365	0.222	0.415	0.207	0.405	0.221	0.415	0.236	0.424
District 6 indicator (1 if crash occurred in District 6, 0 otherwise)	0.149	0.357	0.143	0.351	0.153	0.360	0.158	0.365	0.113	0.317	0.120	0.324
District 7 indicator (1 if crash occurred in District 7, 0 otherwise)	0.168	0.374	0.177	0.381	0.180	0.384	0.188	0.391	0.102	0.303	0.100	0.300
Vehicle characteristics												
Motorcycle indicator (1 if motorcycle, 0 otherwise)	0.052	0.222	0.063	0.244	0.042	0.201	0.043	0.203	0.044	0.205	0.034	0.182
Pickup truck indicator (1 if pickup, 0 otherwise)	0.110	0.314	0.099	0.299	0.107	0.309	0.103	0.304	0.105	0.306	0.123	0.328

Variables	2012		2013		2014		2015		2016		2017	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Passenger car indicator (1 if passenger car, 0 otherwise)	0.556	0.496	0.55	0.497	0.553	0.497	0.567	0.495	0.534	0.498	0.541	0.498
Work zone characteristics												
Lane-closure work zone indicator (1 if lane closure, 0 otherwise)	0.155	0.362	0.168	0.374	0.153	0.360	0.154	0.361	0.154	0.361	0.184	0.388
Lane shift indicator (1 if lane shift work, 0 otherwise)	0.097	0.297	0.106	0.309	0.092	0.290	0.093	0.290	0.087	0.281	0.090	0.287
Shoulder-median work indicator (1 if work on shoulder and median, 0 otherwise)	0.551	0.497	0.515	0.499	0.577	0.493	0.591	0.491	0.603	0.489	0.564	0.495
Absence of enforcement indicator (1 if no enforcement was present, 0 otherwise)	0.899	0.300	0.902	0.296	0.912	0.282	0.914	0.280	0.917	0.274	0.924	0.263
Presence of workers indicator (1 if workers were present, 0 otherwise)	0.319	0.466	0.326	0.468	0.616	0.486	0.382	0.486	0.368	0.482	0.352	0.477
Non-work zone indicator (1 if crashes were identified not related to work zone geometry, 0 otherwise)	0.560	0.496	0.597	0.490	0.648	0.477	0.637	0.480	0.622	0.484	0.639	0.480
Work zone geometry indicator (1 if crashes were identified as work zone geometry related, 0 otherwise)	0.321	0.467	0.272	0.445	0.226	0.418	0.237	0.425	0.233	0.423	0.238	0.426
Transition area indicator (1 if the crash occurred in the work zone's transition area, 0 otherwise)	0.140	0.347	0.151	0.358	0.142	0.349	0.155	0.362	0.146	0.354	0.151	0.358

Variables	2012		2013		2014		2015		2016		2017	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Harmful event characteristics												
Most harmful shoulder indicator (1 if most harmful event occurs in the shoulder, 0 otherwise)	0.336	0.472	0.331	0.470	0.361	0.480	0.354	0.478	0.132	0.339	0.214	0.410
Most harmful median indicator (1 if most harmful event occurs in the median, 0 otherwise)	0.207	0.405	0.246	0.431	0.256	0.436	0.25	0.433	0.092	0.290	0.102	0.302
Harmful event on-road indicator (1 if harmful event occurred on road, 0 otherwise)	0.506	0.500	0.518	0.499	0.460	0.498	0.491	0.499	0.492	0.499	0.515	0.499
Harmful event off-road indicator (1 if the harmful event was off road, 0 otherwise)	0.129	0.336	0.152	0.359	0.174	0.379	0.183	0.387	0.183	0.387	0.202	0.402
Harmful fixed object indicator (1 if harmful event occurred with roadside fixed object, 0 otherwise)	0.551	0.497	0.560	0.496	0.541	0.498	0.509	0.499	0.511	0.499	0.526	0.499
Harmful right shoulder indicator (1 if harmful event occurs on the right shoulder, 0 otherwise)	0.216	0.412	0.224	0.417	0.236	0.425	0.206	0.404	0.220	0.414	0.184	0.388
Harmful median indicator (1 if harmful event occurs in the median, 0 otherwise)	0.095	0.293	0.068	0.253	0.095	0.293	0.091	0.287	0.075	0.264	0.066	0.249
Harmful overturn indicator (1 if harmful event occurs with overturning, 0 otherwise)	0.069	0.254	0.074	0.262	0.090	0.287	0.081	0.273	0.086	0.281	0.059	0.236

Road geometric characteristics												
Large shoulder width indicator (1 if right shoulder width between 6 to 10 ft., 0 otherwise)	0.441	0.496	0.402	0.490	0.404	0.490	0.372	0.483	0.144	0.351	0.236	0.425
Urban interstate indicator (1 if crash occurred on urban interstate, 0 otherwise)	0.299	0.458	0.274	0.446	0.268	0.443	0.269	0.443	0.100	0.300	0.166	0.372
Rural interstate indicator (1 if crash occurred on rural interstate, 0 otherwise)	0.069	0.254	0.094	0.292	0.138	0.345	0.153	0.360	0.019	0.139	0.045	0.208
Urban toll way indicator (1 if crash occurred on an urban toll way, 0 otherwise)	0.069	0.254	0.094	0.292	0.138	0.345	0.153	0.360	0.019	0.139	0.048	0.215
Driver characteristics												
Young driver indicator (1 if driver's age below 30 years, 0 otherwise)	0.383	0.486	0.389	0.487	0.407	0.491	0.395	0.489	0.380	0.485	0.404	0.490
Old driver indicator (1 if driver's age between 50 to 65 years, 0 otherwise)	0.189	0.392	0.184	0.388	0.167	0.373	0.187	0.390	0.181	0.385	0.194	0.395
Older driver indicator (1 if driver's age is 65 years and above, 0 otherwise)	0.076	0.265	0.072	0.260	0.078	0.268	0.077	0.267	0.084	0.278	0.082	0.275
Negligent driver indicator (1 if negligent driving, 0 otherwise)	0.333	0.471	0.317	0.465	0.297	0.457	0.333	0.471	0.337	0.472	0.366	0.481
Over steering indicator (1 if driver actions involved over steering, 0 otherwise)	0.018	0.133	0.023	0.151	0.019	0.136	0.019	0.136	0.022	0.149	0.021	0.144

Table 6.3. Mixed logit model with heterogeneity in mean and variance for single-vehicle work-zone crashes in Florida, 2012.

Variables	**Parameter Estimates**	**t-stat**	**Marginal Effects**		
			No Injury	**Minor Injury**	**Severe Injury**
Constant [SI]	-3.817	-11.42			
Random parameter (normally distributed)					
Constant [MI]	-6.065	-2.59			
(*Standard deviation of parameter distribution*)	(*4.561*)	(*2.04*)			
Heterogeneity in the mean of random parameter					
Constant [MI]: Harmful right shoulder indicator (1 if harmful event occurs on the right shoulder, 0 otherwise)	1.663	1.76			
Heterogeneity in the variance of random parameter					
Constant [MI]: Clear weather indicator (1 if clear weather, 0 otherwise)	0.260	1.76			
Environmental characteristics					
Rain indicator (1 if it is rainy, 0 otherwise) [SI]	-0.667	-1.61	0.0040	0.0004	-0.0044
Traffic characteristics					
Low traffic volume indicator (1 if average annual daily traffic is below 40,000 vehicles/day, 0 otherwise) [SI]	0.765	2.68	-0.0150	-0.0019	0.0169
Average percent of large trucks (1 if large truck volume between 7.5% and 12.5% of all traffic, 0 otherwise) [MI]	1.325	1.97	-0.0384	0.0429	-0.0046
Temporal characteristics					
Late night indicator (1 if time of day is between 8 pm to 11:59 pm, 0 otherwise) [SI]	0.841	2.43	-0.0073	-0.0009	0.0083
Earlier months indicator (1 if crash occurred January to April, 0 otherwise) [MI]	1.101	1.72	-0.0161	0.0181	-0.0020
Spatial characteristics					
District 5 indicator (1 if crash occurred in District 5, 0 otherwise) [NI]	-0.926	-2.81	-0.0140	0.0053	0.0087
Vehicle characteristics					
Motorcycle indicator (1 if motorcycle, 0 otherwise) [SI]	2.762	5.04	-0.0204	-0.0040	0.0244
Pickup truck indicator (1 if motorcycle, 0 otherwise) [MI]	-1.712	-1.76	0.0087	-0.0097	0.0010
Passenger car indicator (1 if passenger car, 0 otherwise) [NI]	0.634	2.42	0.0311	-0.0185	-0.0126
Harmful event characteristics					
Harmful event off-road indicator (1 if the harmful event was off road, 0 otherwise) [SI]	0.878	2.60	-0.0089	-0.0010	0.0099

Harmful overturn indicator (1 if harmful event occurs with overturning, 0 otherwise) [MI]	5.378	2.47	-0.0232	0.0260	-0.0028
Harmful fixed object indicator (1 if harmful event occurred with roadside fixed object, 0 otherwise) [NI]	-1.124	-3.88	-0.0685	0.0324	0.0361
Roadway characteristics					
Large shoulder width indicator (1 if right shoulder width between 6 to 10 ft., 0 otherwise) [NI]	-0.605	-2.37	-0.0276	0.0141	0.0135
Driver characteristics					
Negligent driver indicator (1 if negligent driving, 0 otherwise) [MI]	1.561	1.91	-0.0291	0.0327	-0.0036
Number of observations			1,154		
Log-likelihood at zero			-1267.798		
Log-likelihood at convergence			-850.129		
$\rho^2 = 1 - LL(\boldsymbol{\beta})/LL(0)$			0.329		

SI = Severe Injury; MI = Minor Injury; NI = No Injury

Table 6.4. Mixed logit model with heterogeneity in mean and variance for single-vehicle work-zone crashes in Florida, 2013.

Variables	**Parameter Estimates**	**t-stat**	**Marginal Effects**		
			No Injury	**Minor Injury**	**Severe Injury**
Constant [MI]	-1.935	-11.47			
Constant [SI]	-2.149	-8.60			
Random parameter (normally distributed)					
Shoulder-median work indicator (1 if work on shoulder and median, 0 otherwise) [SI]	-4.616	-1.68	-0.0150	-0.0065	0.0215
(*Standard deviation of should-median work*)	(*4.091*)	(*2.19*)			
Heterogeneity in the mean of random parameter					
Shoulder-median work [SI]: Older driver indicator (1 if driver's age is 65 years and above, 0 otherwise)	2.186	1.67			
Heterogeneity in the variance of random parameter					
Shoulder-median work [SI]: Negligent driver indicator (1 if negligent driving, 0 otherwise)	0.226	1.71			
Environmental characteristics					
Rain indicator (1 if crash occurred at the time of raining, 0 otherwise) [NI]	0.285	1.82	0.0123	-0.0106	-0.0017
Dark indicator (1 if crash occurred at the time of darkness, 0 otherwise) [MI]	-0.519	-2.63	0.0116	-0.0124	0.0008
Traffic characteristics					
Low traffic volume indicator (1 if average annual daily traffic is below 40,000 vehicles/day, 0 otherwise) [NI]	-0.250	-1.87	-0.0176	0.0150	0.0026
Average percent of large trucks (1 if large truck volume between 7.5% and 12.5% of all traffic, 0 otherwise) [MI]	0.252	1.87	-0.0241	0.0258	-0.0017
Temporal characteristics					
Afternoon indicator (1 if time of day is between 12 pm to 2:59 pm, 0 otherwise) [MI]	-0.494	-2.47	0.0104	-0.0113	0.0008
Vehicle characteristics					
Motorcycle indicator (1 if motorcycle, 0 otherwise) [MI]	1.444	5.05	-0.0163	0.0188	-0.0025
Pickup truck indicator (1 if motorcycle, 0 otherwise) [SI]	-2.463	-2.86	0.0030	0.0012	-0.0041
Passenger car indicator (1 if passenger car, 0 otherwise) [SI]	-1.686	-4.44	0.0166	0.0063	-0.0229
Work zone characteristics					
Work zone geometry indicator (1 if crashes were identified as work zone geometry related, 0 otherwise) [SI]	-0.902	-2.40	0.0046	0.0018	-0.0064

Harmful event characteristics					
Most harmful median indicator (1 if most harmful event occurs in the median, 0 otherwise) [NI]	-0.851	-4.82	-0.0444	0.0384	0.0060
Harmful median indicator (1 if harmful event occurred in the median, 0 otherwise) [SI]	1.461	2.77	-0.0039	-0.0020	0.0059
Most harmful shoulder indicator (1 if harmful event occurs in the shoulder, 0 otherwise) [NI]	-0.711	-4.51	-0.0491	0.0425	0.0066
Harmful overturn indicator (1 if harmful event occurred with overturning, 0 otherwise) [MI]	1.775	7.02	-0.0250	0.0271	-0.0022
Harmful fixed object indicator (1 if harmful event occurred with roadside fixed object, 0 otherwise) [NI]	-0.399	-2.66	-0.0454	0.0387	0.0067
Number of observations			1,440		
Log-likelihood at zero			-1582.002		
Log-likelihood at convergence			-1027.924		
$\rho^2 = 1 - LL(\boldsymbol{\beta})/LL(0)$			0.350		

SI = Severe Injury; MI = Minor Injury; NI = No Injury

Table 6.5. Mixed logit model with heterogeneity in mean and variance for single-vehicle work-zone crashes in Florida, 2014.

Variables	Parameter Estimates	t-stat	Marginal Effects		
			No Injury	**Minor Injury**	**Severe Injury**
Constant [SI]	-3.041	-8.83			
Random parameter (normally distributed)					
Constant [MI]	-5.921	-2.97			
(*Standard deviation of parameter distribution*)	(*4.250*)	(*2.75*)			
Heterogeneity in the mean of random parameter					
Constant [MI]: Clear weather indicator (1 if clear weather, 0 otherwise)	-1.221	-2.05			
Heterogeneity in the variance of random parameter					
Constant [MI]: Negligent driver indicator (1 if negligent driving, 0 otherwise)	0.337	2.15			
Environmental characteristics					
Rain indicator (1 if it is rainy, 0 otherwise) [SI]	-1.263	-2.88	0.0065	0.0008	-0.0074
Traffic characteristics					
Average percent of large trucks (1 if large truck volume between 7.5% and 12.5% of all traffic, 0 otherwise) [MI]	1.565	2.40	-0.0407	0.0458	-0.0051
Temporal characteristics					
Earlier months indicator (1 if crash occurred January to April, 0 otherwise) [SI]	-0.867	-2.70	0.0110	0.0013	-0.0123
Spatial characteristics					
District 4 indicator (1 if crash occurred in District 4, 0 otherwise) [NI]	0.695	1.87	0.0117	-0.0081	-0.0036
District 6 indicator (1 if crash occurred in District 6, 0 otherwise) [NI]	1.158	2.81	0.0117	-0.0083	-0.0034
District 7 indicator (1 if crash occurred in District 7, 0 otherwise) [SI]	0.942	3.14	-0.0154	-0.0025	0.0180
Vehicle characteristics					
Motorcycle indicator (1 if motorcycle, 0 otherwise) [SI]	2.592	5.57	-0.0137	-0.0033	0.0170
Work Zone characteristics					
Absence of enforcement indicator (1 if no enforcement was present, 0 otherwise) [MI]	1.655	1.65	-0.0815	0.0912	-0.0097
Harmful event characteristics					
Harmful overturn indicator (1 if harmful event occurs with overturning, 0 otherwise) [MI]	4.141	2.93	-0.0221	0.0282	-0.0062
Most harmful shoulder indicator (1 if most harmful event occurs in the shoulder, 0	-1.680	-4.80	-0.0747	0.0342	0.0405

otherwise) [NI]					
Most harmful median indicator (1 if most harmful event occurs in the median, 0 otherwise) [NI]	-1.147	-3.18	-0.0316	0.0167	0.0149
Roadway characteristics					
Rural interstate indicator (1 if crash occurred on rural interstate, 0 otherwise) [MI]	1.657	1.94	-0.0143	0.0161	-0.0017
Driver characteristics					
Old driver indicator (1 if driver's age between 50 to 65 years, 0 otherwise) [MI]	-1.153	-1.66	0.0086	-0.0098	0.0011
Number of observations			1,111		
Log-likelihood at zero			-1220.558		
Log-likelihood at convergence			-787.273		
$\rho^2 = 1 - LL(\boldsymbol{\beta})/LL(0)$			0.355		

SI = Severe Injury; MI = Minor Injury; NI = No Injury

Table 6.6. Mixed logit model with heterogeneity in mean and variance for single-vehicle work-zone crashes in Florida, 2015.

Variables	**Parameter Estimates**	**t-stat**	**Marginal Effects**		
			No Injury	**Minor Injury**	**Severe Injury**
Constant [SI]	-2.573	-9.49			
Random parameter (normally distributed)					
Constant [MI]	-3.664	-3.24			
(*Standard deviation of parameter distribution*)	(*3.391*)	(*2.64*)			
Heterogeneity in the mean of random parameter					
Constant [MI]: Young driver indicator (1 if driver's age below 30 years, 0 otherwise, 0 otherwise)	0.689	1.90			
Heterogeneity in the variance of random parameter					
Constant [MI]: Rain indicator (1 if it is rainy, 0 otherwise)	0.479	2.03			
Environmental characteristics					
Rain indicator (1 if it is rainy, 0 otherwise) [NI]	1.531	3.99	0.0234	-0.0164	-0.0070
Spatial characteristics					
District 2 indicator (1 if crash occurred in District 2, 0 otherwise) [SI]	-1.345	-2.55	0.0032	0.0004	-0.0036
District 3 indicator (1 if crash occurred in District 3, 0 otherwise) [MI]	0.679	2.78	-0.0065	0.0072	-0.0007
District 6 indicator (1 if crash occurred in District 6, 0 otherwise) [NI]	0.611	1.80	0.0089	-0.0061	-0.0028
Vehicle characteristics					
Motorcycle indicator (1 if motorcycle, 0 otherwise) [MI]	1.216	4.50	-0.0107	0.0122	-0.0016
Passenger car indicator (1 if passenger car, 0 otherwise) [SI]	-1.036	-4.63	0.0190	0.0027	-0.0217
Work zone characteristics					
Presence of workers indicator (1 if workers were present, 0 otherwise) [NI]	0.575	2.65	0.0228	-0.0141	-0.0088
Transition area indicator (1 if the crash occurred in the work zone's transition area, 0 otherwise) [MI]	1.012	2.08	-0.0119	0.0131	-0.0012
Harmful event characteristics					
Most harmful shoulder indicator (1 if most harmful event occurred in the shoulder, 0 otherwise) [SI]	0.963	3.53	-0.0211	-0.0030	0.0241
Harmful overturn indicator (1 if harmful event occurs with overturning, 0 otherwise) [MI]	1.887	2.52	-0.0121	0.0140	-0.0020

Harmful right shoulder indicator (1 if harmful event occurs on the right shoulder, 0 otherwise) [MI]	1.471	2.79	-0.0226	0.0253	-0.0028
Most harmful median indicator (1 if most harmful event occurs in the median, 0 otherwise) [NI]	-1.162	-4.08	-0.0368	0.0198	0.0170
Roadway characteristics					
Urban interstate indicator (1 if crash occurred on urban interstate, 0 otherwise) [SI]	0.567	2.10	-0.0084	-0.0014	0.0098
Rural interstate indicator (1 if crash occurred on rural interstate, 0 otherwise) [NI]	-0.615	-2.26`	-0.0121	0.0060	0.0061
Driver characteristics					
Negligent driver indicator (1 if negligent driving, 0 otherwise) [SI]	0.665	2.77	-0.0132	-0.0021	0.0153
Over steering indicator (1 if driver actions involved over steering, 0 otherwise) [SI]	1.266	2.04	-0.0025	-0.0005	0.0030
Number of observations			1,524		
Log-likelihood at zero			-1674.285		
Log-likelihood at convergence			-1100.637		
$\rho^2 = 1 - LL(\boldsymbol{\beta})/LL(0)$			0.343		

SI = Severe Injury; MI = Minor Injury; NI = No Injury

Table 6.7. Mixed logit model with heterogeneity in mean and variance for single-vehicle work-zone crashes in Florida, 2016.

Variables	Parameter Estimates	t-stat	Marginal Effects		
			No Injury	Minor Injury	Severe Injury
Constant [MI]	-1.054	-7.83			
Constant [SI]	-2.203	-9.68			
Random parameter (normally distributed)					
Young driver indicator (1 if driver's age below 30 years, 0 otherwise) [MI]	-1.124	-1.24	-0.0182	0.0202	-0.0020
(*Standard deviation of parameter distribution*)	(*2.573*)	(*1.79*)			
Heterogeneity in the mean of random parameter					
Young driver (below 30 years) [MI]: Motorcycle indicator (1 if motorcycle, 0 otherwise)	4.458	2.34			
Heterogeneity in the variance of random parameter					
Young driver (below 30 years) [MI]: Earlier months indicator (1 if crash occurred January to April, 0 otherwise)	0.694	2.37			
Traffic characteristics					
Average percent of large trucks (1 if large truck volume between 7.5% and 12.5% of all traffic, 0 otherwise) [NI]	-0.721	-3.60	-0.0244	0.0172	0.0072
Spatial characteristics					
District 1 indicator (1 if crash occurred in District 1, 0 otherwise) [MI]	-0.426	-2.29	0.0090	-0.0102	0.0012
District 3 indicator (1 if crash occurred in District 3, 0 otherwise) [SI]	-1.413	-1.81	0.0013	0.0004	-0.0016
District 6 indicator (1 if crash occurred in District 6, 0 otherwise) [NI]	0.489	2.20	0.0079	-0.0059	-0.0020
Vehicle characteristics					
Motorcycle indicator (1 if motorcycle, 0 otherwise) [SI]	2.047	6.34	-0.0114	-0.0041	0.0155
Passenger car indicator (1 if passenger car, 0 otherwise) [SI]	-0.420	-1.91	0.0088	0.0021	-0.0109
Work zone characteristics					
Lane shift indicator (1 if lane shift work, 0 otherwise) [MI]	0.520	2.26	-0.0069	0.0074	-0.0005
Shoulder-median work indicator (1 if work on shoulder and median, 0 otherwise) [SI]	0.463	2.07	-0.0157	-0.0045	0.0202
Transition area indicator (1 if the crash occurred in the work zone's transition area, 0 otherwise) [SI]	-0.668	-1.84	0.0028	0.0008	-0.0036
Harmful event characteristics					
Harmful event on-road indicator (1 if harmful event occurred on road, 0 otherwise) [NI]	0.506	3.77	0.0390	-0.0284	-0.0106

Harmful overturn indicator (1 if harmful event occurs with overturning, 0 otherwise) [MI]	1.031	4.68	-0.0143	0.0168	-0.0025
Roadway characteristics					
Large shoulder width indicator (1 if right shoulder width between 6 to 10 ft., 0 otherwise) [NI]	0.578	2.43	0.0130	-0.0093	-0.0037
Driver characteristics					
Negligent driver indicator (1 if negligent driving, 0 otherwise) [MI]	0.314	2.14	-0.0141	0.0156	-0.0015
Over steering indicator (1 if driver actions involved over steering, 0 otherwise) [SI]	0.885	1.81	-0.0020	-0.0006	0.0026
Number of observations			1,701		
Log-likelihood at zero			-1868.739		
Log-likelihood at convergence			-1271.449		
$\rho^2 = 1 - LL(\boldsymbol{\beta})/LL(0)$			0.319		

SI = Severe Injury; MI = Minor Injury; NI = No Injury

Table 6.8. Mixed logit model with heterogeneity in mean and variance for single-vehicle work-zone crashes in Florida, 2017.

Variables	**Parameter Estimates**	**t-stat**	**Marginal Effects**		
			No Injury	**Minor Injury**	**Severe Injury**
Constant [MI]	-1.685	-9.22			
Constant [SI]	-4.334	-9.03			
Random parameter (normally distributed)					
Young driver indicator (1 if driver's age below 30 years, 0 otherwise) [MI]	-7.272	-1.97			
(*Standard deviation of parameter distribution*)	(*6.168*)	(*2.58*)			
Shoulder-median work indicator (1 if work on shoulder and median, 0 otherwise) [SI]	-1.104	-0.95	-0.0188	-0.0072	0.0260
(*Standard deviation of parameter distribution*)	(*2.727*)	(*2.72*)			
Heterogeneity in the mean of random parameter					
Shoulder-median work indicator [SI]: Rain indicator (1 if it is rainy, 0 otherwise)	-1.539	-1.75			
Heterogeneity in the variance of random parameter					
Shoulder-median work indicator [SI]: Most harmful median indicator (1 if most harmful event occurs in the median, 0 otherwise)	0.855	3.07			
Traffic characteristics					
Low traffic volume indicator (1 if average annual daily traffic is below 40,000 vehicles/day, 0 otherwise) [SI]	1.286	2.61	-0.0061	-0.0018	0.0079
Temporal characteristics					
Later months indicator (1 if crash occurred October to December, 0 otherwise) [SI]	-1.552	-2.27	0.0036	0.0012	-0.0048
Spatial characteristics					
District 2 indicator (1 if crash occurred in District 2, 0 otherwise) [NI]	0.366	1.95	0.0094	-0.0081	-0.0014
District 3 indicator (1 if crash occurred in District 3, 0 otherwise) [MI]	0.526	2.08	-0.0064	0.0067	-0.0003
District 6 indicator (1 if crash occurred in District 6, 0 otherwise) [NI]	0.492	2.11	0.0076	-0.0069	-0.0007
Vehicle characteristics					
Motorcycle indicator (1 if motorcycle, 0 otherwise) [MI]	2.656	5.96	-0.0148	0.0157	-0.0009
Passenger car indicator (1 if passenger car, 0 otherwise) [MI]	0.314	2.22	-0.0270	0.0280	-0.0010
Work zone characteristics					
Lane-closure indicator (1 if lane shift work, 0 otherwise) [MI]	-0.413	-2.08	0.0087	-0.0090	0.0003
Presence of workers indicator (1 if workers were present, 0 otherwise) [NI]	0.493	3.24	0.0240	-0.0207	-0.0033

Non-work zone indicator (1 if crashes were identified not related to work zone geometry, 0 otherwise) [NI]	-0.299	-2.09	-0.0335	0.0294	0.0041
Harmful event characteristics					
Most harmful shoulder indicator (1 if most harmful event occurs in the shoulder, 0 otherwise) [SI]	1.471	3.05	-0.0132	-0.0040	0.0172
Harmful event off-road indicator (1 if the harmful event was off road, 0 otherwise) [SI]	1.040	2.14	-0.0063	-0.0022	0.0085
Harmful overturn indicator (1 if harmful event occurs with overturning, 0 otherwise) [MI]	0.893	3.58	-0.0099	0.0105	-0.0006
Most harmful median indicator (1 if most harmful event occurs in the median, 0 otherwise) [NI]	-0.507	-2.47	-0.0100	0.0086	0.0013
Roadway characteristics					
Urban interstate indicator (1 if crash occurred on urban interstate, 0 otherwise) [SI]	1.658	3.19	-0.0111	-0.0037	0.0148
Urban toll way indicator (1 if crash occurred on an urban toll way, 0 otherwise) [MI]	-0.483	-1.65	-0.0046	0.0041	0.0005
Driver characteristics					
Negligent driver indicator (1 if negligent driving, 0 otherwise) [MI]	0.281	2.03	-0.0081	-0.0023	0.0104
Older driver indicator (1 if driver's age is 65 years and above, 0 otherwise) [NI]	-0.578	-2.47	-0.0092	0.0079	0.0013
Over steering indicator (1 if driver actions involved over steering, 0 otherwise) [MI]	0.810	2.26	-0.0036	0.0037	-0.0002
Number of observations			1,500		
Log-likelihood at zero			-1647.918		
Log-likelihood at convergence			-984.382		
$\rho^2 = 1 - LL(\boldsymbol{\beta})/LL(0)$			0.403		

SI = Severe Injury; MI = Minor Injury; NI = No Injury

Table 6.9. Temporal comparison of marginal effects for no injuries.

Variables	No Injury					
	2012	**2013**	**2014**	**2015**	**2016**	**2017**
Environmental characteristics						
Rain indicator (1 if crash occurred at the time of raining, 0 otherwise)	0.0040	0.0123	0.0065	0.0234	–	–
Dark indicator (1 if crash occurred at the time of darkness, 0 otherwise)	–	0.0116	–	–	–	–
Traffic characteristics						
Low traffic condition indicator (1 if average annual daily traffic is below 4,000 vehicles/day, 0 otherwise)	-0.0150	-0.0176	–	–	–	-0.0061
Average percent of large trucks (1 if large truck volume between 7.5% and 12.5% of all traffic, 0 otherwise)	-0.0384	-0.0241	-0.0407	–	-0.0244	–
Temporal characteristics						
Late night indicator (1 if time of day is between 8 pm to 11:59 pm, 0 otherwise)	-0.0073	–	–	–	–	–
Afternoon indicator (1 if time of day is between 12 to 2:59 PM, 0 otherwise)	–	0.0104	–	–	–	–
Earlier months indicator (1 if crash occurred January to April, 0 otherwise)	-0.0161	–	0.0110	–	–	–
Later months indicator (1 if crash occurred October to December, 0 otherwise)	–	–	–	–	–	0.0036
Spatial characteristics						
District 1 indicator (1 if crash occurred in District 1, 0 otherwise)	–	–	–	–	0.0090	–
District 2 indicator (1 if crash occurred in District 2, 0 otherwise)	–	–	–	0.0032	–	0.0094
District 3 indicator (1 if crash occurred in District 3, 0 otherwise)	–	–	–	-0.0065	0.0013	-0.0064
District 4 indicator (1 if crash occurred in District 4, 0 otherwise)	–	–		–	–	–
District 5 indicator (1 if crash occurred in District 5, 0 otherwise)	-0.0140	–	–	–	–	–
District 6 indicator (1 if crash occurred in District 6, 0 otherwise)	–	–	0.0117	0.0089	0.0079	0.0076
District 7 indicator (1 if crash occurred in District 7, 0 otherwise)	–	–	-0.0154	–	–	–
Vehicle characteristics						
Motorcycle indicator (1 if motorcycle, 0 otherwise)	-0.0204	-0.0163	-0.0137	-0.0107	-0.0114	-0.0148
Pickup truck indicator (1 if pickup, 0 otherwise)	0.0087	0.0030	–	–	–	–
Passenger car indicator (1 if passenger car, 0 otherwise)	0.0311	0.0166	–	0.0190	0.0088	-0.0270

Work zone characteristics						
Lane-closure work zone indicator (1 if lane closure, 0 otherwise)	–	–	–	–	–	0.0087
Lane shift indicator (1 if lane shift work, 0 otherwise)	–	–	–	–	-0.0069	–
Shoulder-median work indicator (1 if work on shoulder and median, 0 otherwise)	–	–	–	–	-0.0157	–
Absence of enforcement indicator (1 if no enforcement was present, 0 otherwise)	–	–	-0.0815	–	–	–
Presence of workers indicator (1 if workers were present, 0 otherwise)	–	–	–	0.0228	–	0.0240
Non-work zone indicator (1 if crashes were identified not related to work zone geometry, 0 otherwise)	–	–	–	–	–	-0.0335
Work zone geometry indicator (1 if crashes were identified as work zone geometry related, 0 otherwise)	–	0.0046	–	–	–	–
Transition area indicator (1 if the crash occurred in the work zone's transition area, 0 otherwise)	–	–	–	-0.0119	0.0028	–
Harmful event characteristics						
Most harmful shoulder indicator (1 if most harmful event occurs in the shoulder, 0 otherwise)	–	-0.0491	-0.0747	-0.0211	–	-0.0132
Most harmful median indicator (1 if most harmful event occurs in the median, 0 otherwise)	–	-0.0444	-0.0316	-0.0368	–	-0.0100
Harmful event on-road indicator (1 if harmful event occurred on road, 0 otherwise)	–	–	–	–	0.0390	–
Harmful event off-road indicator (1 if the harmful event was off road, 0 otherwise)	-0.0089	–	–	–	–	-0.0063
Harmful fixed object indicator (1 if harmful event occurred with roadside fixed object, 0 otherwise)	-0.0685	-0.0454	–	–	–	–
Harmful right shoulder indicator (1 if harmful event occurs on the right shoulder, 0 otherwise)	–	–	–	-0.0226	–	–
Harmful median indicator (1 if harmful event occurs in the median, 0 otherwise)	–	-0.0039	–	–	–	–
Harmful overturn indicator (1 if harmful event occurs with overturning, 0 otherwise)	-0.0232	-0.0250	-0.0221	-0.0121	-0.0143	-0.0099
Roadway characteristics						
Large shoulder width indicator (1 if right shoulder width between 6 to 10 ft., 0 otherwise)	-0.0276	–	–	–	0.0130	–

Urban interstate indicator (1 if crash occurred on urban interstate, 0 otherwise)	–	–	–	-0.0084	–	-0.0111
Rural interstate indicator (1 if crash occurred on rural interstate, 0 otherwise)	–	–	-0.0143	-0.0121	–	–
Urban tollway indicator (1 if crash occurred on an urban tollway, 0 otherwise)	–	–	–	–	–	-0.0046
Driver characteristics						
Old driver indicator (1 if driver's age between 50 to 65 years, 0 otherwise)	–	–	0.0086	–	–	–
Older driver indicator (1 if driver's age is 65 years and above, 0 otherwise)	–	–	–	–	–	-0.0092
Negligent driver indicator (1 if negligent driving, 0 otherwise)	-0.0291	–	–	-0.0132	-0.0141	-0.0081
Over steering indicator (1 if driver actions involved over steering, 0 otherwise)	–	–	–	-0.0025	-0.0020	-0.0036

Table 6.10. Temporal comparison of marginal effects for minor injuries.

Variables	Minor Injury					
	2012	2013	2014	2015	2016	2017
Environmental characteristics						
Rain indicator (1 if crash occurred at the time of raining, 0 otherwise)	0.0004	-0.0106	0.0008	-0.0164	–	–
Dark indicator (1 if crash occurred at the time of darkness, 0 otherwise)	–	-0.0124	–	–	–	–
Traffic characteristics						
Low traffic condition indicator (1 if average annual daily traffic is below 4,000 vehicles/day, 0 otherwise)	-0.0019	0.0150	–	–	–	-0.0018
Average percent of large trucks (1 if large truck volume between 7.5% and 12.5% of all traffic, 0 otherwise)	0.0429	0.0258	0.0458	–	0.0172	–
Temporal characteristics						
Late night indicator (1 if time of day is between 8 pm to 11:59 pm, 0 otherwise)	-0.0009	–	–	–	–	–
Afternoon indicator (1 if time of day is between 12 to 2:59 PM, 0 otherwise)	–	-0.0113	–	–	–	–
Earlier months indicator (1 if crash occurred January to April, 0 otherwise)	0.0181	–	0.0013	–	–	–
Later months indicator (1 if crash occurred October to December, 0 otherwise)	–	–	–	–	–	0.0012
Spatial characteristics						
District 1 indicator (1 if crash occurred in District 1, 0 otherwise)	–	–	–	–	-0.0102	–
District 2 indicator (1 if crash occurred in District 2, 0 otherwise)	–	–	–	0.0004	–	-0.0081
District 3 indicator (1 if crash occurred in District 3, 0 otherwise)	–	–	–	0.0072	0.0004	0.0067
District 4 indicator (1 if crash occurred in District 4, 0 otherwise)	–	–	-0.0081	–	–	–
District 5 indicator (1 if crash occurred in District 5, 0 otherwise)	0.0053	–	–	–	–	–
District 6 indicator (1 if crash occurred in District 6, 0 otherwise)	–	–	-0.0083	-0.0061	-0.0059	-0.0069
District 7 indicator (1 if crash occurred in District 7, 0 otherwise)	–	–	-0.0025	–	–	–
Vehicle characteristics						
Motorcycle indicator (1 if motorcycle, 0 otherwise)	-0.0040	0.0188	-0.0033	0.0122	-0.0041	0.0157
Pickup truck indicator (1 if pickup, 0 otherwise)	-0.0097	0.0012	–	–	–	–
Passenger car indicator (1 if passenger car, 0 otherwise)	-0.0185	0.0063	–	0.0027	0.0021	0.0270

Work zone characteristics						
Lane-closure work zone indicator (1 if lane closure, 0 otherwise)	–	–	–	–	–	-0.0090
Lane shift indicator (1 if lane shift work, 0 otherwise)	–	–	–	–	0.0074	–
Shoulder-median work indicator (1 if work on shoulder and median, 0 otherwise)	–	–	–	–	-0.0045	–
Absence of enforcement indicator (1 if no enforcement was present, 0 otherwise)	–	–	0.0912	–	–	–
Presence of workers indicator (1 if workers were present, 0 otherwise)	–	–	–	-0.0141	–	-0.0207
Non-work zone indicator (1 if crashes were identified not related to work zone geometry, 0 otherwise)	–	–	–	–	–	0.0294
Work zone geometry indicator (1 if crashes were identified as work zone geometry related, 0 otherwise)	–	0.0018	–	–	–	–
Transition area indicator (1 if the crash occurred in the work zone’s transition area, 0 otherwise)	–	–	–	0.0131	0.0008	–
Harmful event characteristics						
Most harmful shoulder indicator (1 if most harmful event occurs in the shoulder, 0 otherwise)	–	0.0425	0.0342	-0.0030	–	-0.0040
Most harmful median indicator (1 if most harmful event occurs in the median, 0 otherwise)	–	0.0384	0.0167	0.0198	–	0.0086
Harmful event on-road indicator (1 if harmful event occurred on road, 0 otherwise)	–	–	–	–	-0.0284	–
Harmful event off-road indicator (1 if the harmful event was off road, 0 otherwise)	-0.0010	–	–	–	–	-0.0022
Harmful fixed object indicator (1 if harmful event occurred with roadside fixed object, 0 otherwise)	0.0324	0.0387	–	–	–	–
Harmful right shoulder indicator (1 if harmful event occurs on the right shoulder, 0 otherwise)	–	–	–	0.0253	–	–
Harmful median indicator (1 if harmful event occurs in the median, 0 otherwise)	–	-0.0020	–	–	–	–
Harmful overturn indicator (1 if harmful event occurs with overturning, 0 otherwise)	0.0260	0.0271	0.0282	0.0140	0.0168	0.0105
Roadway characteristics						
Large shoulder width indicator (1 if right shoulder width between 6 to 10 ft., 0 otherwise)	0.0141	–	–	–	-0.0093	–

Work zone geometry indicator (1 if crashes were identified as work zone geometry related, 0 otherwise)	–	0.0018	–	–	–	–
Urban interstate indicator (1 if crash occurred on urban interstate, 0 otherwise)	–	–	–	-0.0014	–	-0.0037
Rural interstate indicator (1 if crash occurred on rural interstate, 0 otherwise)	–	–	0.0161	0.0060	–	–
Urban tollway indicator (1 if crash occurred on an urban tollway, 0 otherwise)	–	–	–	–	–	0.0041
Driver characteristics						
Old driver indicator (1 if driver's age between 50 to 65 years, 0 otherwise)	–	–	-0.0098	–	–	–
Older driver indicator (1 if driver's age is 65 years and above, 0 otherwise)	–	–	–	–	–	0.0079
Negligent driver indicator (1 if negligent driving, 0 otherwise)	0.0327	–	–	-0.0021	0.0156	-0.0023
Over steering indicator (1 if driver actions involved over steering, 0 otherwise)	–	–	–	-0.0005	-0.0006	0.0037

Table 6.11. Temporal comparison of marginal effects for severe injuries.

Variables	Severe Injury					
	2012	2013	2014	2015	2016	2017
Environmental characteristics						
Rain indicator (1 if crash occurred at the time of raining, 0 otherwise)	-0.0044	-0.0017	-0.0074	-0.0070	–	–
Dark indicator (1 if crash occurred at the time of darkness, 0 otherwise)	–	0.0008	–	–	–	–
Traffic characteristics						
Low traffic condition indicator (1 if average annual daily traffic is below 40,000 vehicles/day, 0 otherwise)	0.0169	–	-0.0051	–	–	0.0079
Average percent of large trucks (1 if large truck volume between 7.5% and 12.5% of all traffic, 0 otherwise)	-0.0046	-0.0017	-0.0051	–	0.0072	–
Temporal characteristics						
Late night indicator (1 if time of day is between 8 pm to 11:59 pm, 0 otherwise)	0.0083	–	–	–	–	–
Afternoon indicator (1 if time of day is between 12 to 2:59 pm, 0 otherwise)	–	0.0007	–	–	–	–
Earlier months indicator (1 if crash occurred January to April, 0 otherwise)	-0.0020	–	-0.0123	–	–	–
Later months indicator (1 if crash occurred October to December, 0 otherwise)	–	–	–	–	–	-0.0048
Spatial characteristics						
District 1 indicator (1 if crash occurred in District 1, 0 otherwise)	–	–	–	–	0.0012	–
District 2 indicator (1 if crash occurred in District 2, 0 otherwise)	–	–	–	-0.0036	–	-0.0014
District 3 indicator (1 if crash occurred in District 3, 0 otherwise)	–	–	–	-0.0007	-0.0016	-0.0003
District 4 indicator (1 if crash occurred in District 4, 0 otherwise)	–	–	-0.0036	–	–	–
District 5 indicator (1 if crash occurred in District 5, 0 otherwise)	0.0087	–	–	–	–	–
District 6 indicator (1 if crash occurred in District 6, 0 otherwise)	–	–	-0.0034	-0.0028	-0.0020	-0.0007
District 7 indicator (1 if crash occurred in District 7, 0 otherwise)	–	–	0.0180	–	–	–
Vehicle characteristics						
Motorcycle indicator (1 if motorcycle, 0 otherwise)	0.0244	-0.0025	0.0170	-0.0016	0.0155	-0.0009
Pickup truck indicator (1 if pickup, 0 otherwise)	0.0010	-0.0041	–	–	–	–
Passenger car indicator (1 if passenger car, 0 otherwise)	-0.0126	-0.0229	–	-0.0217	-0.0109	-0.0010
Work zone characteristics						
Lane-closure work zone indicator (1 if lane closure, 0 otherwise)	–	–	–	–	–	0.0003

Lane shift indicator (1 if lane shift work, 0 otherwise)	–	–	–	–	-0.0005	–
Shoulder-median work indicator (1 if work on shoulder and median, 0 otherwise)	–	–	–	–	0.0202	–
Absence of enforcement indicator (1 if no enforcement was present, 0 otherwise)	–	–	-0.0097	–	–	–
Presence of workers indicator (1 if workers were present, 0 otherwise)	–	–	–	-0.0088	–	-0.0033
Non-work zone indicator (1 if crashes were identified not related to work zone geometry, 0 otherwise)	–	–	–	–	–	0.0041
Work zone geometry indicator (1 if crashes were identified as work zone geometry related, 0 otherwise)	–	-0.0064	–	–	–	–
Transition area indicator (1 if the crash occurred in the work zone's transition area, 0 otherwise)	–	–	–	-0.0012	-0.0036	–
Harmful event characteristics						
Most harmful shoulder indicator (1 if most harmful event occurs in the shoulder, 0 otherwise)	–	0.0066	0.0405	0.0241	–	0.0172
Most harmful median indicator (1 if most harmful event occurs in the median, 0 otherwise)	–	0.0060	0.0149	0.0170	–	0.0013
Harmful event on-road indicator (1 if harmful event occurred on road, 0 otherwise)	–	–	–	–	-0.0106	–
Harmful event off-road indicator (1 if the harmful event was off road, 0 otherwise)	0.0099	–	–	–	–	0.0085
Harmful fixed object indicator (1 if harmful event occurred with roadside fixed object, 0 otherwise)	0.0361	0.0067	–	–	–	–
Harmful right shoulder indicator (1 if harmful event occurs on the right shoulder, 0 otherwise)	–	–	–	-0.0028	–	–
Harmful median indicator (1 if harmful event occurs in the median, 0 otherwise)	–	0.0059	–	–	–	–
Harmful overturn indicator (1 if harmful event occurs with overturning, 0 otherwise)	-0.0028	-0.0022	-0.0062	-0.0020	-0.0025	-0.0006
Roadway characteristics						
Large shoulder width indicator (1 if right shoulder width between 6 to 10 ft., 0 otherwise)	0.0135	–	–	–	-0.0037	–
Urban interstate indicator (1 if crash occurred on urban interstate, 0 otherwise)	–	–	–	0.0098	–	0.0148
Rural interstate indicator (1 if crash occurred on rural interstate, 0 otherwise)	–	–	-0.0017	0.0061	–	–
Urban tollway indicator (1 if crash occurred on urban tollway, 0 otherwise)	–	–	–	–	–	0.0005

Driver characteristics						
Old driver indicator (1 if driver's age between 50 to 65 years, 0 otherwise)	–	–	0.0011	–	–	–
Older driver indicator (1 if driver's age is 65 years and above, 0 otherwise)	–	–	–	–	–	0.0013
Negligent driver indicator (1 if negligent driving, 0 otherwise)	-0.0036	–	–	0.0153	-0.0015	0.0104
Over steering indicator (1 if driver actions involved over steering, 0 otherwise)	–	–	–	0.0030	0.0026	-0.0002

(Table 6.5), the mean of constant-term parameter specific to minor injury increased if the crash occurred in clear weather and the variance of the parameter increased if negligent driving was involved in the crash. For 2015 (Table 6.6), the mean of the constant-term parameter specific to minor injury increased if a young driver was involved and the variance of the parameter increased if the crash occurred in the rain. For 2016 (Table 6.7), the mean of the younger-driver parameter specific to minor injury increased if a motorcycle was involved and the variance of the parameter increased if the crash occurred in earlier months of the year (January to April). For 2017 (Table 6.8), the mean of the shoulder-median-work parameter specific to severe injury increased if the crash occurred in the rain and the variance of the parameter increased if the most harmful event occurred in the median. The differences in the effect of the constant term and other parameters across injury-severity observations and between years suggests considerable temporal instability in the unobserved heterogeneity over time.

Regarding environmental conditions, rainy weather was found to decrease severe injury crashes relative to other weather categories from 2012 to 2015, with the inconsistency of the marginal effects between years suggesting considerable temporal instability in this effect (Table 6.11). The fact that rainy weather is consistently associated with a lower probability of severe driver injury may reflect drivers being more cautious with their driving through work zones in inclement weather conditions (Osman et al., 2018). In addition, dark conditions at the time of work zone crashes were found to be statistically significant only in 2013, where they increased the likelihood of severe driver injuries.

Table 6.11 shows that work zones with a traffic volumes less than 40,000 vehicles/day were found to increase the likelihood of severe driver injuries in 2012 and 2017, but to decrease severe injuries in 2014, and to be statistically insignificant in other years. Table 6.10 shows work zones that had large-truck (trucks with greater than 10,000 pound gross vehicle weight) percentages between 7.5% and 12.5% of the total traffic volume were found to result in increases in minor driver-injury probabilities in 2012, 2013, 2014, and 2016, but this indicator variable did not have a statistically significant effect in 2015 and 2017. These two traffic-related variables provide information on the effect that traffic characteristics may have on speed and other factors that could affect resulting injuries in single-vehicle crashes, and they both suggest temporal instability to varying degrees.

Looking at temporal characteristics, Table 6.11 shows that crashes occurring between January 1 and April 30 were found to decrease the likelihood of severe driver-injuries in 2012 and 2014, although this early-year effect was statistically insignificant in 2013, 2015, 2016 and 2017. Crashes occurring between October 1 and December 31 were found to decrease the likelihood of severe driver-injuries only in 2017, although this late-year effect was statistically insignificant in 2012-16. In addition, crashes occurring between 8PM to midnight and 12PM to 3PM were found to be statistically significant in 2012 and 2013, respectively, with higher likelihoods of severe driver injury. These findings show an interesting statistical shift in seasonal effects over time, some of which could potentially be explained by the specific nature of the projects undertaken over time.

Turning to the Florida Department of Transportation District effects shown in Tables 6.9, 6.10, and 6.11, it was found that some districts relative to other districts were significant in only one or just a few years over the 6-year analysis period. District 1 (which includes Naples) was significant only in 2016 (resulting in a 0.0012 higher probability of severe injury). District 2 (which includes Jacksonville) was significant in 2015 and 2017 (resulting in a 0.0036 and a 0.0014 lower probability of severe injury, respectively). District 3 (which includes Tallahassee) was significant in 2015, 2016, and 2017 (resulting in a with 0.0007, 0.0016, and 0.0003 lower probabilities of severe injury, respectively). District 4 (which includes Fort Lauderdale) was significant in only 2014 (resulting in a 0.0036 lower probability of severe injury). District 5 (which includes Orlando) was significant in only 2012 (resulting in a 0.0087 higher probability of severe injury). District 6 (which includes Miami) was found to have statistically significant effects from 2014 to 2017, with marginal effects that decrease somewhat over time. The consistency of this finding over these more recent years may suggest differences in work-zone practices or driver behavior in the Miami region relative to other regions of the state. Finally, District 7 (which includes Tampa) was significant only in 2014 (resulting in a 0.018 higher probability of severe injury).

Vehicle characteristics shown in Tables 6.9, 6.10, and 6.11 indicate that, relative to other vehicle types, motorcycles, passenger cars, and pickup trucks were found to be statistically significant in work-zone crashes for 2012-17, 2012-13, and 2015-17, and 2012-13, respectively. Passenger cars were found to decrease the likelihood of severe driver injuries in 2012-13 and 2015-17. Motorcycles were found to increase the likelihood of severe rider injuries in 2012, 2014, and 2016, but to decrease severe injuries in 2013, 2015, and 2017. Pickup trucks were found to increase the likelihood of severe driver injuries in 2012 but decrease them in 2013. Passenger cars also had a notably higher probability of no injuries relative to other vehicle types (not significant in 2014) and to increase with a 0.0311 in 2012, 0.0166 in 2013, 0.019 in 2015, and 0.0088 in 2016, but had a 0.0270 higher probability of minor injuries in 2017 relative to other vehicle types.

Work-zone characteristics in Tables 6.9, 6.10, and 6.11 show that, although statistically insignificant in 2012 to 2015 and 2017, crashes that occurred in work zones with shoulder-median work resulted in a higher probability of severe injury in 2016. Lane-closure and lane-shift work-zone indicators were statistically significant only in 2017 (resulting in a higher probability of severe driver injury) and 2016 (resulting in a lower probability of severe driver injury), respectively. Some of the attributes specific to work zones, such as, enforcement, workers, and non-work zone related factors were found to be statistically significant in different years over the analysis period. The absence of enforcement was found to increase the probability of minor driver injury in 2014, although this indicator variable was found statistically insignificant in 2012, 2013, 2015 and 2017. The presence of workers was found to decrease severe and minor driver injuries in 2015 and 2017. Non-work zone attributes, particularly the factors not related to work zone geometry, were found to increase severe and minor driver injury only in 2017 and found to be statistically insignificant in 2012-16. Having the work zone geometry being identified as the contributing factor of the crash was found to be statistically significant only in 2013 (resulting in lower probability of severe driver injury). Finally, crashes that occurred in the work-zone's transition area had a higher probability of resulting in a minor driver injury in 2016 and 2017.

Overall, these work zone findings show considerable temporal instability, but it is important to exercise some caution in interpreting work zone characteristics. That is, although these characteristics are not endogenous in the classic sense (where reported crashes are used to alter the design characteristics of the work zone), the guidelines used to establish work zone characteristics are based on best practices which may have evolved from past crash data. However, this potential endogeneity would be further mitigated by temporal shifts, which would weaken the link between the effect of work zone characteristics (which were potentially influenced by past crash data) and current crash data.

Regarding the harmful-event indicators shown in Tables 6.9, 6.10 and 6.11, a harmful event involving overturning was found to be consistently statistically significant over the analysis period, with higher probability of severe and minor injuries of drivers. However, most harmful events occurring on the right shoulder and median were found to be statistically significant only in 2013 and 2015, respectively. The most harmful event occurring at shoulder indicator produced some variation in marginal effects over time with a higher probability of severe driver injury relative to other harmful events in 2013-15, and 2017, but relatively lower likelihood of severe driver injury was observed with the most harmful event occurring in the median in the same time period, reflecting some temporal instability (see Table 6.11). Further evidence of temporal instability was suggested with the finding that a harmful event with a fixed object was statistically significant only in 2012-13 but insignificant in rest of the analysis period and harmful events occurring on-road and off-road were statistically significant only in 2016, and in 2012 and 2017, respectively.

Turning to the marginal effects of roadway characteristics, although statistically insignificant in 2013-15, crashes that occurred in work zones that had large right-shoulder widths (between 6 and 10 feet) had a higher probability of resulting in a severe driver injury in 2012 but lower probability of resulting in severe driver injury in 2016 (Table 6.11). Moreover, the urban interstate indicator was found to be statistically significant in 2015 and 2017 (resulting in a higher probability of severe driver injury) and the rural interstate indicator was found to be statistically significant in 2014 (lower probability of severe driver injury) and 2015 (higher probability of severe driver injury). In contrast, the urban tollway indicator was found to be statistically significant only in 2017.

Driver characteristics also played a role in many of the years. For example, relative to other age groups, drivers between 50 to 65 and above 65 years of age were found to have a higher probability of severe injury in 2014 and 2017 (Table 6.11), respectively, but such variables were not statistically significant in other years. Relative to other reasons, if negligent driving was identified as the primary reason for the crash, a higher probability of severe injury was found in 2015 and 2017 (Table 6.11) and a higher probability of minor injury was found in 2012 and 2016

(Table 6.10), with the magnitudes of the marginal effects in these time found to be relatively close. Finally, Table 6.11 shows the driver over steering indicator was found to result in a higher likelihood of severe injury in 2015 and 2016, a lower likelihood of severe injury in 2017 (although the marginal effect was quite small), but was statistically insignificant in 2012, 2013, and 2014.

Looking at the overall injury-severity percentages, at the beginning of the analysis period (2012), the distribution of crashes among severity levels was 54.5% no injury, 36.7% minor injury and 8.8% severe injury. At the end of the analysis period (2017) this distribution among severity levels shifted to 63.9% no injury, 27.2% minor injury and 8.9% severe injury. Thus, while the percent of crashes resulting in severe driver injuries remained relatively constant over this period, there was a shift from minor injuries to no injuries (this is also reflected in Figure 2). To explore this issue further with the estimated models, the parameters from the 2012 work zone model (Table 6.3) were used to forecast 2017 crashes (using actual crash characteristics from the 2017 work zone crashes) and these predictions were compared with predictions based on the estimated 2017 parameters also using 2017 data (to account for random parameters, predictions were made using simulation as was done in the simulated maximum likelihood estimation). This predictive comparison provides an aggregate assessment of how overall injury-severity probabilities have changed over time while controlling for actual crash characteristics. In this case, the 2012 parameters predict 59.5% percent of crashes resulting in no injury in 2017 instead of the observed 63.9% (see Figure 2), 31.0% percent of crashes resulting in minor injury instead of the observed 27.2%, and 9.4% percent of crashes resulting in severe injury as opposed to the observed 8.9%. As mentioned above, in looking at Figure 2 and comparing 2012 and 2017, the observed crashes in both years have about the same percentage of severe injuries, but 2017 has a lower percentage of injury crashes and a higher percentage of no-injury crashes. The predictive comparison also shows that 2012 parameters predict the same trend, although not to the magnitude that was observed (but 2012-parameter predictions also show an uptick in severe injuries relative to the observed value). This suggests that some of the observed injury-proportion trends shown in Figure 2 are due to the specific characteristics of the observed crashes, but that fundamental temporal changes in the influence that crash characteristics are having on injury probabilities are also playing a role. Some of these changes may be the result of improvements in vehicle safety features over time or temporal changes in the risk profiles of individuals becoming involved in crashes as discussed in Mannering et al. (2020). Factors such as these will show up as temporal shifts in unobserved heterogeneity in the models estimated herein.

6.6 Summary and Conclusions

Using single-vehicle crash data in work-zones in Florida from 2012 to 2017, this study used a random parameters logit model (with heterogeneity in mean and variance) to explore the stability of factors determining driver-injury severities for each of year of analysis period: 2012-17. Three driver injury levels were considered: no injury, minor injury (combining possible injury and non-incapacitating injury), and severe injury (combining incapacitating injury and fatal injury). The estimated models find a wide variety of factors significantly influencing driver-injury outcomes including environmental attributes, overall traffic and truck volume, spatial and temporal characteristics, vehicle type, work zone characteristics, harmful event characteristics, roadway geometry, and driver factors.

Although there are some consistencies between each of the years (of the 42 variables found to be statistically significant in at least one of the year, only two of these were found to be statistically significant in all of these years), likelihood ratio tests show that the estimated parameters were temporally unstable for driver injuries in Florida work-zone-related crashes from one year to the next over the 2012-17 analysis period. The cause of this instability is not necessarily clear. Past research has tended to argue that this instability is largely the result of fundamental changes in driving behavior (Mannering, 2018). However, this may not necessarily be the case with work-zones. Although Florida work-zone practices and traffic-control procedures did not change significantly over this time period considered, each work zone has a unique set of characteristics and, with the sample of work zones changing from one year to the next as highway maintenance and construction is undertaken, this work-zone variation could be a substantial source of the observed temporal instability. That is, although we do have a set of variables that describe work-zone characteristics, the unobserved heterogeneity from one work zone to the next could be a major source of the observed temporal instability. This is unlike traditional highway-section data (based on the same highway segments year after year) where the unobserved characteristics of the individual sections do not change drastically over time. The unique characteristics of individual work zones greatly complicate the interpretation of the temporal instability findings.

Chapter 7

A National Estimate of the Zero-Price Effect for Public Electric Vehicle Charging: A Stated Preference Approach

Trang Loung, Michael Maness, Smart Student

7.1 Introduction

Although, many researchers warn about the EVs being only as clean as the energy source, governments recognize their potential in the form of reducing greenhouse effect or atmospheric and noise pollution and thus exploration of factors that may contribute to their further uptake (such as free charging) is worthy investigating.

Electric vehicle (EV) sales and electric vehicle infrastructure have grown extensively over the last decade. No longer dependent on home or workplace charging, public EV charging infrastructure has grown with varying locations, charging network brands, and pricing. The impact of free charging EV usage behavior has seen limited research despite an expanding body of literature on electric vehicles (EVs). Research has shown that "zero cost" items and services tend to be perceived differently (Ariely, 2008). The zero-price-effect can be defined as a phenomenon in which a product becomes more desirable and its demand increases significantly when its price is precisely zero, contrasted to when its price is marginally higher than zero – i.e. a discontinuity in the price to value relationship occurs at zero price (Shampanier et al., 2007; Fruchter et al., 2011; Nicolau et al., 2012). Although individuals tend to perceive free items as having additional benefit just from being free, it was found that as the price of goods and services decreases, their benefits tend to increase making them more attractive choice but once the good or service obtains a price of zero a discontinuity occurs.

Although extensively studied in economics and marketing, research on the zero-price effect in EV charging is scarce (Maness and Lin, 2019). In terms of charging related behavior, Nicholas and colleagues (2019) found substantial differences in charging activity between free and non-free workplace charging stations and concluded that there was evidence that workplace charging can play a larger role in the charging ecosystem of the future, especially for those without home charging. When analyzing results from Chakraborty et al. (2019), their model estimated a monetary value of free workplace charging at 0.39 ¢/kWh per charge. Past work has shown that a short-term free public charging program could possibly increase plug-in electric vehicle sales, decrease oil consumption, and decrease greenhouse gas emissions (Maness and Lin, 2019). There are multiple reasons for the lack of analysis on the value of free charging. As Daina (2014) observed in Zoepf et al. (2013) work, the charging price tends to have low variation which leads to difficulties in estimation. Additionally, the zero-price effect must be explicitly accounted for modeling charger choice and past studies have not done this (Zoepf et al. 2013, Dania 2014, Wen et al. 2016, Latinopoulos et al. 2017). Maness and Lin (2019) explored the possible vehicle ownership and greenhouse gas emissions impact of a national free EV public charging policy. Although they found that long-term costs could potentially reach about $40/metric ton CO_2 saved when an $1/charge zero-price effect was assumed, their study was hampered by a lack of zero-price effect data.

Seeking to fill the data gap, this study establishes an early estimate of the value of free charging for EVs in the United States. To investigate the phenomenon of free charging and arrive at its monetary estimation, a stated preference survey was designed. The collected data was then used to estimate latent class models that assume full attribute attendance and account for attribute non-attendance. The remaining sections of the chapter include literature review on the zero-price effect and stated preference surveys in the context of electric vehicles charging, survey design, and sample weighing. This is followed by methodology, results, analysis, and discussion.

7.2 Literature Review

7.2.1 Why a zero-price effect?

The ‘zero price effect’ can be defined as a phenomenon in which a product becomes more desirable when its price is precisely zero compared to when its price is even slightly higher than zero. Diamond and Sanyal (1990) explained the zero-price effect as a result of prospect theory. They studied people’s choice of action towards supermarket discounts when a product is offered at a discount versus when a free item of equal value is bundled with the product. Diamond and Sanyal proposed that people chose the free product bundle more often because they treated it as a gain in the bundle (i.e. “I gained a can of soup”) whereas the discount was seen as just a reduced loss (i.e. “I paid less for the sauce”).

In other work, Heyman and Ariely (2004) concluded that although people normally operate in a market norms mindset when shopping, when a product is offered for free, the mindset switches to social norms. Under the social norms mindset, they see a free offer as a gift so they believe they should appreciate it when someone gives them a gift. Shampanier et al. (2007) experimented with three additional theories for explaining the zero-price effect: transactional cost, mapping difficulty, and affect. In their experiments using gift cards and candies, they found support for affect as a compelling reason for the zero-price effect. Table 7.1 summarizes the five main proposed theories for the zero-price effect.

Table 7.1. Theorized reasons for the zero-price effect

Reason	Description	Source
Affect	People based their choice on the presence of positive feeling when obtaining a free product/service	Shampanier et al. (2007)
Transactional Cost/Effort	The cost or effort required to take part in a transaction is inhibitive, but the demand increases because free product requires no effort on the individual’s part to transact	Shampanier et al. (2007)
Mapping Difficulties	Individuals’ inability to calculate the net benefit value	Shampanier et al. (2007)
Social Norms	Expectations related to acceptable behavior when receiving gifts may cause decrease in demand for larger quantities of the gift	Heyman and Ariely (2004)
Prospect Theory	Asymmetry caused by Free seen as a benefit whereas a discount is seen as a reduced loss	Diamond and Sanyal (1990)

7.2.2 EVs charging behavior

The current chapter aims to translate the abovementioned ‘zero-price effect’ into the setting of EV charging and consumer behavior. Prior literature has addressed this topic to varying degrees and there have been some studies that attempted to understand the EV public charging choices.

Although, they have not exclusively addressed the value of free charging and its impact on vehicle charging behavior, multiple studies have already confirmed the presence of significant heterogeneity in driver’s behavior and their charging choices (Franke and Krems , 2013; Yang et al., 2016; Zoepf et al., 2013; Sun et al., 2015; Daina et al., 2017). Literature suggests that studies aiming to gain more insights into the heterogeneity of the charging behavior fall into two categories, those that discuss heterogeneity in charging patterns (home, workplace, and public charging) and those that study heterogeneity in the factors that drive charging decisions (e.g., pricing, and behavioral patterns) (Wolbertus and Gerzon, 2018).

There is a wide variety of factors that are being studied with regards to EV charging behavior. In terms of the approach to data collection, prior work has mostly used stated or revealed preference methods. Some researchers used them to gain more insights into factors determining whether charging is initiated in the first place (i.e. Zoepf et al., 2013; Yu and MacKenzie, 2016; Ge and MacKenzie, 2017; Ge et al., 2018) (charging price was included as an independent variable in some, but not all, of the papers) while others studied its frequency and timing (i.e. Daina et al., 2015; Sun et al., 2015). The modeling approaches also vary, however there appears to be less variation and the most common modeling techniques include discrete choice models such as multinomial logit and its extension mixed logit models (i.e. Zoepf et al., 2013; Yu and MacKenzie, 2016; Sun et al., 2015), latent class models (Yu and MacKenzie, 2016; Wen et al. 2016; Kim et al., 2017) as well as extensions and variations (nested logit models as in Yang et al. (2016) or latent class hazard models such as in Kim et al., 2017) To study charging behavior, Daina et al. (2017) proposed a random utility model rooted in discrete choice analysis that is based on the theoretical framework of random utility (Ben-Akiva and Lerman,1985; Train, 2009). Such model assumes that the decision makers choose the alternative that maximizes their own utility from a set of mutually exclusive alternatives and a user will chose a particular alternative if all the other alternatives in their choice set offer a lower utility.

To model the charging behavior response to a time-based charging fee Wen et al. (2016) used mixed and latent class models, in which they included the price of the charging session based upon a stated preference survey among EV drivers. In the latent classes they did find differences in price sensitivity between respondents. Latent class choice models were not only used to capture the unobserved heterogeneity in the data but were also a particularly useful tool in terms of dividing the respondents into different behavioral groups.

Other studies analyzed the influence of pricing on more general charging behavior. Latinopoulos et al. (2017) looked into price in relation to charging decisions combined with parking reservations. They found that EV drivers have higher willingness to pay more to ensure charging station availability and found a connection between free charging and general charging behavior. They also concluded that EV drivers that have been charging their vehicle for free were strongly inclined towards the safer (less risky) charging option. From somewhat a different angle, Wolbertus et al. (2018) found that the EV charging price and the willingness-to-pay) was associated, to a large degree, with time of the day, charging speed, charging time, vehicle type (taxi vs. for personal use), and even parking needs. Other work done by (Pan et al., 2019), in which the authors considered risk attributes and attribute non-attendance, indicated that charging station operators could modify prices to attract EV drivers and increase their charger utilization rates, as charging price and parking price significantly influence EV drivers' stated charging choices. Same authors also concluded that the type of charging (fast or slow charging) and dwell time do not seem to significantly influence charging decisions.

As mentioned in the beginning of the chapter, Chakraborty et al. (2019) estimated a monetary value of workplace charging and arrived at 0.39 ¢/kWh for free workplace charging of BEV (battery electric vehicles) and PHEV (plug-in hybrid electric vehicles) which creates an opportunity to investigate the zero-price effect in non-workplace settings.

7.3 Survey Design and Methodology

This section will describe the survey design and methodology. It begins by describing the probability-based internet panel used to sample US households. Then, the survey sections are briefly described

7.3.1 Sample design: Probability-based internet panel

Since this study sought to develop a national estimate of the zero-price effect, the sampling plan for this study used AmeriSpeak®. Funded and operated by NORC at the University of

Chicago, AmeriSpeak is a probability-based panel designed to be representative of the US household population. Randomly selected US households are sampled using area probability and address-based sampling, with a known, non-zero probability of selection from the NORC National Sample Frame. Sampled households are contacted by US mail, telephone, and field interviewers. The panel provides sample coverage of approximately 97% of the U.S. household population. Those excluded from the sample include people with P.O. Box only addresses, some addresses not listed in the USPS Delivery Sequence File, and some newly constructed dwellings. Households without conventional internet access but having web access via smartphones can participate in AmeriSpeak surveys by web.

7.3.2 General survey methodology and design

Two stated preference surveys exploring aspects of free charging were conducted between June and August 2020. Using a probability-based internet panel, 4,230 panelists were invited to participate, and 1,097 respondents chose to participate and completed the survey. Respondents were contacted up to four times by email with follow-up email reminders, as necessary. The median completion time for the survey was eleven minutes. An overview of survey design methodology for the whole survey is presented in Table 7.2. The survey consisted of four sections:

Stated Choice Experiment – Charger Choice Experiment or Vehicle Choice Experiment

Household Characteristics

Commute characteristics

Vehicle characteristics

Respondents' sociodemographics and some household characteristics were collected by NORC upon entrance into the panel and this information is updated annually.

Table 7.2. Survey design methodology

Characteristic	**Description**
Timeframe	July 10 – August 25, 2020
Target population	Civilian and non-institutionalized adults who are residents of United States households (18 years and older)
Sampling frame	NORC National Sample Frame – a two stage frame with metro area/county first stage and census tract/block group second stage
Sample design	Probability-based internet panel
Sample size	4230 panelists invited
Completed interviews	1097 respondents total with 832 respondents completing charger choice experiment
Non-response Follow-up	Four reminder emails sent in July and August 2020
Respondent Compensation	$3 cash equivalent per response
Use of interviewer	Self-administered
Mode of administration	Self-administered via the internet
Computer assistance	Internet-based survey
Reporting unit	One person (aged 18 or higher) per household
Frequency	One-time response collection
Levels of observation	Individual, household
Survey design platform	Qualtrics

7.3.3 Charger choice stated preference design

The experiment begins by introducing basic definitions and terms related to EVs including their power source, batteries, and charging needs. This was followed by another block that explained the meaning of the attributes used in the experiment: *charging cost*, *charging time*, *detour time*, and *amenities at charging station*. The respondent was instructed of the charging context as follows:

> *Imagine that you are on your way home from a leisure event. Your schedule is free over the next three hours. You are driving an electric car that can only be charged at electric charging stations. When you are 20 miles from home, you notice that your estimated range remaining is 10 miles. You will need to charge your vehicle before you get home.*

Because it was expected that most respondents would have no experience with driving an EV, the charging context chosen was done to ensure that respondents would not need to have experience with range anxiety and state of charge. Additionally, the long time period with a free schedule ensures that respondents could reveal their preferences and tradeoffs for time and cost in an unconstrained context.

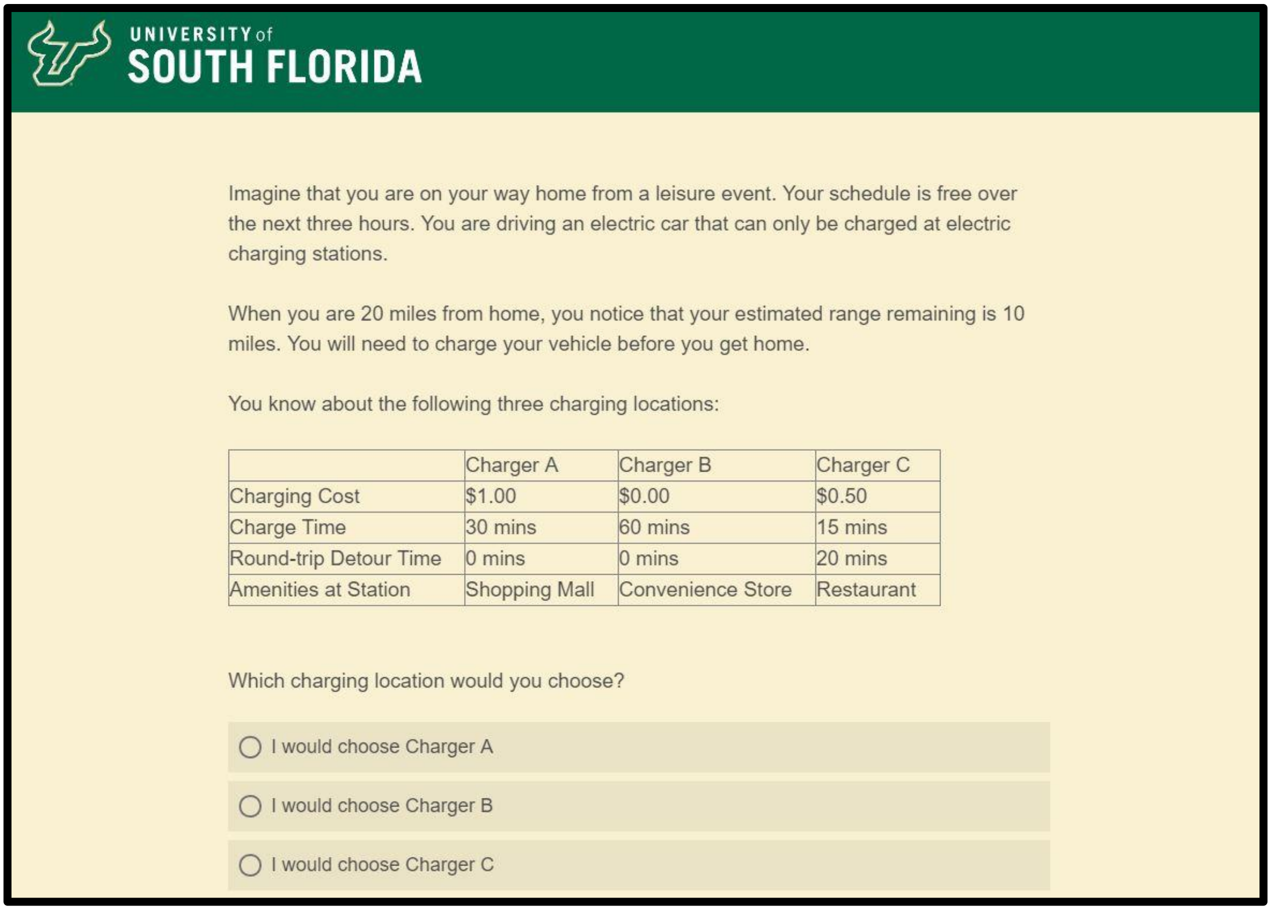
In each scenario, respondents are presented with three charging locations labeled: Charger A, Charger B, and Charger C. Figure 1 present a sample scenario presented to respondents using a personal computer. Mobile users were presented with a stacked 3-table design with each alternative confined to a single table to minimize left-right scrolling for respondents.

The narrative of the hypothetical scenarios was designed to overcome the challenges and the limitations that are encountered frequently such as, in this case, availability of public chargers for EVs and number of people owning a battery electric vehicle (BEV). The respondents were asked to assume that they were driving a battery powered electric vehicle and that they chose to charge their electric vehicle on a trip back home from a leisure activity. They were then asked to choose between three charging station locations based on varying combinations from the four attributes: charging cost, charging time, detour time, and amenities at station (see Table 7.2).

The experimental design for the stated choice experiments is a modified orthogonal design. The initial orthogonal design was generated using NGENE. The four attributes and their attribute levels are presented in Table 7.2. A trial survey was administered to 99 MechanicalTurk users in November 2019. Scenarios were redesign when less than 5 percent of respondents chose a particular charger alternative. The final experimental design used is presented in Table 7.3 with design modification noted for each redesigned scenario.

Table 7.3. Choice experiment attribute levels

Attribute	Attribute Levels
Charging cost	\$0.00, \$0.10, \$0.50, \$1.00, \$2.00, \$4.00, \$5.00, \$6.00
Charging time	15 mins, 30 mins, 60 mins
Round-trip detour time	0 mins, 20 mins
Amenities at station	Parking lot, Restaurant, Convenience store, Mall



UNIVERSITY of SOUTH FLORIDA

Imagine that you are on your way home from a leisure event. Your schedule is free over the next three hours. You are driving an electric car that can only be charged at electric charging stations.

When you are 20 miles from home, you notice that your estimated range remaining is 10 miles. You will need to charge your vehicle before you get home.

You know about the following three charging locations:

	Charger A	Charger B	Charger C
Charging Cost	$1.00	$0.00	$0.50
Charge Time	30 mins	60 mins	15 mins
Round-trip Detour Time	0 mins	0 mins	20 mins
Amenities at Station	Shopping Mall	Convenience Store	Restaurant

Which charging location would you choose?

- I would choose Charger A
- I would choose Charger B
- I would choose Charger C

Figure **7.**1. An example of the charger choice experiment.

The 24 scenarios were divided into 2 blocks of 12 scenarios each. Respondents were randomly assigned to one block for the choice experiment.

Table 7.4. Experimental design

	Charger A				Charger B				Charger C					
Scenario	Price	Charge Time	Detour Time	Amenity	Price	Charge Time	Detour Time	Amenity	Price	Charge Time	Detour Time	Amenity	Block	Modification
1	0	15	0	-3	1	15	20	1	2	15	0	3	2	Alt3: CT to 15 mins from 30
2	0.5	60	0	1	0.1	60	0	-3	0	60	0	3	1	Alt2: CT to 60 mins from 30
3	0.1	60	0	-1	6	15	0	-3	0	30	20	1	2	Alt2: Price to $6 from $0, Alt3: CT to 30 mins from 60
4	0	60	0	3	6	15	20	-3	0	30	20	-1	2	Alt2: CT to 15 mins from 60
5	0.5	30	0	1	0	15	20	-1	0.1	30	0	-3	1	Alt2: CT to 15 mins from 60, DT to 20 mins from 0, cost from $0.10 to $0
6	0.1	30	0	-3	0.1	15	20	3	6	30	20	-3	1	
7	5	15	0	-1	0	30	20	-1	6	60	20	-3	2	
8	2	30	0	3	0.5	15	20	-1	0.5	30	0	1	2	
9	6	30	0	3	2	30	0	1	5	30	0	-1	1	All Alts: CT to 30 mins from 15
10	1	30	0	-1	0	60	0	1	0.5	15	20	-1	2	Alt2: Price from $0.50 to $0
11	2	15	0	1	6	30	20	-3	0	30	0	-3	1	Alt3: CT to 30 mins from 15
12	0	60	0	-3	2	60	0	-1	2	30	20	3	1	Alt1: Price to $0 from $1, Alt3: Price to $2 from $0, Alt2: DT to 0 mins from 20
13	0	60	20	1	4	30	20	3	2	60	20	-1	1	Alt2: Price to $4 from $0
14	0	15	20	-3	5	15	0	1	1	30	0	-1	2	Alt2: CT to 15 mins from 30 and DT to 0 mins from 20
15	0	30	20	3	1	60	0	-3	2	60	0	3	1	Alt3: DT to 0 mins from 20, price from $5 to $2
16	0.5	15	20	-1	6	15	0	3	1	15	20	1	1	
17	0	30	20	-1	0	15	20	-3	5	30	20	1	1	Alt3: Price from $5 to $2
18	0.1	15	20	3	1	30	0	-1	2	15	0	1	2	
19	5	15	0	1	0.1	30	0	1	0	15	20	3	1	Alt3: Price to $0 from $0.10; Alt2: CT to 30 mins from 60, price to $0.10 from 0, DT to 0 mins from 20; Alt1: CT to 15 mins from 30 and DT to 0 mins from 20
20	2	30	20	-3	5	15	20	3	0.5	60	0	1	2	Alt2: CT to 15 mins from 60, Alt1: Price to $2 from $6
21	2	60	20	-1	5	15	0	-1	1	60	0	-3	2	
22	1	30	20	3	0	60	0	3	6	15	0	3	2	Alt1: CT to 30 mins from 60, cost to $1 from $5
23	0	60	0	-3	6	30	0	3	0.1	60	0	-1	1	Alt1: Price to $0 from $6, DT to 0 mins from 20; Alt2: Cost to $6 from $2
24	1	15	20	1	0.5	30	0	1	0	15	20	-3	2	
Note: Alt = Alternatives														

7.4 Sample Weighting

7.4.1 Weighting procedure summary

The analysis of the choice experiment will use study-specific sampling weights. Firstly, *panel base sampling weights* for all sampled housing units are computed as the inverse of probability of selection from the NORC National Frame. These weights are adjusted to account for non-response and unknown eligibility and then post-stratified to external household counts from the Current Population Survey. These weights are applied to each eligible adult in the household and adjusted to account for nonresponding adults within the household. Finally, panel weights are raked to Current Population Survey population totals associated with age, sex, education, race/Hispanic ethnicity, housing tenure, telephone status, and Census Division. The weights adjusted to the external population totals are the final panel weights.

Study-specific base sampling weights are derived using a combination of the final panel weight and the probability of selection associated with the sampled panel member. Adjustment is performed to account for within panel non-response and raked using the same characteristics used in raking the panel weights. Raking and re-raking is done during the weighting process such that the weighted demographic distribution of the survey respondents resembles the demographic distribution in the target population.

7.4.2 Sample descriptive statistics

The sample characteristics (weighted and unweighted) are presented in Table 7.5.

.

Table 7.5. Sample characteristics (n=832)

	Weighted	**Unweighted**
Characteristic	**Mean**	**Mean**
Gender		
Male	0.48	0.48
Female	0.52	0.52
Age		
18-29	0.20	0.19
30-44	0.26	0.35
45-59	0.24	0.20
60+	0.30	0.26
Education		
No HS diploma	0.10	0.06
HS graduate or equivalent	0.28	0.16
Some college	0.28	0.45
BA or above	0.34	0.33
Driven EV/ Not Driven EV		
Not Driven EV	0.88	0.88
Driven EV	0.12	0.12
Employment		
Employed Full Time	0.52	0.54
Employed Part Time	0.12	0.12
Retired	0.19	0.16
Student (not employed for pay)	0.04	0.03
Disabled (not employed for pay)	0.04	0.04
Not employed for pay	0.07	0.07
Other	0.02	0.02
Region		
New England	0.05	0.05
Mid-Atlantic	0.13	0.07
East North Central	0.14	0.16
West North Central	0.06	0.09
South Atlantic	0.20	0.18
East South Central	0.06	0.05
West South Central	0.12	0.10
Mountain	0.08	0.11
Pacific	0.16	0.20
Metro/Not Metro		
Metro Area	0.82	0.84
Non-Metro Area	0.18	0.16

Household Size		
1	0.12	0.13
2	0.32	0.30
3	0.16	0.16
4	0.14	0.14
5	0.10	0.10
6	0.16	0.17
Household Annual Income		
Less than or equal to $25,000	0.19	0.24
$25,001-$50,000	0.22	0.28
$50,001-$100,000	0.34	0.34
$100,001-$200,000	0.19	0.11
More than $200,000	0.03	0.03

7.5 Methodological Approach

The modeling approach used in this study is based on random utility models. In a preference space, each individual *n* is assumed to have a deterministic utility of each charger ($y \in \{A, B, C\}$) that takes the following general form:

$$V_{ni} = \alpha_i + \beta_p P_i + \beta_f F_i + \beta_{ct} CT_i + \beta_{dt} DT_i + \beta_r R_i + \beta_m M_i + \beta_{cs} CS_i$$

Where:

- α_i is an alternative-specific constant for alternative *i*
- *P* denotes the charge price ($),
- *F* is an indicator for a free charger (charge event), i.e. charge price equals $0.00,
- *CT* denotes the charging time (mins),
- *DT* denotes the detour time (mins),
- *R*, *M*, *SC* are indicator for nearby amenities at the charge station: restaurant, mall, and convenience store respectively, and
- $\beta_p, \beta_f, \beta_{ct}, \beta_{dt}, \beta_r, \beta_m, \beta_{cs}$ are model parameters which are generic parameters.

The zero-price effect may be derived by obtaining the money value of a free charge event. With units of \$/charge, the zero-price effect can be derived in preference space as follows:

$$ZPE = \frac{\frac{\partial F_i}{\partial V_i}}{\frac{\partial P_i}{\partial V_i}} = \frac{\beta_f}{\beta_p}$$

Estimating the value of detour time and charge time can be done similarly.

To arrive at the most accurate estimation of the zero-price effect, each independent variable was systematically tested and the final parameter values were obtained from estimations of latent class models. Latent class models are capable of capturing unobserved heterogeneity in the data and have been widely used to model similar datasets (for discussion see Greene and Hensher, 2003). As noted by Wolbertus and Gerzon (2018), models that assume continuous distribution of the preference parameters (e.g. mixed logit) are not capable of connecting the heterogeneity to discretely defined group of users. They also argued that latent class models were best suited for studying such preferences since they can provide richer insights for policy by enabling easier interpretation of the heterogeneity among respondents.

Because the focus of this chapter is to estimate the willingness-to-pay in the context of public EV charging, only a brief description of the modeling approach is described and additional methodological details are available in Hensher et al. (2015) and Washington et al., (2020).

Estimation of the zero-price effect and other willingness-to-pay measured are derived from estimating latent class models. The individuals in the sample are divided into C distinct classes with preferences varying across the classes. Allocating observations to specific classes allows to capture class-specific unobserved heterogeneity (Xiong and Mannering, 2013) without making distributional assumptions (as is required in traditional random parameter models). Latent class models are also readily estimated with maximum likelihood procedures (see Greene and Hensher (2003) and Hensher et al. (2015) for details) and the log-likelihood function is formulated as:

$$\ln L = \sum_{i=1}^{N} \ln P_i = \sum_{i=1}^{N} \ln \left[\sum_{q=1}^{Q} H_{iq} \left(\prod_{t=1}^{T_i} P_{it|q} \right) \right]$$

Where H_{iq} denotes the prior probability for a class q for individual i and $P_{it|q}$is the probability for the specific choice made by an individual i in choice situation t conditional on being in class q.

This study also accounts for attribute non-attendance through a latent class formulation. First, an endogenous attribute attendance model was estimated with 5 potential non-attendance attributes (price, free price, detour time, charging time, and amenities) – i.e. 2^5 model. It was found that amenities were non-attendant in 85% of the sample, so it was decided to limit the use of amenities in one of the model formulations. Another endogenous attribute attendance model was estimated with 2^4 classes (amenities were not included) which showed nearly equal non-attendance among the remaining attributes. Then model specifications were tested and modified such that economic consistency and interpretation was prioritized as well as model fit.

7.6 Modeling Results

This section describes the results from exploratory latent class analysis with full attribute attendance as well as confirmatory latent class models that account for attribute non-attendance. All models are estimated using LatentGOLD version 5.1. The models were estimated using the following estimation conditions:

- Convergence Limits – EM Tolerance: 0.01
- Convergence Limits – Tolerance: 1e-08
- Iteration Limits – EM: 4000
- Iteration Limits – Newton-Raphson: 500
- Start Values – Random Sets: 1200

- Start Values – Iterations: 400
- Start Values – Tolerance 1e-05

7.6.1 Models with full attribute attendance

Exploratory latent-class analysis was performed. Goodness-of-fit and mean zero-price effect are reported in Table 7.6. From the base MNL model up to five classes, the models exhibited a zero-price effect across all classes. All classes exhibited economically consistent behavior as they each had a negative price coefficient and non-negative free price indicator. From six classes and up, the models generally exhibited one class with a small negative price preference which caused unrealistically high willingness-to-pay estimates. Additionally, some of these models also exhibited preference behavior where individuals experienced a disutility from a free price. Since these two cases disagree with economic theory, classes with estimates under two cases were excluded in the calculation of the mean zero-price effect shown in Table 7.6. In terms of model fit, the AIC is seen to increase steadily across all models. BIC acts similarly but decreases between 11-classes and 12-classes. This study postulates that the cause of these inconsistencies is likely due to attribute non-attendance as the models are picking up on individuals who do not care about some attribute(s).

Table 7.6. Exploratory latent class analysis summary

Model	Log-likelihood	BIC	AIC	Parameters	Pseudo R^2	Mean ZPE
MNL	-10213.69	20487.8973	20487.90	9	0.080	-$1.23
2-class LC	-9778.52	19684.7985	19684.80	19	0.165	-$1.67
3-class LC	-9521.68	19238.3456	19238.35	29	0.230	-$1.33
4-class LC	-9354.86	18971.9472	18971.95	39	0.254	-$0.91
5-class LC	-9247.77	18825.0076	18825.01	49	0.287	-$0.79
6-class LC	-9163.91	18724.5338	18724.53	59	0.305	-$0.62*
7-class LC	-9105.19	18674.3283	18674.33	69	0.319	-$0.63*
8-class LC	-9052.60	18636.3728	18636.37	79	0.334	-$0.70*
9-class LC	-8997.02	18592.4575	18592.46	89	0.343	-$0.89*
10-class LC	-8948.18	18562.0186	18562.02	99	0.347	-$1.12*
11-class LC	-8913.94	18560.7714	18560.77	109	0.354	-$1.13*
12-class LC	-8881.04	18562.2236	18562.22	119	0.359	-$0.87*
* denotes the model contained economically inconsistent classes; the calculated class-specific ZPE value was set to 0 for these classes						

7.6.2 Attribute non-attendance without amenities

This section describes attribute non-attendance models were estimated without amenities – with the best model specification shown in Table 7.7. Across all classes except for class 7, price, detour time, and charge time are observed to induce disutility while free charge induces positive utility – unless fixed to no effect. Generally, detour time exhibited greater disutility than charge time. This result is expected since individuals can use charge time to accomplish other tasks (e.g. shopping, talking, eating), while detour time is mostly constrained to only driving.

Heterogeneity in full attribute attendance (excluding amenities) is observed across classes 1 through 4. About 77% of respondents are estimated to exhibit full attribute attendance. A plurality of respondents (42%) were in class 1 with high zero-price effects ($1.93/charge) and below average values of time. Class 2 respondents exhibited low zero-price effects ($0.13/charge) and above average values of time. Class 3 respondents were observed to have a zero-price effect of $0.91/charge with above average valuation of detour time and low values of charging time.

Class 4 respondents (with a 10% class share) had large zero-price effects ($6.91/charge) and exhibited high values of time, with value of charging time about one-fifth larger than value of detour time.

Non-attendance of free charge (Classes 5 and 8) occurs across about 15% of respondents. Accordingly, these respondents have no zero-price effect. Class 5 respondents (~6%) showcased low sensitivity to detour and charging time. Class 8 respondents have non-attendance to price, free price, and detour time with almost 12% of respondents observing this state. Class 6 respondents exhibited non-attendance to detour time and near non-attendance to charge time. Non-attendance for time attributes may be attributed to the leisurely nature of the scenario's context; even in the longest charge event, respondents still had 100 minutes remaining before their schedule was no longer free.

Non-attendance to price (class 7) was observed among about 1% of respondents. These respondents had strong sensitivity to free prices with a price worth over 20 minutes of charge time and almost 20 minutes of detour time.

Table 7.7. Latent class non-attendance model without amenities

	Class1		Class2		Class3		Class4		Class5		Class6		Class7		Class8	
Attributes	Est.	t-stat	Est.	t-stat	Est.	t-stat	Est.	t-stat	Est.	t-stat	Est.	t-stat	Est.	t-stat	Est.	t-stat
Price ($)	-0.11	-5.16	-0.79	-4.12	-1.30	-10.19	-0.20	-1.22	-1.00	-5.35	-3.72	-2.70	--	--	--	--
Free Charge Indicator	0.21	2.89	0.10	0.50	1.18	2.99	1.40	1.26	--	--	0.72	1.64	-1.08	-7.26	--	--
Charge Time (10-mins)	-0.06	-1.80	-1.27	-6.32	-0.86	-6.13	-1.74	-1.73	-0.82	-4.98	-0.02	-0.17	-0.28	-4.60	-0.55	-5.64
Detour Time (10-mins)	-0.16	-3.73	-2.26	-6.58	-2.62	-7.15	-2.04	-1.79	-0.17	-0.66	--	--	0.39	6.30	--	--
ASC Charger B	0.24	4.50	0.24	2.07	0.40	1.88	0.60	0.83	-0.09	-0.29	-0.65	-1.97	-3.83	-4.01	-0.05	-0.47
ASC Charger C	-0.07	-1.35	0.06	0.39	0.33	1.23	0.20	0.42	0.22	1.28	0.22	0.57	-8.00	-13.67	0.36	2.04
Class Membership Model																
Class Probability	43.7%		15.5%		12.2%		5.8%		6.0%		4.3%		0.9%		11.6%	
Willingness-to-Pay																
Zero-Price Effect	$1.93		$0.13		$0.91		$6.91		--		$0.19		--		--	
Value of Detour Travel Time	$8.82		$17.12		$12.08		$60.44		$1.02		--		--		--	
Value of Charging Time	$3.09		$9.61		$3.96		$51.50		$4.95		$0.03		--		--	
Model Statistics																
Log-likelihood	-9180.94															
AIC	18459.87															
BIC	18691.34															
McFadden R2	0.309															
Number of Observations	9984															
Number of Individuals	832															
Number of Parameters	48															

7.6.3 Attribute non-attendance with amenities

Attribute non-attendance models were estimated with amenities as well, with the best model specification shown in Table 7.8. Across all classes except for class 8, price, detour time, and charge time are observed to have induce disutility while free charge induces positive utility. Heterogeneity in full attribute attendance (including amenities) is observed across classes 1 and 2 with 20% of respondents exhibiting full attribute attendance. Class 1 respondents exhibited a low zero-price effect ($0.23/charge) with above average valuations of time. Class 2 respondents (~7% of sample) had low zero-price effects and low valuations of time (about $2 for detour time and $4.50 for detour time). The remaining classes include attribute non-attendance for amenities.

Attribute non-attendance to only amenities (classes 3, 4, and 5) was observed in 61% of respondents. Class 3 respondents exhibited a high zero-price effect ($2.26/charge) with low valuations on time. A plurality of respondents was in Class 3 with 42% of respondents showing this preference structure. Class 4 respondents exhibited a zero-price effect of $0.88/charge with an average value of detour time of about $12/h. Class 5 respondents exhibited a high zero-price effect of $4.05 and high valuations of time (about $37 and $38 for detour and charge time respectively). Seven percent of respondents exhibited this preference structure.

Attribute non-attendance to a free price and amenities behavior (class 6) was observed across 14% of the sample. These respondents had no zero-price effect and above average values of time. Attribute non-attendance to amenities, charge time, and detour time (class 7) was observed across 4% of the sample. These respondents exhibited a low zero-price effect ($0.24/charge) and thus were largely only concerned with price. Attribute non-attendance to charge time and amenities (class 8) was rarely observed (1% of respondents). Additionally, these respondents were observed to have economically inconsistent behavior where they preferred higher prices and longer detour times. It is possible this class covers respondents who were inattentive to the choice tasks.

Table 7.8. Latent class non-attendance model with amenities

	Class1		Class2		Class3		Class4		Class5		Class6		Class7		Class8	
Attributes	Est.	t-stat	Est.	t-stat	Est.	t-stat	Est.	t-stat	Est.	t-stat	Est.	t-stat	Est.	t-stat	Est.	t-stat
Price ($)	-0.22	-2.08	-1.11	-3.18	-0.09	-3.67	-1.26	-10.14	-0.25	-1.42	-0.89	-5.32	-3.90	-2.02	0.30	5.08
Free Charge Indicator	0.05	0.33	0.14	0.34	0.19	2.58	1.12	3.22	1.01	3.10	--	--	0.79	1.36	-0.38	-3.04
Charge Time (10-mins)	-0.37	-4.43	-0.84	-4.78	-0.05	-1.30	-0.83	-6.12	-1.54	-3.64	-1.38	-7.26	--	--	--	--
Detour Time (10-mins)	-0.44	-2.23	-0.36	-0.92	-0.08	-2.04	-2.52	-7.47	-1.60	-3.37	-2.44	-9.27	--	--	0.65	7.43
Convenience Store Indicator	0.64	1.62	0.37	1.12	--	--	--	--	--	--	--	--	--	--	--	--
Restaurant Indicator	1.65	4.30	-0.43	-0.71	--	--	--	--	--	--	--	--	--	--	--	--
Shopping Mall Indicator	2.29	4.12	0.64	1.24	--	--	--	--	--	--	--	--	--	--	--	--
ASC Charger B	0.51	2.92	-0.13	-0.29	0.17	3.25	0.41	2.12	0.29	1.08	0.21	1.63	-0.68	-2.21	-4.09	-4.72
ASC Charger C	0.25	1.50	0.04	0.17	-0.02	-0.43	0.34	1.50	-0.09	-0.52	0.03	0.24	0.23	0.70	-8.19	-13.75
Class Membership Model																
Class Probability	13.4%		6.5%		41.7%		12.6%		7.1%		13.8%		4.0%		0.9%	
Willingness-to-Pay																
Zero-Price Effect	$0.23		$0.12		$2.26		$0.88		$4.05		--		$0.20		--	
Value of Detour Travel Time	$12.13		$1.95		$5.63		$11.97		$38.33		$16.56		$0.00		--	
Value of Charging Time	$10.07		$4.51		$3.73		$3.95		$36.97		$9.38		$0.00		--	
Model Statistics																
Log-likelihood	-9083.28															
AIC	18280.56															
BIC	18549.82															
McFadden R2	0.3231															
Number of Observations	9984															
Number of Individuals	832															
Number of Parameters	57															

7.7 Zero-Price Effect Analysis

The primary goal of this study was to present an estimate of the zero-price effect across adults in the US. Four estimates will be showcased in this section:

- Lowest BIC model of exploratory latent-class analysis
- Lowest BIC model of exploratory latent-class analysis which is economically consistent
- Zero-price effect estimate accounting for non-attendance without amenities
- Zero-price effect estimate accounting for non-attendance with amenities

7.7.1 Lowest BIC model

The model with the best BIC was the 11-class model. The mean zero-price effect for this model was $1.13. The distribution of the zero-price effect is heavily weighted at two locations: $0.00 and $2.16. The median zero-price effect was $0.27. Table 7.9 shows the nonparametric distribution of the zero-price effect for the lowest AIC model.

Table 7.9. Zero-price effect distribution (best BIC)

	Class1	Class2	Class3	Class4	Class5	Class6	Class7	Class8	Class9	Class10	Class11
Zero-Price Effect	$5.09	$2.16	$0.27	$0.22	$0.05	--	--	--	--	--	--
Class Probability	0.06	0.37	0.11	0.04	0.15	0.05	0.06	0.07	0.02	0.05	0.01
-- denotes that no zero-price effect was calculated due to economic inconsistencies											

7.7.2 Lowest BIC and economically consistent

The model with the lowest BIC which had all economically consistent classes was the five-class model. The mean zero-price effect for this model was $0.79 with a median zero-price effect of $0.75. The distribution of the zero-price effect is heavily weighted in class 1: $1.04; this is the mean and mode of the distribution. Table 7.10 shows the nonparametric distribution of the zero-price effect for the lowest BIC model.

Table 7.10. Zero-price effect distribution (economically consistent)

	Class1	Class3	Class2	Class4	Class5
Zero-Price Effect	$1.04	$0.75	$0.67	$0.48	$0.39
Class Probability	0.44	0.16	0.20	0.07	0.14

7.7.3 Non-attendance without amenities

As discussed in section 5.2, the best fitting and behaviorally sound model that excluded amenities was estimated and presented. For this model, the mean zero-price effect is $1.38 with a median zero-price effect of $0.91 (Class 3). The general shape of the distribution is decreasing from the median point in both directions. About 6% of people have no zero-price effect and 13% of respondent have an inconclusive zero-price effect (since they were price non-attendant).

Table 7.11. Zero-Price Effect Distribution (Attribute Nonattendance without Amenities)

	Class4	Class1	Class3	Class6	Class2	Class5	Class7	Class8
Zero-Price Effect	$6.91	$1.93	$0.91	$0.19	$0.13	--	--	--
Class Probability	0.06	0.44	0.12	0.04	0.15	0.06	0.12	0.01
-- denotes that no zero-price effect was calculated due to non-attendance on the free or price attribute								

7.7.4 Non-attendance with amenities

As discussed in section 5.3, the best fitting and behaviorally sound model that included amenities was estimated and presented. For this model, the mean zero-price effect is $1.39 with a median zero-price effect of $0.88 (Class 1). The general shape of the distribution is generally decreasing from the median point in both directions. About 14% of people have no zero-price effect.

Table 7.12. Zero-Price Effect Distribution (Attribute Nonattendance with Amenities)

	Class5	Class3	Class4	Class1	Class7	Class2	Class6	Class8
Zero-Price Effect	$4.05	$2.26	$0.88	$0.23	$0.20	$0.12	--	--
Class Probability	0.07	0.42	0.13	0.13	0.04	0.06	0.14	0.01
-- denotes that no zero-price effect was calculated due to non-attendance on the free or price attribute								

7.8 Discussion

7.8.1 Summary

This research explores the possibility of a free-price effect in electric vehicle charging. A charger choice stated preference survey was developed and distributed through a probability-based internet panel. Using latent class models weighted according to US population demographics, it was found that a substantial proportion of the population likely experiences a zero-price effect in regards to electric vehicle charging. The national average mean zero-price effect was estimated to range between about \$0.90 and \$1.40. There was substantial heterogeneity in this zero-price effect. Across the models that account for attribute non-attendance, 14% to 19% of the population is estimated to have no zero-price effect or an inconclusive zero-price effect. Even among the population with a zero-price effect, a plurality of the population has a zero-price effect around \$2 per charge.

7.8.2 Policy Implications

The zero-price effect is the minimum partial discount for maximizing consumer surplus. If the discounted price is greater than the zero-price effect, then each additional cent of discount confers the same benefit to the consumer. But once a discount price becomes less than the zero-price effect, although that cent of discount still confers a cent of benefit, there is a missed opportunity benefit. This opportunity benefit is the difference between the zero-price effect and the discounted price. At this point, greater consumer surplus would be obtained by fully discounting the charging event since then the consumer would gain the benefit of the zero-price effect.

For example, assume a zero-price effect of \$1 on a \$5 dollar charging event. If an entity discounts the charging event by \$2.50, the consumer gains \$2.50 in benefit. If an entity fully discounts a charging event, the consumer gains $5.00 of benefit from the discount and an additional $1.00 of benefit from the zero-price effect. If an entity discounts that charging event by $4.50, the consumer gains $4.50 of discount benefit. But the consumer is missing out on that $1.00 from the zero-price effect which would only cost $0.50 more to obtain – an 100% return on that additional cost. Understanding this breakpoint, the zero-price effect size, has implications for cost-benefit analysis of charging pricing policies. This research provides a basis for more accurately performing such cost-benefit analyses of national and local infrastructure pricing policies (Maness and Lin, 2019). This study supports an effect size of that order of magnitude.

Free charging may cause concerns with inefficient charging behavior. Motoaki and Shirk (2017) found that flat-rate fees incentivized charging station users to increase charger occupancy time. Particularly, with fast chargers, this creates additional inefficiency beyond capacity constraints since DC fast chargers have diminishing charge rates over time.

7.8.3 Sample, Modeling, and Experiment Limitations

The sample collected exhibits some limitations on the practical implications of the zero-price effect estimate. The study assumes that all American adults could charge EVs which is likely not the case, at least not in the medium term. This suggests that studying the demographics of the different classes could be helpful. By understanding the likelihood of each class to purchase an EV, a zero-price effect estimate weighted by EV purchase likelihood could be obtained. This measure would likely be more relevant for understanding the benefits of such a policy when EV penetration rates are low. As the penetration rate rises, this “self-selected zero-price effect” would converge towards the general population zero-price effect. Additionally, there may be some timing effects associated with the survey collection time. Occurring during an economic downturn, the temporal stability of the measured zero-price effect is uncertain.

Modeling limitations involved the choice model structures. Although a mixed logit form could potentially be more parsimonious, a latent class approach is more non-parametric accounting for prior uncertainty about the distribution of a free-price effect. Additionally, the mixed logit formulation has difficulties dealing with attribute non-attendance. Attribute non-attendance leads to distributions more strongly tilted towards low zero-price effect values. Creating a "zero-inflated" form with continuous distributions is difficult to obtain. Future work could attempt to estimate latent class mixed logit formulations which could reduce the number of attendant classes.

This study did not include sociodemographics in the model specifications. This was chosen due to the study's focus on policy and measurement rather than behavioral explanation. For early analysis of free charging policies and pricing structures, the cost-effectiveness of the policy is a greater focus than the exact structure of the policy, so a looser mean-focus and distribution-focused approach brings greater value. Future work can look at the demographic characteristics of the preference classes. This allows for tailoring the policies and business plan around groups that may see greater benefit from the program or groups most likely to use chargers from companies that offer free charging in some form. Understanding the demographic also may have implication for equity analysis as disparities were observed in the value of time between different classes. Since value of time is seen as partially determined by wage rate, understanding the characteristics of these group may be important for creating more equitable public charging infrastructure.

The analysis has been written such that attribute non-attendance is a population-level trait. This is rather difficult to disentangle true attribute non-attendance from experiment-level attribute non-attendance in a stated-preference study. As hypothetical bias decreases, the observed non-attendance should move closely match true non-attendance.

The limitations of the choice experiment relate to the choice context, hypothetical bias, and the presence of other relevant attributes. The choice context present is low pressure and with no time constraint. This helps to explains the low values of time observed as respondents did not have to balance the prospects of being late or reduced time at a follow up activity with price. Under a choice context with time constraint, a different zero-price effect distribution would be expected. Future work will look to change the context of the choice scenarios to create a fuller understanding of the impact of free prices on charging behavior. The prospect of hypothetical bias is always present in stated choice experiments and this analysis is limited by individuals' limited experience with electric vehicle charging. The value of times obtained aid in understanding possible hypothetical bias and individuals did exhibit expected time valuations (lower values-of-time due to the lack of time constraints/leisure activity context and charging time exhibiting lower valuation than travel time). Additionally, other attributes could be relevant to the choice of charger location. Charger type was excluded due to expected unfamiliarity with charging technology. Since charger type mostly impacts charging rate, charging time was included rather than type. State of charge after the charging is rather important in a charging event. The researchers decided that simplification of the task was more important, as different respondents would interpret the importance/needs for charge differently and this would be more difficult to clearly differentiate in the modeling task.

Chapter 8

A Multi-City Investigation of the Effect of Holidays on Bikeshare System Ridership

Trang Loung, Michael Maness

8.1. Introduction

Bikeshare provides important first mile last mile, commuting, circulation, and sightseeing options in many cities. In 2018, there are 67 U.S. cities with bikeshare systems consisting of 45.5 million bikeshare trips—80% of these trips are station based and 20% are floating. In the U.S. in 2018 there were a total of 57,000 station-based bikes available (NACTO, 2019). Bikeshare can also be healthy and convenient for users. Bikeshare ridership includes both leisure and work-related trips. It is important and useful to gain insights into how and why travelers choose bikeshare—this is valuable for planning and for network design and operations (e.g. for pre- and re-positioning of the bikes). There is substantial research on characteristics of bikeshare commute trips, but there is limited research on leisure-based trips during weekends, holidays, and special events. During a typical year there are 10 federal holidays, accounting for 3% of all calendar days. For example, in 2018 in Washington, D.C.'s Capital Bikeshare system there are approximately 68 thousand trips taken during federal holidays that account for about 2% of all trips. These types of days are interesting for analysis because they are when users characteristically change their daily routines and alter their typical bikeshare behavior.

Past investigations of the impacts of holidays on bikeshare usage show disparate results due to differences in holiday definitions, user composition across systems, and modeling specifications. Past studies found a range of directional effects of the combined weekend and holiday indicator variables including increased (El-Assi et al., 2015, Corcoran et al., 2014) and decreased ridership (Mattson and Godavarthy, 2017; Sun et al., 2018; Gebhart and Noland, 2014). Although research has found that subscribing members and non-members exhibit different bikeshare usage behavior (Sun et al., 2018; Wang and Lindsey, 2019; Zhou, 2015), no studies have analyzed holiday impact by user type. This study identifies the following four gaps in understanding bikeshare usage behavior during holidays:

1. *Holidays are defined differently across studies.* Different scales of holidays, including federal holidays (Gebhart and Noland, 2014), public holidays (Corcoran et al., 2014; Kim, 2018), school holidays (Corcoran et al., 2014; Mattson and Godavarthy, 2017; Kim 2018), and local holidays (Kaltenbrunner et al., 2010) give different effects on bikeshare usage resulting in incomparable results.
2. *Existing studies neglect to account for user's heterogeneity in system-level ridership patterns on holidays.* Member and non-members have different usage profiles, which likely leads to different holiday ridership. But, looking at just total ridership obscures this relationship.
3. *Are there differences between holidays and weekend ridership patterns?* Federal holidays and weekends share some similar effects due to them both being non-workdays. It was found that no past studies have clearly answered this question.
4. *Are there differences in ridership patterns between individual holidays?* Holiday traditions have differing social and recreational activity space and preference profiles. But no previous research on daily bikeshare ridership has analyzed the effects of holidays individually.

To better understand these four gaps, this study analyzes daily ridership patterns across five U.S. bikeshare systems in Washington D.C., Chicago, Boston, Minneapolis, and Los Angeles. To address the first gap, this study defines holidays as federal and non-work holidays across five systems. By conducting a multi-city analysis with clearly defined and consistent holidays, general ridership patterns on the same types of holidays can be drawn from the results. For the second gap, to address all user types, the effects of holidays on member and non-member ridership are considered separately rather than addressing only total ridership. This could explain the impact of who is using the bikeshare system on holidays. To understand gap three, this study found support that ridership behavior varies between holidays and weekends. For gap four, this study found differences between different holiday ridership patterns. For example, Thanksgiving and Christmas result in less total ridership than Labor Day and Memorial Day.

8.2 Literature Review

In past investigations of bikeshare use in the U.S. and abroad, there have been some disparities in holiday bikeshare use across different systems. Holidays occur on weekdays and weekends, but the approach as to how to define them in the econometric model differs amongst studies.

Some prior bikeshare studies have distinguished between weekend and weekday variables in their econometric models to account for calendar days. In Montreal, for example, Faghih-Imani et al. (2014) found that weekends had a negative effect on arrival and departure rates, along with Zhou (2015) who found that weekend ridership was much lower than weekday ridership in Chicago. Conversely, Hyland et al. (2018) found that both weekends and weekdays had positive effects on total trip counts in Chicago. In Washington D.C., Younes et al. (2020) incorporated separate day of the week fixed effects. Relative to Sunday, their models revealed that Monday had fewer trips; Tuesday through Thursday had similar trips; and Friday and Saturday had more trips. Further, a study discovered that Friday evening trip counts were like Saturday counts, which indicates that Friday evenings are similar to weekend days, and therefore Fridays can exhibit ridership patterns like a weekday and a weekend day in different cities (Faghih-Imani et al., 2014).

Whether the user is a subscribing member or non-member of the bikeshare system can also impact the ridership results on different calendar days. Typically, bikeshare members have annual or monthly passes while non-members have single day passes. Wang et al. (2019) found that annual members took fewer bikeshare trips on weekends in St. Paul, Minnesota. Likewise, Sun et al. (2018) found that on weekends in Seattle, member ridership declined while non-member ridership increased. Zhou (2015) observed that member users were more commute-oriented and non-member users were more recreational oriented. This could explain why past analyses found weekend ridership to be lower for members who were not working/commuting and higher for casual non-members since special events and leisure may attract locals and tourists to bikeshare on weekends. Analyzing trips made by members and non-members separately can provide more information in a model on the types of riders utilizing the system on different calendar days.

A limitation in the above studies is that they did not include holiday indicator variables in their econometric models. For example, missing holiday variables could result in omitted variable bias and unobserved heterogeneity issues, as not incorporating any form of a holiday variable could make it unclear whether the magnitude and direction of the weekend and weekday parameter estimates are caused by the weekend and weekday itself or caused by the unobserved effects of the holidays.

There is limited consensus, however, on how to incorporate holidays in the analysis of calendar day effects. This occurs across two factors:

- What is defined as a holiday?
- How are holidays incorporated into the model?

Across the studies reviewed, the definition of a holiday varies. Cultural differences can play a role in the definition of holidays. Across studies internationally, Kim (2018) in Daejeon, South Korea, and Corcoran et al. (2014) in Brisbane, Australia looked at public and school holidays, Kaltenbrunner et al. (2010) in Barcelona included local holidays, El-Assi et al. (2015) in Toronto statutory holidays and Zhang et al. (2017) in Zhongshan, China undefined holidays. Undefined holidays are when studies use a holiday variable in the analysis, but do not specify what holidays are being analyzed. Across studies in the U.S. alone, one used school holidays in Fargo, North Dakota (Mattson and Godavarthy, 2017), one used national holidays (Sun et al., 2018), one used federal holidays in Washington D.C. (Gebhart and Noland, 2014), and two used undefined holidays (Wang and Lindsey, 2019; Younes et al., 2020).

Past studies found a range of directional effects of the combined weekend and holiday indicator variables. To account for unobserved heterogeneity due to holidays, a common approach is to combine holidays with weekends into a single fixed effect. Across models with weekend-holiday fixed effects, decreased bikeshare ridership was observed for school holidays (Mattson and Godvarthy, 2017), national holidays (Sun et al., 2018), and federal holidays (Gebhart and Noland, 2014). In one study, increased ridership occurred on statutory holidays (El-Assi et al., 2015). Corcoran et al. (2014) also found that weather (ambient temperature) was associated with increased ridership on public holidays and weekends as compared to weekdays. Using shared variables in a model indicates that the modeler assumes the effects of weekends and holidays are mutual.

To distinguish between weekends and holidays, some studies grouped holidays into a binary variable separate from the weekend variable. Kim (2018) found that public holidays had a negative effect on system-hour-daily ridership, but Corcoran et al. (2014) found that public holidays wielded a positive effect on total system ridership. A commonality in the literature is that school holidays exerted no significance on ridership; this is often attributed to age restrictions on bikeshare membership (Corcoran et al., 2014; Kim, 2018). Younes et al. (2020*)* and Zhang et al. (2017) included a holiday indicator variable, but they did not specify in the analysis which holidays were included. Both found that holidays induced no effect on both member and non-member trips.

The impact on ridership appears to change based on how holidays are classified and grouped. While separating the holiday and weekend indicator variables may strengthen the accuracy of the results, failure to ungroup the holidays based on the holiday (Christmas, Veterans Day, Labor Day, etc.) could mask the actual effects of each individual day. There is a lack of research regarding ungrouping holidays, but it may not always be feasible in some analyses due to the time frame of the study. For example, Zhang et al. (2017) had a grouped holiday variable in their model, but the study only occurred over a four-month period. None of the studies reviewed had more than two years of data, which makes the inference of separate holiday effects difficult since there is limited or no replication of individual holidays in the analysis.

As these studies evaluated the factors that influenced bikeshare ridership that change based on weekdays, weekends, and holidays, very few studies incorporate special events in their model local to the bikeshare system analyzed. In Washington D.C., Younes et al. (2020) found that the annual Cherry Blossom Festival had a substantial positive impact on all trips. Many users utilized the bikeshare system to participate in the activities at the Cherry Blossom Festival. Kaltenbrunner et al. (2010) found that on a local holiday, the Feast of Sant Joan, Barcelona's bikeshare system exhibited ridership patterns more like a typical Sunday although the holiday occurred on a Tuesday that year. Therefore, different holidays and special events may cause changes in bikeshare decision making. Table 8.1 summarizes the holiday and special event variables from the literature review in a table with the corresponding directional effects.

Table 8.1 Literature review table – holidays and special events

Author	Model	Location	Time Period	Dependent Variable	Independent Variable (relevant to this study)	Directional Effect
Corcoran et al.	Poisson Regression Model	Brisbane, Australia	Nov. 2010 to Jul. 2012	Total Trips (log)	Public Holiday, School Holiday, Weekend	No effect on public or school holidays
El-Assi et al.	Weekday and/Weekend Distributed Lag Model; Multi-Level Mixed-Effects Regression Model	Toronto, Canada	2013	Natural Logarithm of the Trip Counts	Statutory Holidays and Weekends, Weekday	Increased ridership occurred on statutory holidays
Faghih-Imani et al.	Multilevel Linear Mixed Model	Montreal, Canada	Apr. to Aug. 2012	Arrivals or Departures at a Station	Weekends, Weekdays, Friday, and Saturday Nights	Bicycle usage decreased during weekends in Montreal, but increased on Friday and Saturday Nights
Gebhart and Noland	Negative Binomial Model	Washington D.C.	Oct. 1, 2010, to Dec. 31, 2011	Trips per Hour	Weekends and Federal Holidays, Month Year	Usage on weekends and holidays is not significantly different than on weekdays
Hyland et al.	Multilevel Mixed-Effect Regression Model-Hybrid Cluster-Regression Approach	Chicago, Illinois	2016	Log (Trip Count)	Dec. 2016, Weekdays in Month, Weekend Days in Month, Month	Weekends and weekdays had positive effects on total trip counts in Chicago
Mattson and Godavarthy	Regression Model	Fargo, North Dakota	2015 and 2016	Ridership (log)	Weekend or Holiday, Year	Negative effect for weekend or holidays
Kaltenbrunner et al.	Auto-Regressive Moving Average (ARMA) Model	Barcelona, Spain	May 15, 2008, to Jul. 3, 2008	Number of Bicycles Available	Time of Day, Day of Week, Local Holiday	Feast of Sant Joan has bicycle patterns more like a typical Sunday
Kim	Negative Binomial Regression Model	Daejeon, South Korea	2015	Number of Bicycle Rentals	Weekend, Public Holidays, School Holidays	Negative effect on public holidays; School holidays do not have a significant impact on bikeshare usage

Sun et al.	Generalized Additive Mixed Model (GAMM)	Seattle, Washington	Oct. 15, 2014, to Aug. 31, 2016	Total Counts of Pickups and Returns	Workday, National Holiday and Weekend	Non-working days are negatively correlated with member pickups but positively correlated with short-term pass holder pickups
Wang et al.	Linear Mixed-Effects Models and Multinomial Logistic Models	Minneapolis-St. Paul, Minnesota	2017	Average Daily Trip Frequency, Average Daily Trip Frequency on Weekends, Average Daily Trip Frequency on Weekdays	Weekends and Holidays	Annual members took fewer bikeshare trips on weekends in St. Paul, Minnesota
Zhou	Hierarchical Clustering Method	Chicago, Illinois	Jul. to Dec. 2013; Jul. to Dec. 2014	Total Over-Demand Numbers for Docks and Bikes	Time of Day, Day of Week, Subscribers vs. Customers	Weekend usage was much less than on weekdays
Younes et al.	Negative Binomial Regression, Log-Linear OLS Regression	Washington, D.C.	Dec. 2018 to Jun. 2019	Number of Trips per Hour, Medium Duration of Trips per Hour	Cherry Blossom Festival, Government Shutdown, Day of Week	Holidays had a positive effect on casual bikeshare trips and fewer member trips; The Cherry Blossom Festival had a significant positive impact on all types of bike share trip activities
Zhang et al.	Multiple Linear Regression Models	Zhongshan, China	Feb. to Jun. 2014	Ln[D/S] of Weekdays, Weekends, and Holidays	Weekdays, Weekends and Holidays	Negative impact on daily D/S at stations on weekdays and no influence on weekends and holidays

8.3 Hypothesis Testing

This study proposes to answer the research question: Why does existing literature have disparities amongst different bikeshare system-level ridership results on holidays? The following hypotheses will suggest possible explanations as to why these disparities occur between holidays across different systems and why there is no clear effects on bikeshare usage. The hypotheses are classified into three ridership types for holidays: Member (HM), Non-member (HN) and Total (HT).

8.3.1 Grouped non-work holidays effects

Since members can be more likely to use bikeshare for commuting, it is expected that member users would use bikeshare less during their non-workdays, such as federal holidays, and weekends.

HM1. *Federal holidays exhibit lower member ridership compared to non-holidays.*

HM2. *Federal holidays have similar member ridership with weekend days.*

In contrast with member users, non-member users are assumed to generally use bikeshare for mostly leisure-based trips. Since non-workdays increase the time available for leisure, non-member trips are assumed to increase during non-work holidays.

HN1. *Federal holidays exhibit greater non-member ridership compared to non-holidays.*

HN2. *Federal holidays have greater non-member ridership than a weekend day.*

Similarly, variations in holiday observance likely will impact ridership levels on non-work holidays. It is expected that as holiday observance increases member-level ridership would decrease because of fewer commuting trips. Simultaneously, non-member ridership would increase because of increased leisure.

HM3. *Non-work holidays with higher levels of observance have fewer member trips than holidays with lower levels of observance.*

HN3. *Non-work holidays with higher levels of observance have more non-member trips than holidays with lower levels of observance.*

8.3.2 Holidays-specific effects

It is expected that different federal holidays could have dissimilar effects across systems. Hence, this hypothesis is built for testing whether each individual federal holiday has a different effect on total ridership.

HT1. *Ridership levels vary across specific federal holidays.*

When federal holidays fall on a weekend day, most often the holiday is observed on the Friday before or the Monday after. Therefore, users may have a three-day weekend and therefore may be more likely to engage in outdoor or leisure activities and utilize bikeshare.

HT2. *Federal holiday on the weekend induces higher ridership compared to the same federal holiday on a weekday.*

8.3.3 Cherry blossom festival effects

The Cherry Blossom Festival in Washington, D.C. is in an area away from public transit, across a large outdoor space and attracts international and regional tourism. This could cause non-member ridership to rise due to the transient nature of the riders. It assumes that non-members are typically tourists or infrequent users.

HN4. *The Cherry Blossom Festival in Washington D.C. attracts significant tourism which induces higher ridership for non-members.*

8.4 Data Description

Five station-based bikeshare systems were selected as case studies in this analysis: Capital Bikeshare (Washington, D.C.), Bluebike (Boston), Divvy (Chicago), Nice Ride (Minneapolis) and Metro Bikeshare (Los Angeles). In 2018, these systems account for approximately 13,800 station-based bikes (nearly 24% of the total in the U.S.) and 9.5 million total trips (nearly 26% of total in U.S.) (NACTO, 2019). To maintain consistency in the estimation results, the first three systems were selected because of their similarities in system size and weather characteristics. Looking at Nice Ride and Metro Bikeshare with unique characteristics may help explain the analysis as it can see whether system disparities affect holiday ridership. Nice Ride is somewhat unique as it closes its operations in the winter, so it does not account for the winter holidays in Minneapolis. Also, member and non-member trips have an almost even distribution in ridership, while members dominate bikeshare culture in the other studied systems (Table 8.2) – thus increasing variation in the membership ratio distribution across the systems studied. This increases the robustness of the study's findings in helping to differentiating total ridership effect versus membership effects. To control for possible seasonal variations in the outdoor leisure activity space, Metro Bikeshare in Los Angeles was included due to its subtropical climate to increase the robustness of the hypothesis testing for the holiday-specific effects.

Table 8.2 Summary of bikeshare systems

City, State	System	Year	Total Trips	Member Trips	Non-Member Trips
Washington, D.C.	Capital Bikeshare	2012-2019	24,716,073	19,620,321	5,095,752
Chicago, IL	Divvy	2015-2019	18,028,625	13,845,616	4,183,009
Boston, MA	Bluebike	2015-2019	7,883,904	6,211,713	1,672,191
Minneapolis, MN	Nice Ride	2012-2019	3,150,889	1,677,022	1,473,867
Los Angeles, CA	Metro Bikeshare	2017-2019	783,733	526,569	257,164

Daily ridership data was compiled for each system using disaggregate trip data from each system's respective websites (Capital Bikeshare, 2020; Bluebikes, 2020; Nice Ride, 2020; Metro Bikeshare, 2020; Divvy 2020). The individual trip information from each system included: start and end time, origin and destination station, and user type (member or non-member). Depending on system inception dates, the systems were analyzed over different time periods. Additionally, the data at the beginning of each system will be excluded in this study to avoid instability of the systems during their growing phases. Furthermore, the trips whose user types were unknown were left out of the data. Table 8.2 presents the summary of all five systems in this chapter. Data was removed in each system near the beginning of the system's existence. This was to account for how system demand is low and highly variable upon system introduction. The authors did not think that the system ridership demand generation process was similar enough to include in a single model. The removed data is listed as follows:

- Capital Bikeshare: Removal of 2011 data.
- Divvy: Removal of 2013 and 2014 data.
- Bluebike: Excluded data from the first four months of 2015.
- Nice Ride: Excluded data from 2010 and 2011.

- Metro Bikeshare: Excluded data from the first four months of operation in 2017.
- Excluded observations with unknown user types.

The individual trip data of each system later was aggregated into daily trip count data with the start of the day occurring at 4:00 am of the calendar day and ending at 3:59 am the following day. This daily bikeshare trip data was also categorized into daily member and non-member trip counts. The purpose of changing the start of day from midnight to 4:00 am is to capture late night bikeshare trips in the systems. Finally, the weather data was obtained from the National Oceanic and Atmospheric Administration (2020) and Weather Underground websites (2020). Since the weather at an airport is similar to the weather of the city, weather data from each city's closest international airport weather station was collected for this study. The weather data used in this analysis included daily maximum temperature, average wind speed, maximum dewpoint, snowfall, snow depth and precipitation in inches.

In order to help visualize the variation of bikeshare ridership throughout a calendar year, the calendar heatmap in Figure 1 shows the difference between the actual number of non-member bikeshare trips on a specific day and the mean non-member trips of 14 days around that day in Washington D.C. The mean trips were calculated by using a centered moving average with a 14-day time window. By looking at this calendar, one would see whether the actual non-member bikeshare trips are above or below the average non-member trips of 14 days around that day. It appears that 14 days is a reasonable span length as it can help standardize the effects of holidays and unusual weather conditions. The non-member ridership was generally higher on weekends than weekdays. The non-member ridership remained similar or decreased on weekdays. The heatmap generally shows that federal holidays on weekend had higher non-member ridership than the same federal holidays on weekday such as New Year's Day, Christmas Day, Independence Day, and Veterans Day.

In Figure 1, the Cherry Blossom Festival on Saturday generally show high levels of non-member ridership. For Washington D.C., approximately from the third week of March to the second week in April, Capital Bikeshare participates in the National Cherry Blossom Festival by offering the corral service. This service encourages event attendees by guaranteeing a space to dock. In 2018, the average daily trips during the Cherry Blossom Festival was 11.5% higher than average annual daily trips. Since the Cherry Blossom Festival is extended over a long period of time and is in the outdoor activity space, it is expected that this event could significantly affect the bikeshare usage in Washington D.C.. This park-based events may increase ridership due to users possibly associating parks with bicycling. The primary area for the Cherry Blossom Festival (around the Tidal Basin) is serviced by only two metro stations with significant egress distances (over 0.5 miles). Additionally, the festival has three trail loops for viewing with distances varying from 2.1 to 4.1 miles. This makes bikeshare a significantly viable mode of transportation for this event.

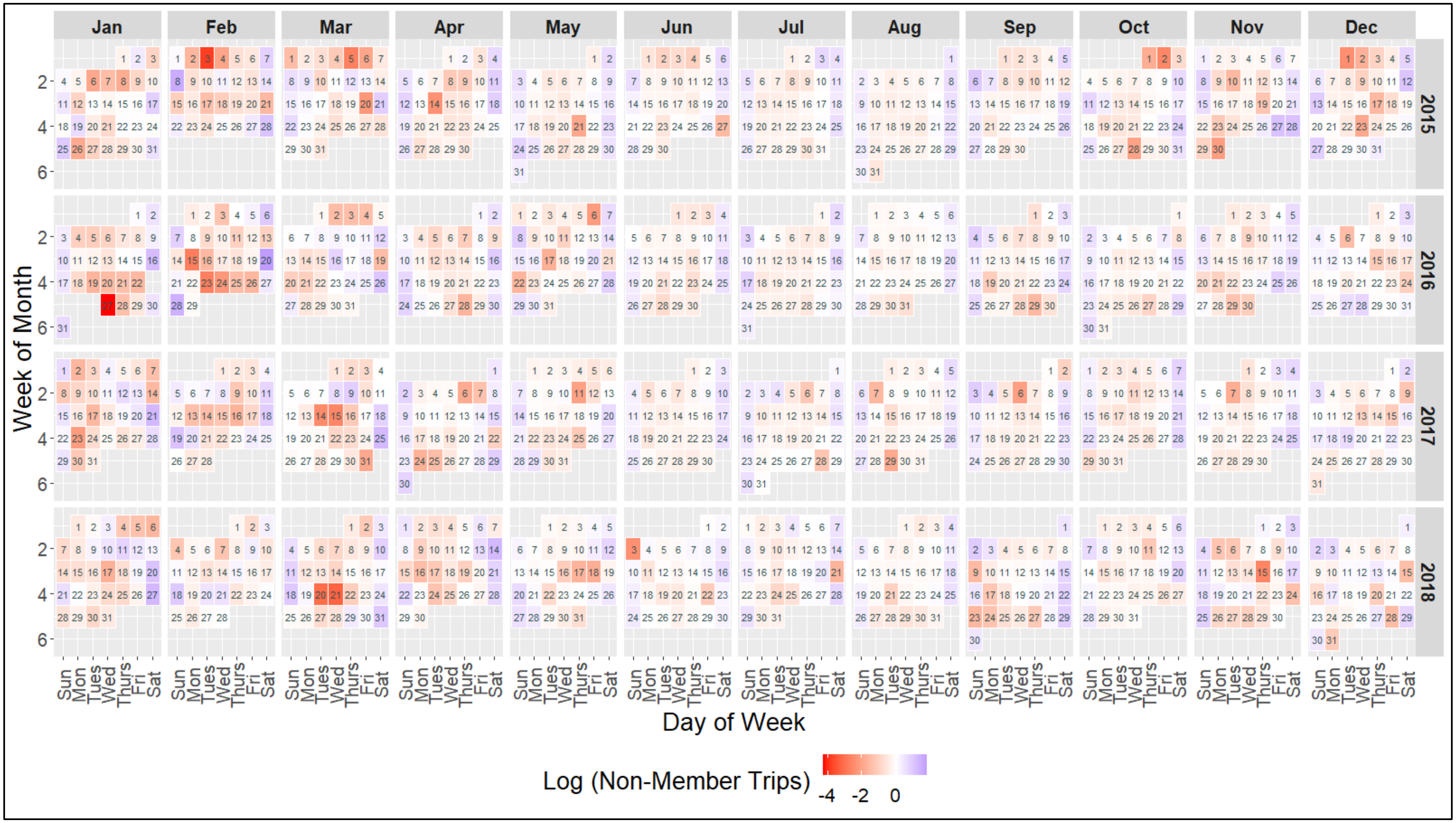


Figure 8.1 Capital Bikeshare calendar heatmap for non-member ridership – log (ratio of daily non-member ridership to 14 day moving average non-member ridership)

8.5 Methodology

This study consists of a multi-city investigation of the effect of holidays on system-level bikeshare ridership in five U.S. cities using a log-linear regression with robust standard errors model. A log-linear regression model resulted in more normally distributed error terms because the system grows over time and has high seasonal variation. Also, this allows the coefficients to be compared across systems because the estimated coefficients show the percent increase or decrease for every one-unit increase in the independent variable. To account for additional heteroscedasticity, the Eicher-Huber-White robust standard errors are reported. The general model specification formula is presented in (8.1):

$$log(Trips) = \beta_0 + \beta_{\mathrm{H}}(Holiday) + \beta_{\mathrm{F}}(CherryBlossom) + \beta_{\mathrm{T}}(Temporal) + \beta_{\mathrm{W}}(Weather) \quad (8.1)$$

Where β_0, β_{H}, β_{F}, $\beta_{\mathrm{T}}, and\ \beta_{\mathrm{W}}$ are estimated coefficients regarding the constant, holidays, the Cherry Blossom Festival, temporal fixed effects, and weather, respectively.

Yearly fixed effects were included to account for changes in the supply of bikes and stations along with the familiarity and demand of the system. Monthly fixed effects were used to account for seasonal variation. The reference variables were the first year of the study period for yearly fixed effects, and July for monthly fixed effects. Variations in ridership over the week were accounted for with daily fixed effects with Wednesday serving as the reference day.

For testing hypotheses, five different model specifications were estimated for total, member, and non-member ridership in each system (with the corresponding hypotheses tested in parentheses).

- Federal holidays grouped plus weekend variable (HM1, HM2, HN1, HN2)

- Federal holidays grouped and Saturday and Sunday separately (HT1 - restricted model)
- Federal holiday-specific Effects (HT1 - unrestricted model)
- Non-work holidays grouped by Business Closure Rate (HM3, HN3)
- Holiday-specific Effects (HT2, HN4)

To test the significance of the hypotheses, either an unpaired one-sided *t*-test, unpaired two-sided *t*-test of the equality of coefficients, or a likelihood ratio test was conducted using the coefficients of the inferential models.

The levels of observance for non-work holidays were obtained from the Society for Human Resource Management which forecasted the percentage of U.S. businesses closed by holiday in 2015 (SHRM, 2014). Based on these percentages, non-work holidays were grouped ordinally by observance into three categories from high to low observance for testing hypotheses HN3 and HM3:

- High observance non-work holidays: New Year's Day (95%), Memorial Day (94%), Labor Day (95%), Thanksgiving (97%) and Christmas (97%).
- Medium observance non-work holidays: Independence Day (76%), Black Friday (76%), and Christmas Eve (47%)
- Low observance non-work holidays: Birthday of Martin Luther King (37% of businesses closed), Washington's Birthday (35%), Good Friday (28%), Columbus Day (16%), Veterans Day (20%), and New Year's Eve (23%)

For the holiday-specific effects models, all the holidays listed above were included individually. Fixed effects were added for each holiday on its corresponding day of the year the holiday falls on to account for the ridership impact of that holiday. For federal holidays which can occur on a weekend day, an additional fixed effect was added for its official day of federal observance. Additionally, when a federal holiday fell on a weekend day, a fixed effect was added to account for differences in ridership compared to when the respective holiday occurs during a weekday.

Weather conditions were controlled in the model structure with temperature and precipitation modeled non-linearly. For example, bikeshare ridership is expected to increase with temperature (Eren and Uz, 2020), however ridership may drop above a threshold due to excess heat.

8.6 Analysis – Hypothesis Tests

All hypotheses in this study are tested at a 95% confidence interval across all five systems. The results of the hypothesis testing are summarized in Table 8.3.

Table 8.3 Hypothesis test results

Hypothesis	**Test**	**Alternative Hypothesis**	**Capital Bikeshare**	**Divvy**	**Bluebike**	**Nice Ride**	**Metro Bikeshare**
Non-work holidays effects							
HM1	One-sided *t*-test	Grouped federal holidays < 0	-11.699*	-8.989*	-8.537*	-5.505*	-6.880*
HM2	Two-sided *t*-test of the equality of two coefficients	Grouped federal holidays ≠ Weekend	-4.922*	-1.228	-1.463	0.181	-0.478
HN1	One-sided *t*-test	Grouped federal holidays > 0	14.717*	7.121*	0.105	6.126*	4.888*
HN2	One-sided *t*-test of the equality of	Grouped federal	-3.26	-3.998	-2.332	-2.141	-0.925

	two coefficients	holidays > Weekend					
HM3	One-sided *t*-test of the equality of two coefficients	High Non-work < Medium non-work	-3.331*	-2.653*	-0.744	-1.893*	-2.019*
		Medium non-work < Low non-work	-7.128*	-6.853*	-5.210*	-3.354*	-2.798*
HN3	One-sided *t*-test of the equality of two coefficients	High Non-work > Medium non-work	0.532	1.923*	0.046	0.297	2.086*
		Medium non-work > Low non-work	-0.167	0.253	-1.875	1.801*	-0.309
Holiday-specific effects							
HT1	Likelihood ratio test	Separate federal holidays ≠ Grouped federal holidays	χ^2 = 288.6* DF = 13	χ^2 = 229.9* DF = 13	χ^2 = 93.9* DF = 13	χ^2 = 50.2* DF = 7	χ^2 = 17.4 DF = 10
HT2	One-sided *t*-test	New Year's Day on weekend > New Year's Day on weekday	6.717*	8.437*	3.945*	-	-
		Christmas Day on weekend > Christmas Day on weekday	12.082*	5.274*	6.157*	-	-
		Independence Day on weekend > Independence Day on weekday	2.232*	4.482*	3.400*	-1.812	-
		Veterans Day on weekend > Veterans Day on weekday	0.765	-0.169	0.171	-4.160	0.375
Cherry Blossom Festival effects							
HN4	One-sided *t*-test	Cherry Blossom Saturday > 0	3.551*	-	-	-	-
		Cherry Blossom weekday > 0	3.681*	-	-	-	-
* indicates a *p*-value <= 0.05 - denotes that a test was not performed for the corresponding test and system							

8.6.1 Grouped non-work holidays effects

Hypothesis HM1

The results of the one-sided *t*-tests supports this hypothesis that federal holidays induce lower member ridership versus a non-holiday across all five systems.

Hypothesis HM2

The results of two-sided *t*-test mostly supports the hypothesis that federal holidays and weekends have similar member ridership in four out of five systems. Capital Bikeshare experienced significantly lower member-level ridership on federal holidays compared to weekends.

Hypothesis HN1

The one-sided *t*-test results supports this hypothesis that federal holidays have higher ridership for non-members in all cities except Boston. Therefore, it could be roughly concluded the effects of federal holidays and non-holiday days are equivalent in Boston.

Hypothesis HN2

The results of the one-sided *t*-test does not support this hypothesis. It was found that federal holidays induce lower non-member ridership than a weekend day, which is opposite to this hypothesis. Therefore, the results from this hypothesis are not statistically significant with a p-value greater than 0.05 shown in Table 8.3.

Hypothesis HM3

The results of the one-sided *t*-test supports the hypothesis that the number of member trips decreases proportionally to the levels of observance for non-work holidays in all systems except in Boston. This means there are no statistically significant differences between the effects of different levels of observance non-works holidays on member ridership in Boston.

Hypothesis HN3

After conducting a one-sided *t*-test, the results do not support the hypothesis that the number of non-member trips increases proportionally to levels of observance for non-work holidays in all systems. This means there are no statistically significant differences between the effects of different levels of observance non-works holidays on non-member ridership in all systems.

8.6.2 Holidays-specific effects

Hypothesis HT1

The results generally support this hypothesis at 95% confidence level in all systems except Metro Bikeshare (p-value = 0.056); the effects of individual federal holidays on total bikeshare ridership in each system are different.

Hypothesis HT2

This hypothesis tests whether federal holidays on a weekend induce higher ridership compared to the same federal holiday on a weekday. To test this hypothesis, a one-sided *t*-test was conducted based on the results of the econometric model displayed in Table 8.4. New Year's Day, Christmas Day, Independence Day, and Veterans Day can occasionally fall on a weekend. After conducting a one-sided *t*-test, the hypotheses are generally supported on New Year's Day, Christmas Day, and Independence Day.

8.6.3 Cherry blossom festival

Hypothesis HN4

After conducting a one-sided *t*-test, the results support the hypothesis that the Cherry Blossom Festival in Washington, D.C. induces higher non-member ridership on Saturdays and on weekdays.

8.7 Analysis – Individual Holiday Effects Models

In this section, results from the holiday-specific effects models are presented with system-level ridership level aggregated by total, member, and non-member ridership. Fifteen models are presented in Tables 8.4-8.6 covering the five bikeshare system and three ridership aggregation levels.

8.7.1 Total trips

The baseline models of total bikeshare trips in all five systems were estimated with indicator variables for individual holidays (including observed holidays and holidays on weekends). All other factors such as weather characteristics, time characteristics whose effects on bikeshare demand could be substantial were also included in the model estimations. The purpose of estimating these baseline models is to capture the general effects of individual holidays on bikeshare ridership across five systems. The statistically insignificant variables are kept in the models when relevant to the tested hypotheses.

Regarding effects of federal holidays, it was found that there was a negative relationship between total ridership and federal holidays. The total ridership was found to be significantly lower than other federal holidays on New Year’s Day, Christmas Day, and Thanksgiving Day in all systems.

For the weekday effects, it was found that all five systems had lower ridership on Mondays than Wednesdays. For other weekdays, the results show that Thursdays had positive effects on bikeshare trips for Metro Bikeshare, and Fridays had positive effects on ridership for Metro Bikeshare, Nice Ride, and Capital Bikeshare. It was found that Thursdays and Fridays had no effects on the other systems. The positive effects of Fridays could be due to evening trips.

In terms of the weekend, the results show that both Saturdays and Sundays had negative effects on total ridership for all systems except Nice Ride. It was found that Saturday induce higher total ridership in Nice Ride. Due to the equal distribution between member and non-member, the leisure trips which were made by non-member users on Saturday could account for this finding.

Concerning potential seasonal effects, it was found that all systems except Capital Bikeshare exhibited common patterns. The results show that total ridership was lower in winter and spring, higher in summer, and then reduced again in fall. On the other hand, in Capital Bikeshare the results indicate that there is little variation in the total ridership across all four seasons.

Regarding weather effects on bikeshare ridership, the total bikeshare trips decrease as average wind speed increases for all systems except Metro Bikeshare. It was also found that precipitation from rain (no snow) had negative effects on bikeshare trips in all cities. Similarly, the results show that as snowfall and snow depth increases, total ridership decreases in the systems that experience snow. For temperature effects, total ridership increased quadratically for these systems as the maximum temperature increased. But, once the maximum temperature reached approximately 85 degrees F, ridership dropped in a cubic fashion. Finally, the results show that maximum dewpoint has negative effects on bikeshare trips in all systems.

Table 8.4 Log-linear regression model results of total bikeshare trips in five systems with robust standard errors

Variable Description	**Capital Bikeshare Washington D.C.**		**Divvy Chicago**		**Bluebike Boston**		**Nice Ride Minneapolis**		**Metro Bikeshare Los Angeles**	
	Mean	**t-stat**	**Mean**	**t-stat**	**Mean**	**t-stat**	**Mean**	**t-stat**	**Mean**	**t-stat**
Constant	6.96	23.10	7.30	34.13	6.65	21.09	4.61	5.30	1.22	0.42
New Year's Day (actual days, including weekend)	-0.90	-9.74	-1.25	-25.02	-0.86	-3.66	-	-	-0.22	-1.42
New Year's Day Observed (only observed days, Friday, or Monday)	-0.84	-6.01	-0.64	-17.10	-0.20	-3.76	-	-	-	-
New Year's Day*Weekend	0.33	3.60	0.51	8.48	0.81	3.44	-	-	-	-
New Year's Eve	-0.25	-2.45	-0.54	-4.07	-0.57	-4.34	-	-	0.16	0.83
Thanksgiving Day	-1.10	-23.32	-1.46	-44.53	-1.38	-18.83	-1.85	-9.56	-0.53	-2.49
Day After Thanksgiving (Black Friday)	-0.55	-12.01	-0.87	-7.12	-0.93	-18.22	-0.84	-6.17	-0.27	-3.40
Christmas Day	-1.32	-21.63	-1.58	-12.70	-1.59	-8.86	-	-	-0.22	-1.81
Christmas Day Observed	-0.86	-31.72	-0.69	-14.10	-1.20	-23.11	-	-	-	-
Christmas Day*Weekend	0.45	7.58	-0.21	-1.67	0.49	2.75	-	-	-	-
Christmas Eve	-0.89	-14.60	-1.00	-9.39	-1.02	-13.85	-	-	-0.27	-8.72
Birthday of Martin Luther King	-0.31	-3.91	-0.34	-2.27	-0.44	-1.77	-	-	-0.02	-0.31
Memorial Day	-0.11	-3.81	0.00	0.05	-0.38	-2.73	-0.08	-0.48	0.07	0.90
Washington's Birthday	-0.18	-1.57	-0.21	-3.76	-0.28	-1.05	-	-	-0.19	-1.38
Independence Day	0.02	0.33	-0.19	-2.89	-0.34	-3.93	0.26	3.86	-0.04	-0.27
Independence Day Observed	0.00	0.13	0.28	8.68	-0.22	-6.31	0.02	0.60	-	-
Independence Day*Weekend	0.34	5.53	0.77	11.04	0.53	5.94	-0.06	-0.77	-	-
Labor Day	-0.19	-5.02	-0.21	-2.34	-0.25	-4.49	0.17	2.90	-0.08	-0.93
Columbus Day	-0.10	-2.13	0.00	0.09	-0.30	-4.02	0.03	0.32	-0.11	-3.13
Veterans Day	-0.07	-1.40	-0.13	-0.56	-0.06	-1.24	0.60	1.83	-0.05	-1.80
Veterans Day Observed	0.00	0.01	0.03	0.24	-0.04	-0.46	-0.11	-0.45	-0.04	-0.72
Veterans Day*Weekend	0.07	0.78	-0.19	-0.83	-0.01	-0.25	-1.74	-8.53	0.05	0.51
Good Friday	0.00	0.04	-0.11	-0.94	0.03	0.49	-0.03	-0.32	-0.04	-0.79
Cherry Blossom Festival Saturday	0.20	3.87	-	-	-	-	-	-	-	-

Cherry Blossom Festival Weekday	0.03	1.38	-	-	-	-	-	-	-	-
Monday	-0.02	-1.56	-0.04	-2.08	-0.06	-2.24	-0.07	-2.91	-0.03	-1.67
Tuesday	0.01	1.08	0.00	0.17	-0.02	-0.79	-0.01	-0.36	0.02	1.02
Thursday	0.01	0.90	0.00	0.09	0.01	0.38	0.01	0.27	0.04	2.50
Friday	0.04	3.21	0.01	0.29	-0.02	-0.85	0.11	4.15	0.07	4.22
Saturday	-0.09	-5.68	-0.31	-12.18	-0.37	-12.52	0.18	7.37	-0.08	-3.98
Sunday	-0.22	-13.23	-0.44	-16.40	-0.52	-17.75	-0.02	-0.66	-0.14	-5.19
Average wind speed	-0.01	-7.56	-0.02	-7.44	-0.02	-7.11	-0.01	-5.78	0.00	-0.74
Rainfall indicator	-0.06	-4.79	-0.13	-6.79	-0.08	-4.19	-0.09	-4.83	-0.02	-0.51
Precipitation due to rain (inches)	-0.59	-9.86	-0.35	-6.36	-0.72	-9.24	-0.42	-7.02	-1.44	-5.42
Quadratic effects of rainfall	0.10	3.36	0.01	0.47	0.17	4.01	0.07	2.81	0.72	3.92
Snowfall indicator	-0.08	-0.72	-0.04	-0.88	-0.06	-0.69	-0.13	-1.14	-	-
Snowfall (inches)	-0.28	-2.36	-0.08	-4.70	-0.21	-5.33	-0.18	-4.32	-	-
Snow Depth (inches)	-0.14	-3.58	-0.05	-7.97	-	-	-0.18	-9.67	-	-
Maximum temperature (in Fahrenheit)	0.00	-0.11	0.04	3.95	0.07	5.89	-0.01	-0.21	0.17	1.55
Quadratic effects of maximum temperature ÷ 100	0.00	3.16	0.00	-0.30	0.00	-2.64	0.00	2.06	0.00	-1.25
Cubic effects of maximum temperature ÷ 10,000	0.00	-4.95	0.00	-0.97	0.00	0.72	0.00	-3.22	0.00	0.95
3-day moving average of maximum temperature ÷ 100	0.00	2.96	0.00	3.51	0.01	3.45	0.01	3.43	0.00	0.24
30-day moving average of maximum temperature ÷ 10	0.01	7.08	0.01	4.12	0.00	1.10	0.01	4.54	0.00	1.37
Maximum dewpoint (°F) ÷ 10	-0.01	-9.11	-0.01	-7.58	-0.01	-4.51	-0.01	-9.41	0.00	-1.74
Number of Observation	2917		1826		1704		1727		975	
R-Squared	0.8464		0.9057		0.8951		0.8685		0.6546	
Adjusted R-Squared	0.8431		0.9028		0.8916		0.8650		0.6379	
Durbin Watson's Test	1.6114		1.4055		0.9168		1.3784		1.2043	
MAPE	0.1715		0.2053		0.2400		0.2092		0.1233	

Note: Year and Month Indicator Variables are excluded from this table but were included in the model estimation.
- Indicates that the variable is not included in the respective model

8.7.2 Member and non-member trips

New Year's Day and Independence Day on the weekend generally induces higher total and member ridership but lower non-members ridership. For Labor Day, total and member ridership mostly exhibit negative effects whereas non-members exhibit positive effects. Memorial Day was found to be statistically significant for both member and non-member trips but displayed opposite effects. Member trips were lower on Memorial Day in all five systems, but higher for non-member trips. These findings may reflect that non-members use this holiday for leisure trips. On the other hand, Memorial Day is a high observance holiday and therefore, members are not commuting. The results show that the ridership was lower on Thanksgiving Day for total and member but are mixed for non-member ridership. These findings reflect the fact that members are the predominant percentage of the bikeshare population. Therefore, breaking down member and non-member trips is crucial for understanding which type of users are using the bikeshare system on these special days.

For the effects of Independence Day, it was found that both the actual day and the observed day generally display mixed effects on total trips. However, the effects of these days showed clearer patterns when broken down into member and non-member trips. There was a significant negative relationship between Independence Day (for both actual and observed days) and member trips. For non-member, there was a significant positive relationship except for Bluebike on observed Independence Day.

A non-federal holiday, such as the day after Thanksgiving (Black Friday) generally has lower ridership for both members and non-members. But, for non-members in Washington D.C., there is increased ridership on Black Friday. Generally, Good Friday exhibits no effect on bikeshare ridership.

There is significantly higher ridership during the Cherry Blossom Festival in Washington, D.C. on Saturday for both member and non-members. It could be because this is a large annual event that attracts both residents and tourists. But, on the weekdays during this event, members show no increase in ridership. This may propose that on the weekday, members are using the bikeshare system to go to work, rather than leisure activities. Conversely, non-member ridership is high on the weekday of this festival. This may be that these non-members are tourists.

The results in Table 8.5 and Table 8.6 show that on Fridays, Saturdays and Sundays, member trips were lower compared to Wednesdays, while non-member trips were higher. This could reflect the findings of lower total ridership on Fridays and weekends because members dominate the bikeshare culture. However, in Nice Ride, since the member and non-member trips share an even distribution amongst the system, total ridership was higher on the weekend.

Table 8.5 Log-linear regression model results of member bikeshare trips in five systems with robust standard errors

Variable Description	**Capital Bikeshare Washington D.C.**		**Divvy Chicago**		**Bluebike Boston**		**Nice Ride Minneapolis**		**Metro Bikeshare Los Angeles**	
	Mean	**t-stat**	**Mean**	**t-stat**	**Mean**	**t-stat**	**Mean**	**t-stat**	**Mean**	**t-stat**
Constant	6.88	24.98	7.07	37.73	6.83	21.41	4.11	4.76	2.37	0.81
New Year's Day (actual days, including weekend)	-1.14	-15.46	-1.38	-25.12	-0.96	-4.74	-	-	-0.94	-4.27
New Year's Day Observed (only observed days, Friday, or Monday)	-0.90	-7.90	-0.78	-25.15	-0.33	-5.91	-	-	-	-
New Year's Day*Weekend	0.46	6.32	0.68	11.15	0.82	4.10	-	-	-	-
New Year's Eve	-0.33	-4.30	-0.53	-3.27	-0.63	-3.99	-	-	-0.13	-0.64
Thanksgiving Day	-1.44	-43.36	-1.74	-47.90	-1.50	-22.09	-2.36	-11.70	-0.91	-8.47
Day After Thanksgiving (Black Friday)	-0.87	-26.30	-1.02	-12.40	-1.00	-16.14	-1.02	-6.84	-0.53	-3.13
Christmas Day	-1.77	-24.93	-1.96	-32.65	-1.64	-6.22	-	-	-0.93	-3.83
Christmas Day Observed	-1.20	-43.81	-0.94	-21.13	-1.25	-22.39	-	-	-	-
Christmas Day*Weekend	0.35	5.08	0.34	5.71	0.45	1.72	-	-	-	-
Christmas Eve	-0.97	-14.63	-1.00	-11.02	-1.01	-12.41	-	-	-0.33	-3.46
Birthday of Martin Luther King	-0.41	-5.29	-0.40	-2.64	-0.50	-2.11	-	-	-0.15	-1.14
Memorial Day	-0.47	-14.49	-0.48	-11.81	-0.52	-2.75	-0.62	-3.99	-0.33	-2.30
Washington's Birthday	-0.34	-3.02	-0.28	-6.37	-0.34	-1.16	-	-	-0.38	-6.92
Independence Day	-0.35	-7.66	-0.65	-12.27	-0.82	-7.77	-0.29	-4.10	-0.45	-3.73
Independence Day Observed	-0.42	-18.08	-0.35	-13.33	0.00	-0.03	-0.42	-15.81	-	-
Independence Day*Weekend	0.33	5.98	0.56	10.03	1.44	13.08	0.22	2.96	-	-
Labor Day	-0.49	-16.03	-0.70	-11.73	-0.43	-5.24	-0.47	-4.51	-0.47	-2.79
Columbus Day	-0.21	-4.99	-0.09	-1.90	-0.33	-4.49	0.01	0.15	-0.11	-3.81
Veterans Day	-0.12	-2.11	-0.15	-0.59	-0.09	-1.14	0.58	1.78	-0.05	-1.62
Veterans Day Observed	-0.07	-2.36	-0.03	-0.29	-0.05	-0.52	-0.41	-1.71	-0.27	-2.86
Veterans Day*Weekend	0.12	1.56	-0.08	-0.29	0.02	0.26	-2.18	-10.58	-0.07	-0.31
Good Friday	-0.08	-2.06	-0.16	-1.70	-0.01	-0.14	-0.08	-0.95	-0.05	-0.46
Cherry Blossom Festival Saturday	0.09	2.22	-	-	-	-	-	-	-	-

Cherry Blossom Festival Weekday	0.02	1.36	-	-	-	-	-	-	-	-
Monday	-0.05	-3.48	-0.06	-3.54	-0.07	-2.28	-0.08	-3.50	-0.04	-1.85
Tuesday	0.01	0.70	0.00	0.14	-0.03	-0.87	0.00	-0.18	0.00	-0.19
Thursday	0.01	0.34	-0.01	-0.48	0.01	0.33	-0.01	-0.60	0.02	1.05
Friday	-0.01	-0.49	-0.05	-3.01	-0.05	-1.53	-0.03	-1.21	-0.02	-0.97
Saturday	-0.34	-22.74	-0.60	-31.08	-0.53	-17.03	-0.40	-16.95	-0.44	-17.08
Sunday	-0.46	-30.61	-0.73	-35.67	-0.66	-21.73	-0.49	-21.76	-0.49	-15.24
Number of Observation	2917		1826		1704		1727		975	
R-Squared	0.8436		0.9234		0.8745		0.8619		0.6756	
Adjusted R-Squared	0.8403		0.9209		0.8703		0.8582		0.6599	
Durbin Watson's Test	1.5550		1.4332		0.7834		1.093		0.7622	
MAPE	0.1590		0.1603		0.2605		0.1999		0.1599	

Note: Weather, year, and month variables are excluded from this table but were included in the model estimation.
- Indicates that the variable is not included in the respective model

Table 8.6 Log-linear regression model results of non-member bikeshare trips in five systems with robust standard errors

Variable Description	Capital Bikeshare Washington D.C.		Divvy Chicago		Bluebike Boston		Nice Ride Minneapolis		Metro Bikeshare Los Angeles	
	Mean	**t-stat**	**Mean**	**t-stat**	**Mean**	**t-stat**	**Mean**	**t-stat**	**Mean**	**t-stat**
Constant	1.78	4.12	3.65	13.82	1.28	2.43	2.63	3.24	-1.83	-0.30
New Year's Day (actual days, including weekend)	0.56	4.49	0.40	3.31	-0.35	-0.70	-	-	0.77	9.14
New Year's Day Observed (only observed days, Friday, or Monday)	-0.29	-1.14	0.55	8.87	1.11	13.26	-	-	-	-
New Year's Day*Weekend	-0.63	-4.97	-0.28	-2.15	1.33	2.65	-	-	-	-
New Year's Eve	0.52	2.19	-0.02	-0.08	0.02	0.09	-	-	0.64	3.68
Thanksgiving Day	0.36	3.92	0.23	2.16	-0.47	-1.49	-1.28	-6.66	0.03	0.07
Day After Thanksgiving (Black Friday)	0.69	8.54	0.01	0.02	-0.35	-1.69	-1.12	-7.50	0.06	0.22
Christmas Day	0.52	3.53	0.23	1.02	-1.44	-2.40	-	-	0.76	4.07
Christmas Day Observed	0.71	15.40	1.09	14.34	-0.62	-7.16	-	-	-	-
Christmas Day*Weekend	-0.26	-1.75	-2.24	-9.77	0.88	1.47	-	-	-	-
Christmas Eve	-0.23	-1.26	-0.45	-1.21	-0.82	-3.65	-	-	-0.12	-0.72
Birthday of Martin Luther King	0.49	3.50	-0.31	-1.75	0.12	0.28	-	-	0.32	3.27
Memorial Day	0.73	13.01	1.05	8.76	0.14	0.64	0.60	3.17	0.80	11.87
Washington's Birthday	0.85	4.62	0.03	0.12	0.32	1.43	-	-	0.21	0.50
Independence Day	0.96	12.33	0.84	5.68	0.73	7.95	0.86	9.25	0.55	2.68
Independence Day Observed	0.83	19.43	0.85	18.18	-1.00	-15.05	0.19	5.14	-	-
Independence Day*Weekend	-0.13	-1.39	-0.57	-3.73	-2.25	-20.59	-0.82	-8.35	-	-
Labor Day	0.72	9.68	0.94	7.18	0.53	1.79	0.96	12.45	0.55	2.80
Columbus Day	0.41	3.79	0.53	7.72	0.02	0.15	0.05	0.33	-0.08	-0.78
Veterans Day	0.29	3.36	-0.14	-0.82	0.23	0.72	0.83	2.79	-0.09	-1.51
Veterans Day Observed	0.56	8.51	-0.07	-0.19	0.12	1.02	0.85	3.93	0.49	10.36
Veterans Day*Weekend	-0.02	-0.14	0.05	0.21	0.07	0.23	-1.42	-7.25	0.29	4.22
Good Friday	0.29	3.40	0.36	1.45	0.34	1.45	0.20	1.75	0.03	0.27
Cherry Blossom Festival Saturday	0.30	3.55	-	-	-	-	-	-	-	-

Cherry Blossom Festival Weekday	0.14	3.68	-	-	-	-	-	-	-	-
Monday	0.15	5.00	0.15	4.49	-0.02	-0.36	-0.02	-0.43	0.00	-0.12
Tuesday	0.05	1.66	0.03	0.87	0.00	-0.05	0.01	0.34	0.07	1.86
Thursday	0.09	3.44	0.12	3.61	0.07	1.34	0.06	1.92	0.10	2.59
Friday	0.36	13.41	0.46	13.88	0.21	4.22	0.40	11.34	0.30	7.99
Saturday	0.94	33.01	0.98	25.94	0.41	7.51	0.93	28.75	0.56	14.11
Sunday	0.79	27.02	0.80	19.35	0.19	3.42	0.64	18.69	0.50	12.50
Number of Observation	2916		1826		1698		1727		975	
R-Squared	0.8851		0.9422		0.8598		0.8877		0.6072	
Adjusted R-Squared	0.8826		0.9404		0.8551		0.8847		0.5881	
Durbin Watson's Test	1.3087		1.5097		1.029		1.3091		0.8747	
MAPE	0.3150		0.3614		0.508		0.3031		0.2684	

Note: Weather, year, and month variables are excluded from this table but were included in the model estimation.
- Indicates that the variable is not included in the respective model

8.7.3 Individual holiday effects summary

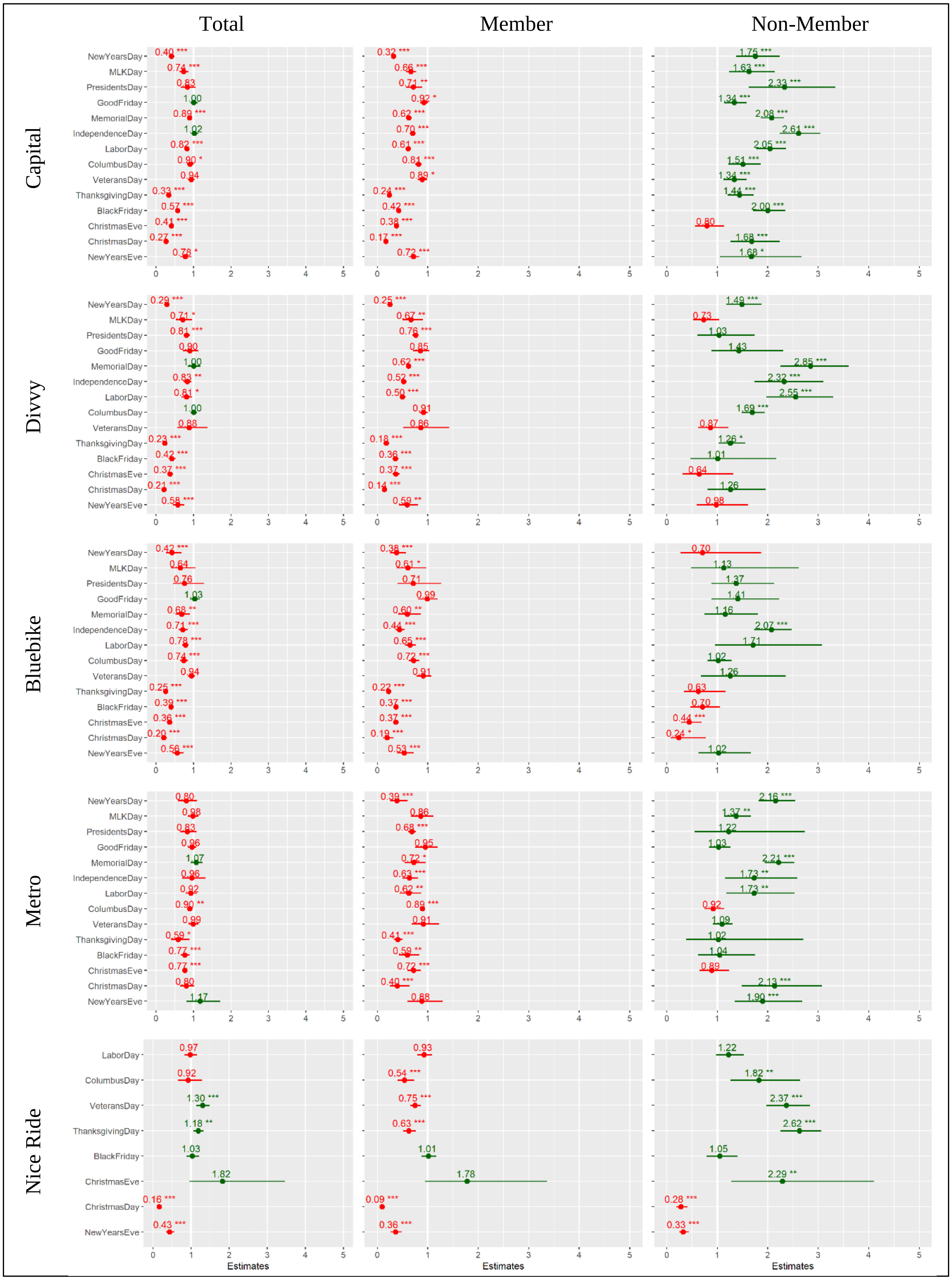


Figure 8.2 Holiday-specific expected ridership ratios by bikeshare system

A summary of total, member, and non-member ridership during federal and non-work holidays across five systems in chronological order is presented in Figure 8.2. It appears that holidays-specific ridership effects are seasonal for non-member trips. Holidays during the spring and summer months, such as Independence Day and Labor Day, display higher expected ridership ratios compared to the cooler months. For total and member ridership, these ridership ratios are generally negative, but warmer month holiday effects still tend to be higher than the cooler months.

8.8 Conclusions and Recommendations

This study primarily focused on exploring the impacts of holidays on daily bikeshare ridership of five U.S. bikeshare systems. Compared to prior literature reviewed in this chapter, the findings of this study are more spatially and temporally robust as this study considers holiday effects over multiple years and across multiple locations. Log-linear regression models were estimated to infer the impacts of holidays on total, member, and non-member bikeshare trips across hypotheses.

Similar to prior studies, it was found that total system-level ridership tends to decrease on federal holidays compared to comparable non-holidays. But when accounting for heterogeneity in user types, it is observed that non-member riders take more trips on federal holidays while members take fewer trips. To address the differences in ridership patterns between holidays and weekends, this study found support that both federal holidays and weekends share similar effects on ridership, but the direction of the effects are controlled by user types. It is seen that member ridership is lower on federal holidays and weekends while non-member ridership is higher.

The study generally found that the effects of total and member ridership on holidays were negative while non-member ridership was positive. But the magnitude of the effects varies based on the individual holidays. Thanksgiving and Christmas, for example, experience lower total ridership than Memorial Day and Labor Day. This study also found that the total system-level ridership is likely to increase on federal holiday on a weekend compared to the same federal holiday on a weekday. Lastly, it was found that holidays with higher observance by businesses resulted in lower member-level ridership which could be due to less commuting-based trips.

These findings on bikeshare ridership patterns during special calendar days have implications for the management of bikeshare systems, local economies, and public health. Because of increased non-member ridership on holidays, municipalities and bikeshare systems can concentrate information and advertising campaigns around non-users on holidays. Also, offering a differing ticketing structure, such as reducing the cost of an annual bikeshare pass if one subscribes on a holiday may increase bikeshare ridership on holidays. Local restaurants can aid in this effort by offering discounts or other special offers during holidays if a user rides bikeshare to the restaurant. Efforts to increase public awareness and experience with bikeshare systems during holidays may lead to improved public health and additional system membership. The general trip patterns of bikeshare systems, which induce shorter trips and slower travel speeds, can increase local business activity, and encourage local patronage in locations with bikeshare. Lastly, the results of this study can aid bikeshare systems in developing general system-level pre-and repositioning efforts and for determining maintenance and cleaning schedules.

Future work could examine bikeshare ridership on holidays at the station-level. Analyzing holiday effects at the station-level could aid in understanding where users are traveling which could broaden the understanding of travel behavior patterns on these days. Future work may also explore additional bikeshare systems to strengthen hypothesis results further.

One of the limitations in this study is that although most of the Durbin Watson test statistics show little autocorrelation, some systems show positive autocorrelation, especially for the non-member models. Analyzing systems with the same model specifications result in more comparable results across five systems. Although some of the models show autocorrelation, to have consistent and comparable results, the same model specifications were used across all five systems. To account for possible autocorrelation, autoregressive integrated moving average (ARIMA) model could be used.

In terms of special events, it was found that the Cherry Blossom Festival induces higher ridership for both members and non-members in Washington D.C. But there is very limited research on the effects of special events on bikeshare system ridership. There is no theoretical framework to explain which events are expected to impact bikeshare usage. The authors analyzed other special events, but these were not included in this chapter to conserve space. Using the Managing Travel for Planned Special Events Handbook (Latoski et al., 2003) to characterize special events, it was found that event operation characteristics impact special event bikeshare ridership. Park-based events had more positive effects on ridership than street-based events. Multi-day events generally showed increases in ridership compared to single day events, and national events contributed to the highest bikeshare numbers. In general, festivals have higher impacts on bikeshare ridership compared to other event types. The National Cherry Blossom Festival – being both a national event and a festival – had increased ridership across all bikeshare users' types. Future work could seek to build and expand a testable framework using event characteristics to explore these impacts. These characteristics could include proximity to bikeshare stations, level of outdoor involvement, the size of the event space/area, event time span, and the distribution of attendees' time use. This future work could be useful to urban planners and civic leaders in the consideration of road closures and other traffic changes to ensure pedestrian safety on these special days. Corral services can be offered during high bikeshare volume events to reduce the stress of parking and encourage other active modes. Lastly, planning future events to market to encourage greater member usage could increase bikeshare ridership success.

For special event selection, future work can also include analyzing data at the microscopic level. Qualitative study, including interviews, focus group, ethnographies, and surveys, may help to extract the types of special events that are highly correlated to bikeshare usage. This can be done for non-registered users as well to understand bikeshare ridership during special events. Future work could also seek to build a testable framework using event characteristics and machine learning techniques to explore these impacts.

8.9 Acknowledgments

This study is funded by the Center for Urban Transportation Research at the University of South Florida, along with the FHWA Dwight David Eisenhower Transportation Fellowship. The authors are thankful for Eren Yuksel and Brian Staes for brainstorming and developing ideas with the authors and for insights on this work. The authors would also like to thank Christina Van Allen for editorial support. Lastly, the authors would like to thank the bikeshare agencies that provided the data for this study.

Chapter 9

Comprehensive Analysis of Leisure Activity Variety as an Instrumental Outcome of Social Capital

Trang Loung, Michael Maness

9.1 Introduction

The need for travel is often perceived as demand derived from participation in various activities. From an activity-travel perspective, most trips are categorized by purpose such as mandatory, maintenance or discretionary/leisure travel. Leisure activities and their induced travel are differentiated by their voluntary and often social nature. Leisure activity is often informed by, motivated by, and performed with connections within an individual's social network. Measuring and modeling social and socially motivated activity have increasingly been recognized within the transportation research community as important for understanding leisure activity behavior. Gathering evidence of the linkage between social networks and activity generation, Kim et al. (2018) reviewed transportation studies that analyzed the impacts of individuals' social network characteristics on the frequency of social activity participation across three measures:

- Network size (Carrasco et al., 2008; Carrasco and Miller, 2006; Sharmeen et al., 2014; Van den Berg et al., 2009, 2010, 2015)
- Relationship type (Carrasco et al., 2008; Carrasco and Miller, 2006; Sharmeen et al., 2014; Van den Berg et al., 2009, 2012; Carrasco and Miller, 2009; Frei and Axhausen, 2008)
- Tie strength (Carrasco et al., 2008; Carrasco and Miller, 2006; Sharmeen et al., 2014; Frei and Axhausen, 2008; Van den Berg et al., 2012)

Regarding network size, more frequent activity participation was associated with larger networks. Regarding relationship type, there was no clear consensus on its impact on ego-alter activity frequency due to varying methodologies and classification schemes. Regarding tie strength, several studies asked specifically about whether respondents had strong, medium, or weak ties with their alters. Each study found that higher social activity frequency was generated by stronger ties. Among those considerations of social relationships, there is a lack of a cohesive theory linking social network characteristics to leisure activity outcomes (Parady et al., 2019):

> *"Although several studies have analysed leisure activity generation, and the relationship between social network and social interactions, to date there is no well-established and validated theory on the nature of this relation, hence most studies, including the present one, are of an exploratory nature."* (p. 546)

Maness (2017b) attempted to develop such a theory inductively through social tie generation principles (Kadushin, 2012). Maness theorized that larger strong-tie networks and weak-tie diversification increases activity variety and frequency. More specifically, strong ties were hypothesized to increase activity frequency due to similarity in interests. The weak-ties role in expanding one's social circle allows for diversification in activities and status-seeking behavior thus leading to greater activity variety. Using a name generator for strong-tie characteristics and a position generator for weak-tie characteristics, the theory was supported by improved model fit. Maness' study, however, was restrained by survey data with a limited activity space as well as lacking mobility data and direct measures of accessible resources (Maness, 2017).

To address the limitations of a coherent theory linking social network characteristics to leisure activity outcomes, this chapter aims to further develop a social capital theory of leisure activity behavior. Lin's conception of individual-level social capital provides a schema which links the concepts of network characteristics (size, density, range, structural holes) with access to socially embedded resources that result in purposive outcomes. The research proposes that such a schema can potentially serve to link prior research in activity-travel behavior on the impacts of social networks on leisure activity behavior.

This study examines a sample's access to social capital and its association with the leisure activity outcome of activity diversity. This research proposed two research questions to test this social capital theory of leisure activity behavior:

1. Do individuals with greater social capital (i.e. access to social resources) have greater diversity in their leisure activities?
2. Does instrumental support play a more significant role than expressive support in enabling different leisure activity participation, thus suggesting that activity variety is an instrumental outcome rather than an expressive outcome?

This study aims to answers those questions using insights gathered from a self-administered web-based survey designed specifically to test differences in social capital and its relevance in a leisure activity context.

9.2 Literature Review

This section starts by highlighting transportation research on leisure activity, then describes the interpretation of how social capital is defined as benefits from embedded social resources and concludes by providing linkages between leisure activity behavior and social capital.

9.2.1 Leisure activity outcomes

Leisure was categorized as activities encompassing social, recreational, physical, and cultural events that are not bounded by work or maintenance tasks (Ettema and Schwanen, 2012).

As leisure activity engagement offers considerable benefits for personal wellbeing and social connections, the 2018 American Time Use Survey conducted by the US Bureau of Labor Statistics revealed that Americans spent an average of 5.27 hours on daily leisure activities compared to only 3.57 work hours. Stauffacher et al. (2015) surveyed the diversity of leisure activities and the different contributing motives for increasing activity variety. Calastri et al. (2020) examined the influence of social relationship strength on leisure activity participation.

9.2.2 Social capital as access to resources

The concept of social capital describes how individuals acquire beneficial assets and services through social interactions. Among various definitions of social capital, Lin's formulation of social capital as embedded social resources has a strong methodological synergy in activity-travel behavior since it aligns well with the individual-level basis of most activity and travel research. Specifically, Lin (2001) proposes three primary elements of social capital: a) resource embeddedness in social networks, b) resource accessibility, and c) resource use for action-oriented aspects. Lin further defines three processes involved in the creation and use of social capital: a) *investment* in social capital, b) *access to* and *mobilization* of social capital, and c) *returns* of social capital.

Häuberer's (2011) schema (Figure 9.1) clarifies Lin's theory of the three processes and thus provides causal relationships between preconditions, social capital, and outcomes. Individuals are preconditioned in a societal context and have access to individually owned resources and assets. Access to social resources is mobilized through social networks and their structural properties. Smaller, denser networks help maintain social connections and promote continued access to group resources through trust and reciprocation. This leads to more resources for expressive actions and subsequently, capitalization of expressive outcomes. Lin classifies expressive outcomes as mental health, physical health, and life satisfaction. In contrast, larger, wider social networks enable new contacts but results in less intimate social support (Lin 2001). These diverse, lower-maintenance connections can broaden access to new resources for profit or influential gain and lead to more resources for instrumental action. Lin classifies instrumental outcomes as wealth, power, and status (Lin 2001).

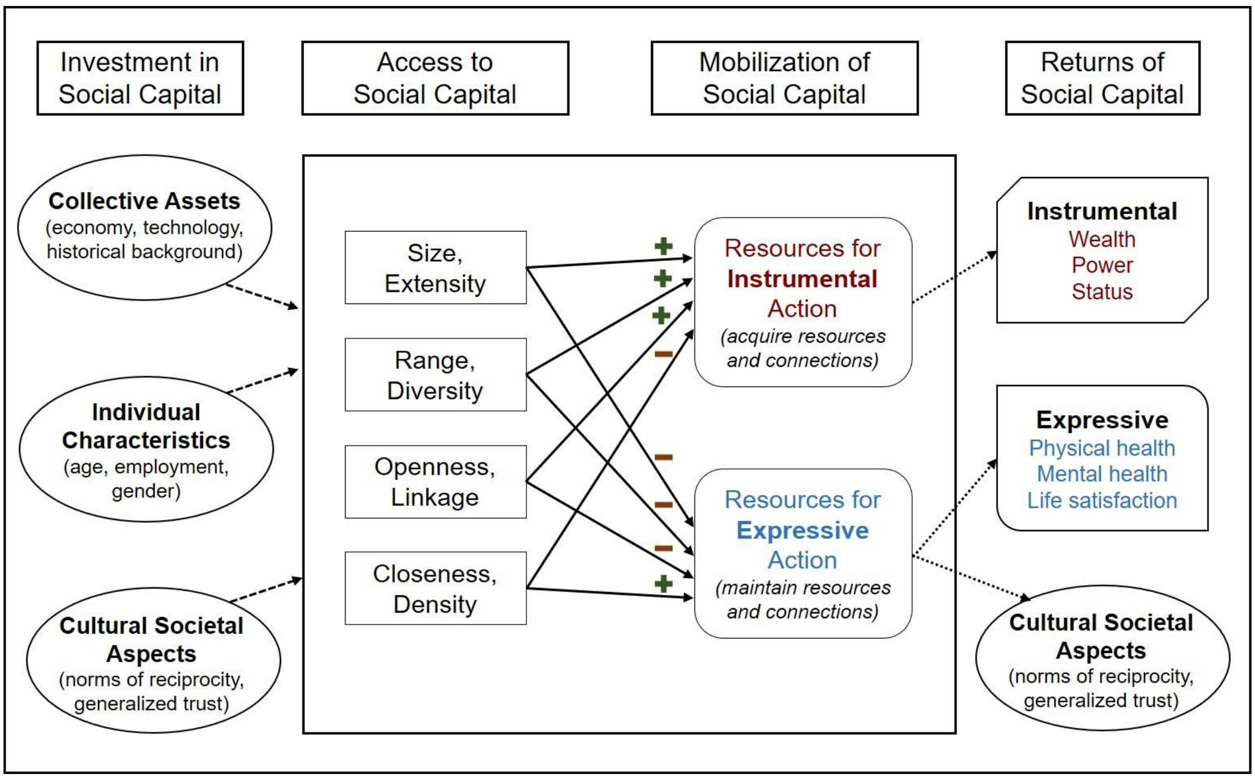


Figure 9.1 Conceptual diagram of social capital and outcomes – adapted from Häuberer (2011)

To operationalize Lin's conception of social capital as socially embedded resources in an activity-travel context, *activity behavior* is considered as *returns of social capital*. Hence, aspects of activity behavior must be described as instrumental or expressive outcomes (or combinations thereof). As a starting point, suppose leisure activity behavior is described through the variety of activity types participated in and the frequency at which someone participates in those activities.

As instrumental actions focus on acquiring new connections and resources from social interactions, leisure activity behavior that enables these actions will be described as instrumental leisure activity outcomes. Seeking out new connections is seen as a form of social tie creation due to brokerage and status seeking. This enables access to resources from people outside of one's own social circles, often through participating in the interests of those people and their extended networks. Subsequently, an individual may need to expand their activity space to participate in different activities to reap instrumental benefits. An expanded activity space increases participation in a greater variety of activity types, since the individual would participate in both their preferred activities and their weak/loose social ties' preferred activities. Additionally, as information flow is enabled through brokerage over weak ties, the individual's knowledge of new activities increases through their instrumental actions.

9.2.3 Theory of leisure activity variety as an instrumental outcome

Carrasco and Cid-Aguayo (2012) and Maness (2017a) attempt to link social capital to activity behavior through measuring social network characteristics. Parady et al. (2019) also links network size and club membership to social activity variety. Their efforts have, however, been limited by an unclear linkage between social resources and leisure activity preferences. By using Lin's social capital concept with the ability to measure structural and mobilized resources, the effects of leisure activity for enabling expressive and instrumental outcomes could be explored.

9.2.4 Conceptualization of leisure activity variety as an instrumental outcome

Numerous studies have attempted to understand leisure behavior for its beneficial impacts on individuals, households, and society (Lloyd and Auld, 2002). Characterized by a voluntary and social nature, leisure activities are often enabled by interpersonal relationships. Even activities that can be performed individually in solitude such as playing guitar, reading fiction, or working puzzles can be more enjoyable when joined by others. Our research interprets Lin's social capital definition in an activity context and hence, conceptualizes leisure activity as the capitalization of social resources. Lin specified wealth, power, and social status as three instrumental elements that can be mobilized from the investments in social networks.

While leisure activities can be perceived as personal hobbies without connections to one's instrumental benefits, this study proposes and tests the theory that leisure variety (e.g. a collection of unique leisure activities) exemplifies instrumental outcome and is achieved by instrumental resources embedded in one's social network. When an individual's leisure behavior encompasses a wide range of social, recreational, entertainment activities, this variety of leisure activity extends beyond personal hobbies and can be an indicator of wealth, power, and status. Wealth can be attained through the creation of new connections and richness in information, opportunities, and idea exchange. One's affluence can also be manifested as the accumulation of pleasures, enjoyments, and self-improvement that were missing from work or maintenance tasks. Leisure activities that offer positive experience can also enhance individuals' productivity and earning potential (Lyubomirshy et al., 2005).

According to Tinsley and Eldredge (1995), service-based leisure activities can offer power to people who have a sense of responsibility to help, comfort, or inspire others. Sports can particularly empower individuals for gratifying one's desire to overcome challenge. Many social activities such as attending church, dancing, dining out, visiting friends and family promote attention and feeling of importance. One can particularly enhance their social status by coordinating leisure activities with others.

9.2.5 Measuring social capital and social resource access

Lin (2001) suggests measuring social capital as resources (assets) in social networks. A simplified approach to enumerate the instrumental and expressive resources embedded in a social network would involve asking individuals about their social contacts and each contacts' available resources. Called the name generator approach, this is the primary technique used in transportation studies of social capital (Kim et al., 2018). While the name generator approach provides extensive, detailed information, it has practical limitations in data collection due to respondent burden and contact recall biases. Additionally, there are concerns about its sensitivity to survey mode, particularly in self-administered formats (Joye et al., 2019).

The position generator technique primarily measures access to instrumental support that can help to attain wealth, power, and status. This approach focuses on hierarchal measures of people's access to resources by relating them with their contacts' societal positions. The disparities in resource allocation across many societies are attributable to societal hierarchies. Generally, those with higher societal position have more resources and can create new connections more easily (Lin, 2001). To determine instrumental connections, the position generator relies on the tendency of societies to associate occupations with prestige and status. The position generator measures a person's ties with individuals across various occupations – which have varying levels of prestige and status across society. A position generator is a list of various occupations for respondents to indicate whether they have individuals in their social network with those occupations. Thus, the position generator can provide an indirect measure of social capital access and specifically focuses on instrumental resources.

The resource generator approach combines the name generator/interpreter and position generator to directly measure social resource access. Using a list of specific resources, "the resource generator asks if [respondents] would have anyone to turn to should they need to access one or more of a range of resources" (Crossley et al., 2015: p. 49). This enables the resource generator flexibility in answering a range of research questions. This is particularly important since all social capital is not equivalent and cannot be mobilized for all purposes. Van der Gaag and Snijders' resource generator surveyed in the Netherlands identified four types of social resources: personal support, political/financial skills, personal skills, and prestigious/educational related social capital (Van der Gaag and Snijders, 2005). Additionally, since the resources are specific, resource generators can also directly measure differences between instrumental and expressive resource access.

By combining the position generator and resource generator approaches, an individual's social integration and network range can be measured alongside his or her "concrete resources available through social relations" (Joye et al., 2019: p. 23). Joye and colleagues explain that: "considering social networks and the social resources embedded in them in the form proposed here is a way to escape from an individualistic survey perspective and to defend the perspective of 'life in context', the context being the network of relations with family or friends but also the society in which the individual is living" (pp. 11-12).

9.3 Hypotheses

To test the proposed theory of leisure activity variety as an instrumental outcome, this research explores the following hypotheses:

1. Individuals with greater social capital (i.e. greater access to social resources) participate in more types of leisure activities than those with less social capital.
2. Activity diversity is primarily an instrumental outcome in a social capital process with greater reliance on instrumental resource than expressive outcomes.

9.4 Survey Design and Social Capital Measures

9.4.1 Data collection

A cross-sectional survey was designed to better understand social factors influencing the leisure activity participation. The survey design and administrations are outlined in Table 9.1. Table 9.1 also describes the selection of three different survey distributing platforms to ensure a qualified and diverse group of respondents. As participants from Amazon Mechanical Turk (MTurk) were younger and less female than the Qualtrics panel, an equal gender quota was set for the Qualtrics panel and the survey was also distributed to 118 female participants in the Prolific platform. Responses were rejected if the respondents spent less than five minutes on the survey or had substantial missing or inconsistent/invalid answers.

The survey consists of sections on activity space, social capital, mobility/accessibility, individual and household characteristics. The activity space section of the survey asks about: 1) leisure activity variety and frequency, 2) household mandatory and maintenance activities, and 3) work and school demand.

Table 9.1 Survey methodology summary

Characteristic	**Description**
Survey name	Leisure Activity and Social Resources Survey
Time frame	November - December 2019
Target population	US adults aged 18 years and older
Sampling frame	*Qualtrics Panels*: Adults with internet in an internet-based survey panel *Prolific*: Women with internet in an internet-based survey panel *Amazon MTurk*: Registered US MTurk workers with task approval rates > 90% and at least 100 approved tasks
Recruitment	*Qualtrics Panels*: Email recruitment with varied incentive unknown to researchers *Prolific and MTurk*: Advertised task with $3.00 incentive
Sample size	1,297 responses after data cleaning
Sample design	Non-probability samples with quota-based (gender, age) for Qualtrics Panels and Prolific, and no quota for MTurk
Sampling source	Qualtrics Panel (46% of the data), MTurk (46%), and Prolific (8%)
Administration mode	Self-administered via the internet
Time dimension	Cross-sectional survey
Level of observations	Individual, household

9.4.2 Leisure activity variety

Survey respondents were presented an activity list and asked to choose the specific activities they participated in over the last three months. Leisure activity variety was asked using a list of 86 unique activity types. Adopted from Tinsley and Eldredge, 77 out of their 82 activities were adopted, while arcade games, collecting bottles, shortwave radio listening, volunteering for crisis intervention, and watching television were excluded due to being outdated, dependent on specific crisis events, or overabundance (Tinsley and Eldredge, 1995). Nine additional leisure activities were added including: attending festivals and parades, board gaming, joyriding, gambling, gardening in community gardens, softball, singing karaoke, video games, and visiting amusement/theme parks. The list of 86 activities was presented across four pages. Activities that are similar such as hiking and backpacking were listed adjacently to reduce the likelihood of inaccurate counts.

9.4.3 Social capital measures

Instrumental Support

As outlined in previous section, this study utilized a position generator to measure the instrumental support of social capital. To maintain comparability with the 2004 Social Capital Surveys described in Lin et al. (2013), a list of 22 occupations for the position generator was applied in this survey to measure access to instrumental social resources. Respondents were asked to indicate if they personally knew someone (a relative, friend, or acquaintance) on a first-name basis with that occupation. Each occupation also has a prestige score determined by the Standard International Occupational Prestige Scale that was later used to calculate each occupation's prestige/status (Lin and Ao, 2008). The list of 22 occupations is reported in decreasing order of prestige score [in brackets] but were presented to respondents randomly ordered:

1. Professor [78]
2. Lawyer [73]
3. Chief Executive Officer [70]
4. Congressman or Congresswoman [64]
5. Production manager [63]
6. Middle school teacher [60]
7. Personnel manager [60]
8. Writer [58]
9. Nurse [54]
10. Computer programmer [51]
11. Bookkeeper [49]
12. Administrative assistant in a large company [49]
13. Police officer [40]
14. Farmer [38]
15. Receptionist [38]
16. Operator in a factory [34]
17. Hairdresser [32]
18. Taxi driver [31]
19. Security guard [30]
20. Full-time babysitter or nanny [23]
21. Janitor [21]
22. Hotel bellhop [20]

In order to comprehensively assess individual's access to instrumental support, principal component analysis was applied on three measures of position generator (network occupational volume, highest reach, and range of reach) to derive a network occupational composite score. In particular, the first metric, the *network occupational volume*, is the sum of all different occupations a respondent knew in his or her social network on a first name basis. The network occupational volume was normalized by dividing the sum by the maximum number of known occupations, which is 22 based on the given list of position generator. Second, *network occupational highest reach* is the highest prestige score of all occupations reached in social network. The network occupational highest reach was also normalized by dividing the highest prestige score by 78, which is the prestige score for a professor. About 41% of the survey respondents reported knowing a professor. Third, *network occupational range of reach* is defined as the difference between the maximum and minimum prestige scores of the occupations reached in social network. The network occupational range of reach was also normalized by dividing the range of prestige score by 58, which is the difference between the most prestigious occupation and the least prestigious occupation in the given list. About 12% of survey respondents knew no one or only one occupation, which resulted in a range of zero. Subsequently, *network occupational composite score* was computed as the sum of the principal rotations for the three aforementioned measures as follows:

$$Network\ occupational\ composite\ score = (0.42 * \ number\ of\ occupations + 0.54 * highest\ prestige\ score\ + \ 0.73 * range\ of\ prestige\ scores)$$

Expressive Support

In order to measure the availability of resources that individuals can access through their social network; an 11-item resource generator proposed by Joye et al. (2019) was included in the questionnaire. Survey respondents were advised: “This section is about who you would turn to for help, if you needed it, in different situations. For each situation, please choose who you would turn to first for help. (If there are several people you are equally likely to turn to, please choose the one who you feel is closest to you).” Joye et al. (2019) defined three dimensions of social support that can be offered by immediate family, other family member, close friend, neighbor, someone I work with, other friend or acquaintance, or no one:

<u>Practical support:</u>

a. Help you for a household or a garden job that you can’t do yourself
b. Help you around the house if you were sick and had to stay in bed for a few days
c. Look after you if you were seriously ill

<u>Informational support:</u>

d. Help you with finding a job
e. Help you with finding a new place to live
f. Help you look for information about a serious personal health issue

g. Help you if you needed advice on administrative formalities and on other legal matters

<u>Emotional support:</u>

h. Be there if you felt a bit down or depressed and wanted to talk about it
i. Give you advice on family problems
j. Make you feel appreciated for who you really are
k. Be there if you just wanted to talk about your day

To further account for the emotional support provided by an individual's core network size, respondents were asked: "From time to time, most people discuss important matters with other people. Looking back over the last three months, think about the people whom you discussed matters that are important to you. How many people were you able to recall?" This core network size (counted as the number of people who they discussed important matters over the last three months) was a generalized version of Burt's name generator in the General Social Survey (Burt, 1984). Using the confirmatory factor analysis described in Joye et al. (2019), three latent variables of social support were obtained from the list of 11 items as follows:

$$Practical\ support = 1.000a + 1.002b + 0.899c$$

$$Informational\ support = 1.000d + 0.967e + 0.954f + 0.925g$$

$$Emotional\ support = 1.000h + 1.127i + 1.182j + 0.845k$$

Subsequently, the expressive social support offered by social network was computed based on the three latent dimensions of social support and the core network size as follows:

$$Expressive\ support = (1.000 * practical + 1.458 * informational + 1.025\ emotional + 0.501 * core\ network\ size)$$

The two indicators for instrumental and expressive support were subsequently used as explanatory variables in models estimating the number of unique leisure activities.

9.5 Data Descriptions

Survey respondents reported participating in between zero and 56 different activities over the last three months (Figure 9.2). About 50% of respondents participated in between 5 and 15 different activities over the three-month period.

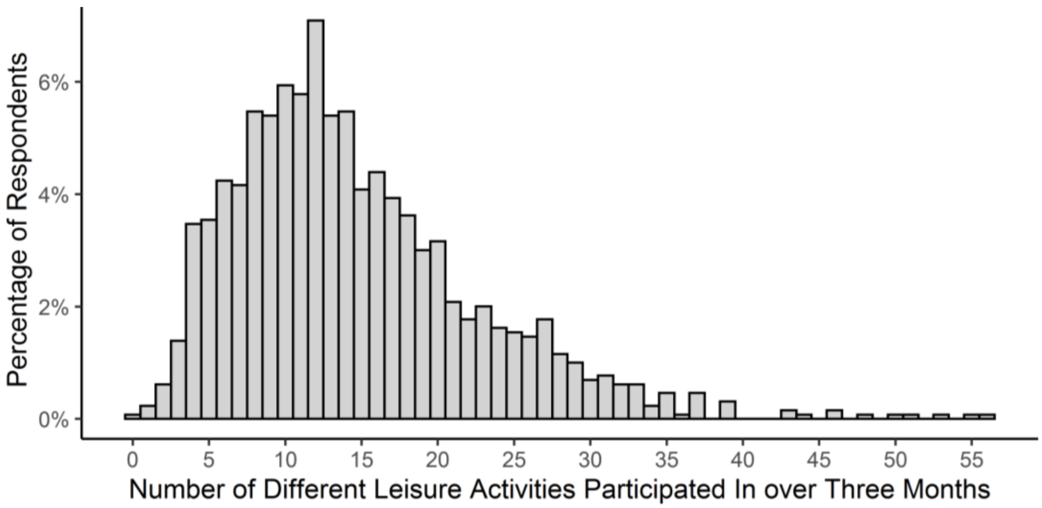


Figure 9.2 Activity variety distribution

Survey correspondence was recorded after participants accepted the survey consent and met the quota on gender and age. After data cleaning, descriptive statistics of variables used in regression models and characteristics of 1,297 survey respondents are provided in Tables 9.2 and 9.3.

Table 9.2 Descriptive statistics of model variables (N=1,297)

Variable Description	**Min**	**Max**	**Mean**	**SD**
Number of different activity types	0	56	14.6	8.1
Social Capital Measures				
Network occupational composite	0	1	0.57	0.24
Social support composite	0	1	0.86	0.20
Sociodemographic Attributes				
Having a driver license and at least one motorized vehicle in the household	0	1	0.88	0.33
Having no driver license and no access to public transit	0	1	0.01	0.10
Having a disability or illness affecting the ability to travel	0	1	0.07	0.26
Personality score for being extraverted	0	7	3.53	1.64
Personality score for being open to experience	0	7	4.91	1.35
Age of respondents	19	91	46.91	16.90
Having a bachelor’s degree	0	1	0.55	0.50
Median income (in $1,000)	0	275	73.20	55.60
Identified as white	0	1	0.82	0.38
Widowed marital status	0	1	0.04	0.20
Female respondent	0	1	0.50	0.50
Hours spent on cooking and chores per week	0	80	9.42	7.95
Work hours per week	0	100	26.68	19.44
Sampled from Prolific panel	0	1	0.08	0.28
Sampled from Qualtrics panel	0	1	0.46	0.50

Table 9.3 Descriptive statistics of sample characteristics (N=1,297)

Age	Mean	46.9
	Median	44.0
	Standard deviation	16.9
Education	Less than high school	0.7%
	High school graduate/GED	12.2%
	Some college, no degree	18.4%
	Associate degree	9.5%
	Bachelor's degree	35.4%
	Graduate degree	19.6%
Employment	Full-time	57.0%
	Part-time	12.7%
	Retired	15.9%
	Student (not employed for pay)	1.9%
	Disabled (not employed for pay)	2.9%
	Not employed for pay	6.7%
Gender	Female	50.3%
	Male	49.2%
Household income	Under $15,000	5.5%
	$15,000–$24,999	9.6%
	$25,000–$34,999	10.6%
	$35,000–$49,999	16.0%
	$50,000–$74,999	22.3%
	$75,000–$99,999	14.1%
	$100,000–$149,999	12.5%
	$150,000–$199,999	4.9%
	$200,000–$249,999	2.2%
	$250,000 or more	2.0%
Household size	One person	22.2%
	Two people	38.2%
	Three or more people	39.6%
Marital status	Married/domestic partnership	48.1%
	Widowed	4.0%
	Divorced	9.6%
	Separated	1.0%
	Living with a partner	7.2%
	Never been married	29.8%
Race/ethnicity	American Indian or Alaska Native	1.5%
	Asian	6.9%
	Black or African American	9.7%
	Hispanic, Latino or Spanish origin	0.6%
	White	82.0%
	Other race/ethnicity	2.5%
Household vehicles	No vehicle	7.8%
	One	37.2%
	Two	39.1%
	Three or more	15.7%

9.6 Methodology

Activity variety is defined as the number of different leisure activities that survey respondents had participated in over the last three months. With a mean of 14.6 unique leisure activities and a standard deviation of 8.1 unique leisure activities, the over-dispersion of the data distribution suggests the use of negative binomial regression. The model is specified through the following mathematical expectation:

$$E(y_n|x_n, s_n, p_n, r_n) = \exp(\beta x_n + \gamma I_n + \delta E_n) \qquad (1)$$

where

y_n = activity variety for individual n,

x_n = individual and household characteristics for individual n,

I_n = measures of instrumental support

E_n = measures of expressive support, and

β, γ, δ = model parameters.

In NB regression, an individual's probability $P(y_n)$ of participating in y_n different activities is defined as follows:

$$P(y_n) = \frac{\Gamma(1/\alpha + y_n)}{\Gamma(1/\alpha) y_n!} \left(\frac{1/\alpha}{(1/\alpha) + \lambda_n}\right)^{1/\alpha} \left(\frac{y_n}{(1/\alpha) + \lambda_n}\right)^{y_n} \qquad (2)$$

where $\Gamma(\cdot)$ is the gamma function, $\lambda_n = \exp(\beta(x_n + I_n + sE_n) + \varepsilon_i)$, and $exp\ (\varepsilon_n)$ is a Gamma-distributed disturbance term with unit mean and variance given by the dispersion parameter α. Model parameters were estimated using quasi-maximum likelihood estimation using the *MASS* package in R.

9.7 Estimation Results

In order to distinguish the impacts of social capital and individual/household characteristics on leisure activity variety, model estimation results of a non-social model and models with social capital measures are presented in this section. The variables used in the estimated negative binomial models are summarized in Table 9.4 and sample characteristics' influence on activity variety outcome are provided in Table 9.4. Eleven respondents with missing data were excluded in the model.

The non-social model was used as the base to compare the significant improvement of activity variety model's prediction capability. The likelihood ratio test substantiates the hypothesis that social capital are strong determinants in higher variety of leisure activity outcome. Specifically, while the non-social model captures a wide range of socio-demographics, household, and mobility characteristics in predicting a number of different activities, the model converged at -4,186. Likelihood ratio tests for model with social capital measures showed significant model fit. Random-parameter models were tested for the unobserved heterogeneity of these social capital measures, but no significant likelihood improvement suggested the use of fixed-parameter models.

As this study hypothesized that leisure activity variety was primarily an instrumental outcome of a social capital process, quasi-likelihood ratio tests were performed, and the results show support for the hypothesis that higher social capital correlates with higher leisure activity variety.

Table 9.4 Negative binomial model estimation results of activity variety

Parameter description [units]	Nonsocial model		Model with social capital		
	Mean estimate	Standard error	Mean estimate	Standard error	Marginal effect
Social Capital Measures: Network occupational volume*	—	—	**0.646**	**0.059**	**8.96**
Social support composite*	—	—	**0.222**	**0.071**	**3.07**
Other Attributes: Licensed driver with vehicle availability in the household indicator	0.085	0.045	0.075	0.043	1.01
Non-driver with no access to public transit indicator	-0.241	0.140	-0.159	0.135	-2.04
Having disability to travel indicator	-0.067	0.054	-0.089	0.052	-1.19
Personality score for extraversion	**0.022**	**0.008**	0.009	0.008	0.12
Square of personality score for openness to experience	**0.007**	**0.001**	**0.006**	**0.001**	**0.08**
Age of respondents [year]	**-0.004**	**0.001**	**-0.004**	**0.001**	**-0.06**
Having bachelor's degree indicator	**0.098**	**0.029**	**0.057**	**0.028**	**0.78**
Square root of median income [$1,000]	0.006	0.005	0.001	0.005	0.01
Identified as white indicator	0.039	0.036	0.041	0.034	0.56
Widowed indicator	-0.104	0.075	**-0.105**	**0.071**	**-1.39**
Female respondent indicator	-0.027	0.061	-0.064	0.058	-0.89
Weekly cooking and chores time [10h]	**0.381**	**0.090**	**0.344**	**0.086**	**4.77**
Weekly cooking and chores time for women [10h]	-0.187	0.099	-0.157	0.095	-2.18
Square of weekly cooking and chores time [$(10h)^2$]	**-0.129**	**0.035**	**-0.123**	**0.034**	**-1.71**
Square of weekly cooking and chores time for women [$(10h)^2$]	**0.111**	**0.036**	**0.105**	**0.034**	**1.46**
Weekly work time [40h]	0.029	0.033	-0.019	0.031	-0.27
Sampled from Qualtrics panel indicator	-0.096	0.052	**-0.277**	**0.035**	**-3.81**
Sampled from Prolific indicator	**-0.299**	**0.037**	**-0.120**	**0.050**	**-1.58**
Intercept	**2.357**	**0.088**	**1.997**	**0.100**	**N/A**
Dispersion parameter	**6.649**	**0.386**	**7.852**	**0.482**	N/A
Model Statistics: Number of parameters	20		22		
Log likelihood at convergence	-4,185.7		-4,112.8		
Number of observations	1,286		1,286		

Note: * = normalized measure with values [0,1]; bold numbers = estimate p-value ≤ 0.05

9.8 Results

Social Capital Measures

For equivalent comparisons of the social resource effects, the two social measures were normalized as a ratio of reported values distance from the minimum value to the value's range;

this results in a value of zero at the minimum observed value and one at the maximum observed value. Both social capital indicators had statistically significant and strong positive effects on activity variety. The models with instrumental social capital and expressive social support had significantly greater model fit than the non-social models, thus supporting the hypothesis of social capital's conducive effect on an individual's leisure activity participation. Compared to the indicator of expressive resources, the instrumental measures had the greatest normalized impact. A respondent with maximal instrumental social capital had an expected increase in activity variety of 91% more than someone with minimal instrumental social capital. As the network occupational composite score accounted for the volume, range, and highest occupation known in a person's network, this score has the biggest influence on activity variety and contributed to the most significantly improved model fit. The strongest effect derived from position generator attested the hypothesis that leisure activity variety (being more an instrumental than expressive outcome) could be enabled through broader instrumental support. Regarding the expressive resources offered by social network, a respondent with maximal expressive social support experienced about 25% more activity variety than those with minimal expressive social support.

Mobility Impacts

Mobility was hypothesized to have a positive relationship with leisure activity participation due to increasing the size of an individual's activity space. Licensed driver with in-home automobile availability had significantly increased activities variety. In contrast, non-drivers with no access to public transportation had decreased activity variety. Although the activity list had many activities that did not require travel, people who self-reported travel-limiting disabilities or illnesses were also restricted in their activity space. These effects were as expected since auto availability/usage provided greater mobility, and thus increased out-of-home activity variety.

Personality Types

To account for concerns with endogeneity of social capital due to differences in social personality *(12)*, personality types were included in the analysis. In particular, openness to experience was the personality type that exhibited the strongest potential effect on activity variety. It was expected that people who are more extraverted may be more inclined to participate in socially oriented leisure activities. All models demonstrated by positive and significant effects of higher scores on being extraverted as well as open to experience and increased activity variety. Based on the Ten-Item Personality Inventory proposed by Gosling et al. (2003) included in this survey, respondents with high score on extraversion showed much smaller impacts of only 0.009 relative to the social capital measures. Computed from the same personality inventory(Gosling et al., 2003), respondents who agreed that they were “open to new experiences, complex” and disagreed on being “conventional, uncreative” had a significant correlation with increase in more diverse activities. Being open to experience was considered having a non-linear effect on activity outcome.

Sociodemographic Attributes

Age had a significantly negative effect on individuals’ variety of activities. Individuals who earned a bachelor’s degree or higher participated in higher activity variety. Education can be seen as an indicator of cultural capital (Joye et al., 2019) with the expectation that higher cultural capital would be correlated with higher activity variety. Regarding marital status, widowed respondents experienced reduced activity variety which may be attributed to other life changes beyond accessible social support. While increasing disposable income was hypothesized to expand the activities space, the income effect was found to be positive but statistically insignificant. Minority groups were not found to have significantly less activity variety than whites. Other socio-demographic factors, such as the number of workers and children in the household, were also tested (not included in presented models) but did not have significant effects on activity variety.

Gender-based impacts were found to be insignificant for the gender fixed-effect. But the analysis found that the impact of housework varied between genders. It was hypothesized that greater levels of housework would reduce the available time to participate in leisure activities and could therefore reduce leisure activity variety. this relationship was found to hold generally for men above about 28 to 30 hours of housework, under a quadratic formulation. But for the women surveyed, a decrease in leisure activity variety was not found for throughout the observed range of housework hours.

Survey Administration Mode Effects

Finally, results show that respondents in the Prolific and Qualtrics panels reported less activity variety than Mechanical Turk. It is unclear what unobservable factors caused this differentiation between samples but perhaps related to differing participation motivations/needs.

9.9 Conclusion

A comprehensive analysis was conducted to explore the impacts of social capital on leisure activity behavior – particularly activity variety. A social capital theory of activity behavior was presented two hypotheses that 1) social capital is an integral determinant of leisure activity participation, and 2) having access to instrumental social support promotes instrumental outcomes demonstrated by increased in more leisure activity variety. This theory contributes to the limited understanding of activity and travel behavior using social capital concepts discussed by Lin (2001), Häuberer (2011), and Joye et al. (2019). There have been a few studies accounted for social capital influence on leisure activity participation.

This study has also been one of the first attempts to shed deeper insights on a wide range of leisure activities adopted from Tinsley and Eldredge (1995) instead of a small number of common activities. A refined survey was then designed to obtain an extensive collection of leisure activities, social capital measures as well as mobility and sociodemographic characteristics. All of these factors were tested by numerous estimations of count data regression to extract the most significant determinants. Among the variables that have positive correlations with activity variety, social capital measures significantly improved the model fit and have the strongest effects on increasing more diverse activities. Thus, including social capital consideration on an individual's activity space will both help to unravel the unobserved heterogeneity across similar socioeconomic groups and reduce the biases as demonstrated in the non-social model. Instrumental support measured by the position and resource generators indeed had the largest influence on predicting activity variety outcome. The core network size and accessible expressive resources resulted in considerable explaining power, besides the sampling source effect. Several traditional sociodemographic attributes, such as race, gender, and employment status became insignificant or have negligible effects on activity variety after accounting for social network's role in activity participation. The addition of mobility indicators and personality traits were also essential to better predictions of increased numbers of unique leisure activities. The positive correlation of household auto availability with increased out-of-home activities demonstrates the need to assess the effect of alternative modes of transportation on activity participation.

This work contributes to a growing interest of considering the effects of social network characteristics on activity-travel behavior. As social capital has distinct impacts even among homogeneous groups, transportation modelers can derive more refined characteristics from social capital measures to build more socially and behaviorally realistic models.

The results of this empirical study may not be representative of the larger population due to the nonprobability sampling technique. In addition, the self-administrative nature of the survey can amplify the lack of attention or accuracy, which was mitigated during the survey refinement and data validation process. To capture missing activities due to incomplete recollection, an activity diary with detailed temporal and spatial information would provide a more complete picture of each activity as well as its transportation components. Endogenous effects of people who participate in different activities to create more connections and subsequently enrich their social capital are difficult to disentangle from a cross-sectional survey.

Subsequently, future work can focus on examining the causal effects of social capital, mobility, and activity participation. From another perspective, individuals may report the same number of different leisure activities but those sets of activities may emerge from distinct social and psychological motives. More analysis can be conducted to measure the impacts of social capital on various aspects of each activity because leisure activities may or may not be conducive to a person's social interactions (e.g. visiting friends vs. reading books).

Furthermore, since this study focuses on operationalize instrumental outcomes, the next immediate step would be to explore expressive outcomes as the mobilization of expressive social support. As expressive outcomes are essential to strengthening mental wellbeing, physical health, and life satisfaction, the impact of leisure activities that can be conducted with no travel or personal interaction requirements is an emerging need given the concerns of public health and lack of access to public venue

Conclusions and Policy Implications

Several critical issues have emerged in recent years in the fields of highway safety, alternative transportation modes, and activity and travel behavior modeling. The Chapter 2 of this report addressed highway safety, where issues relating to big data, traditional data and the tradeoffs between prediction and causality in highway-safety analysis. The analysis of highway accident data is largely dominated by traditional statistical methods (standard regression-based approaches), advanced statistical methods (such as models that account for unobserved heterogeneity), and data-driven methods (artificial intelligence, neural networks, machine learning, and so on). These methods have been applied mostly using data from observed crashes, but this can create a problem in uncovering causality since individuals that are inherently riskier than the population as a whole may be over-represented in the data. In addition, when and where individuals choose to drive could affect data analyses that use real-time data since the population of observed drivers could change over time. This issue, the nature of the data, and the implementation target of the analysis imply that analysts must often tradeoff the predictive capability of the resulting analysis and its ability to uncover the underlying causal nature of crash-contributing factors. The selection of the data-analysis method is often made without full consideration of this tradeoff, even though there are potentially important implications for the development of safety countermeasures and policies. This chapter shows the issues involved in this tradeoff with regard to specific methodological alternatives and presents researchers with a better understanding of the trade-offs often being inherently made in their analysis.

Chapter 3 addressed issues related to the recent growth in the popularity of mobility-on-demand (ridehailing), which has substantially disrupted the transportation market by providing a variety of new transportation options. The chapter presented a statistical model of individuals’ usage rates of ridehailing services. Using a sample of recently collected data, a mixed logit model (multinomial logit model with random parameters) of ridehailing-usage rate was estimated and, in addition to traditional socio-demographic factors, several travel and health-related variables were found to play statistically significant roles for ridehailing usage. Specifically, age, gender, income, household size, vehicle ownership, typical parking time, and the nature of commutes were some of the significant variables found in model estimation results. In addition, self-assessed health, high body mass index, and registration for other shared mobility services were all found to play roles in ridehailing usage. The results suggest that ridehailing usage tends to be driven by a wide variety of individual characteristics and lifestyle choices.

Chapter 4 considered the emerging phenomena of carsharing and specifically the renting of personal vehicles (peer-to-peer carsharing), which has become increasingly popular in the U.S. The chapter studied the attitudes, perceptions, and decision process through which individuals decide to offer their car for rent in such peer-to-peer carsharing. A stated preference survey was designed and disseminated where survey respondents were asked how likely they would be to rent their car (extremely unlikely, unlikely, unsure, likely, extremely likely). The survey questionnaire also collected detailed socio-demographic information, as well as data on travel behavior and travel patterns. These data were then used to estimate a random parameters ordered probit model of their likelihood of renting their car. Some of the variables found statistically significant determinants of the willingness to rent a personal vehicle were gender, age, income, household composition, vehicle ownership, living location with respect to a grocery store, and participation in other shared mobility modes. These findings and especially the gender and income related variables were found to complement prior literature and offered additional layer of understanding of the factors determining the supply side of peer-to-peer carsharing.

Chapter 5 turned to the area of safety, modeling issues relating to aggressive driving, which has become a national traffic-safety concern. Looking at single-vehicle crashes, this study investigates differences between resulting crash-injury severities when aggressive and non-aggressive driving behavior is observed, and how these differences changed over time by estimating random parameters multinomial logit models with unobserved heterogeneity in means and variances. Model estimates show that there were significant differences in driver-injury severities resulting from aggressive and non-aggressive driving, and that the effect of factors that determine injury severities changed significantly over time (statistically significant temporal instability). However, it is noteworthy that crashes involving non-aggressive drivers had many explanatory variables that produced temporally stable marginal effects, whereas crashes involving aggressive drivers had only one such variable (restraint belt usage). Importantly, this suggests the possibility that temporal instability found in many recent safety studies may be driven by a subset of the crash population, and that there may be temporal stability in many crashes.

Chapter 6 continued the safety emphasis by looking at issues relating to work zone safety, a critical issue with likely nationwide infrastructure initiatives. Using Florida work zone data from the 2012 to 2017 time period, resulting driver-injury severities in single-vehicle work zone crashes were studied by estimating random parameters logit models that allow for possible heterogeneity in the means and variances of parameter estimates. The model estimates produced significantly different parameters for each of the year from 2012 to 2017, and a fundamental shift in unobserved heterogeneity, suggesting statistically significant temporal instability. In addition, in several key instances, the marginal effects of individual parameter estimates show marked differences between one year and the next. However, these findings may not be the sole result of variations in driver behavior over time as has been argued in past research that has found temporal instability. This is because each work zone has a unique set of characteristics and, with the sample of work zones changing from one year to the next as highway maintenance and construction is undertaken in different locations, this work-zone sample variation could be a substantial source of the observed temporal instability.

Chapter 7 shifted to the behavioral analysis of the zero-price effect phenomenon. Prior research has shown that a short-term free public charging program could possibly increase plug-in electric vehicle sales, decrease oil consumption, and decrease greenhouse gas emissions. To deepen the understanding of consumer behavior relating to free charging, this research analyzed the zero-price effect to estimate a monetary value of free charging. To arrive at accurate estimation, data from stated preference survey were used to estimate latent class models of attribute non-attendance. The values calculated via different computations methods were then compared. The national mean zero-price effect for public charging ranged from $0.95 to $1.40 across the models. Because the collected sample was correctly weighted and national representativeness was achieved, the findings from this work can help to assess policies which offer free public charging infrastructure.

Chapter 8 provided an analysis of bikeshare behavior during holidays. Existing literature showed mixed results relating to the ridership impacts of holidays, as some research showed that these days may result in higher ridership, while others showed no effect. To control for these aspects, this chapter used a multi-city study of the effect of holidays on system-level ridership using a log-linear regression model with robust standard errors. The results showed the impacts of holidays on bikeshare system ridership for different user types among systems in the Washington DC, Chicago, Boston, Los Angeles, and Minneapolis metro areas. Several hypotheses were built and tested for examining the expected effects of holidays on bikeshare usage. A major finding from this chapter is that federal holidays negatively affect member ridership and positively affect non-member ridership. It was also found that different federal holidays have dissimilar effects on total ridership.

Finally, in Chapter 9, social capital was explored in relation to leisure activity behavior. Motivated by the influence of social capital on leisure activity behavior, this chapter proposed a theory that leisure activity variety is an instrumental outcome and thus mostly affected by instrumental social resources. The theory underlined two hypotheses that 1) social capital is an integral determinant of leisure activity participation, and 2) having access to instrumental social support promotes instrumental outcomes demonstrated by increased leisure activity variety. This theory was comprehensively tested on the number of different unique leisure activities collected from 1,297 survey respondents. This refined and specially designed survey is the first in the transportation literature to use both position generator and resource generator to measure social capital. Results from negative binomial regression models demonstrated that instrumental support indeed had the largest influence on predicting activity variety outcome. This chapter showed that as social capital has distinct impacts even among homogeneous groups, transportation modelers can derive insights from social capital measures to build more socially and behaviorally realistic models.

By addressing critical contemporary modeling issues, this report provides practical insights into several emerging modeling issues in the transportation field. The insights provided herein can form the basis for effective transportation policies relating to highway safety, the effectiveness of ridehailing as a mode of travel, peer to peer car sharing, electric vehicle adoption, bikesharing, and the impact of social capital on travel decisions.

References

AAA Foundation for Traffic Safety, 2016. Prevalence of Self-Reported Aggressive Driving Behavior: United Sates, Washington, DC.

AAA Foundation for Traffic Safety, 2009. Aggressive Driving: Research Update. AAA Foundation for Traffic Safety, Washington, DC. Available at: http://www.aaafoundation.org/pdf/AggressiveDrivingResearchUpdate2009.pdf.

AASHTO, 2010. Highway Safety Manual. American Association of State Highway and Transportation Officials, Washington, DC.

Abay, K., Mannering, F., 2016. An empirical analysis of risk-taking in car driving and other aspects of life. Accident Analysis and Prevention 97, 57-68.

Abay, K., Paleti, R., Bhat, C., 2013. The joint analysis of injury severity of drivers in two-vehicle crashes accommodating seat belt use endogeneity, Transportation Research Part B 50, 74-89.

Abdel-Aty M., Pande A., 2005. Identifying crash propensity using specific traffic speed conditions. Journal of Safety Research 36(1), 97-108.

Abdel-Aty M., Pande A., Das A., Knibbe W., 2008. Assessing safety on Dutch freeways with data from infrastructure-based intelligent transportation systems. Transportation Research Record 2083, 153-161.

Abdelwahab, H., Abdel-Aty, M., 2001. Development of artificial neural network models to predict driver injury severity in traffic accidents at signalized intersections. Transportation Research Record 1746, 6-13.

Ajzen, I., 1991. The theory of planned behavior. Organizational Behavior and Human Decision Processes 50(2), 179-211.

Al-Shihabi, T., Mourant, R., 2007. Toward more realistic driving behavior models for autonomous vehicles in driving simulators. Transportation Research Record 1843, 41-49.

Alemi, F., Circella, G., Handy, S., Mokhtarian, P., 2018. What influences travelers to use Uber? Exploring the factors affecting the adoption of on-demand ride services in California. Travel Behaviour and Society 13, 88–104.

Alnawmasi, N., Mannering, F., 2019. A statistical assessment of temporal instability in the factors determining motorcyclist injury severities. Analytic Methods in Accident Research 100090, 1-20.

Alonso, F., Esteban, A., Montoro, L., Serge. A. 2019. Conceptualization of aggressive driving behaviors through a perception of aggressive driving scale (PAD), Transportation Research Part F 60, 415-426.

Anastasopoulos, P., Mannering, F., 2009. A note on modeling vehicle accident frequencies with random-parameters count models. Accident Analysis and Prevention 41(1), 153-159.

Anastasopoulos, P., Mannering, F., 2011. An empirical assessment of fixed and random parameter logit models using crash- and non-crash-specific injury data. Accident Analysis and Prevention 43(3), 1140-1147.

Anderson, D., 2014. Not just a taxi? For-profit ridesharing, driver strategies, and VMT. Transportation 41(5), 1099–1117.

Angrist, J., Pischke, J., 2009. Mostly harmless econometrics: An empiricist's companion. Princeton University Press, Princeton, NJ.

Angrist, J., Pischke, J., 2015. Mastering metrics: The path from cause to effect. Princeton University Press, Princeton, NJ.

Angrist, J., Pischke, J., 2017. Undergraduate econometrics instruction: Throughout classes, darkly. Working paper 23144, National Bureau of Economic Research, Cambridge, MA.

Ariely, D., 2008. Predictably irrational. Harper Collins, New York.

Ballús-Armet, I., Shaheen, S., Clonts, K., Weinzimmer, D., 2014. Peer-to-peer carsharing exploring public perception and market characteristics in the San Francisco Bay area, California. Transportation Research Record: Journal of the Transportation Research Board 2416, 27-36.

Barbour, N., Menon, N., Zhang, Y., Mannering, F., 2019b. Shared automated vehicles: a statistical analysis of consumer likelihoods and concerns. Transport Policy 80, 86-93.

Barbour, N., Zhang, Y., Mannering, F., 2019a. An exploratory analysis of the role of socio-demographic and health-related factors in ridesourcing behavior. Working paper, University of South Florida, Tampa, FL.

Barbour, N., Zhang, Y., Mannering, F., 2019c. A statistical analysis of bikesharing use and its potential as an auto-trip substitute. Journal of Transport and Health 12, 253-262.

Bardhi, F., Eckhardt, G., 2012. Access-based consumption: The case of car sharing. Journal of Consumer Research 39(4), 881–898.

Beck, K., Wang, Q., Moser M., Mitchell, M., 2006. Concerns, dispositions and behaviors of aggressive drivers: what do self-identified aggressive drivers believe about traffic safety? Journal of Safety Research 37(2), 159-165.

Behnood A., Mannering F., 2016. An empirical assessment of the effects of economic recessions on pedestrian-injury crashes using mixed and latent-class models. Analytic Methods in Accident Research 12, 1-17.

Behnood, A., Mannering, F., 2015. The temporal stability of factors affecting driver-injury severities in single-vehicle crashes: Some empirical evidence. Analytic Methods in Accident Research 8, 7-32.

Behnood, A., Mannering, F., 2017a. The effect of passengers on driver-injury severities in single-vehicle crashes: A random parameters heterogeneity-in-means approach. Analytic Methods in Accident Research 14, 41-53.

Behnood, A., Mannering, F., 2017b. Determinants of bicyclist injury severities in bicycle-vehicle crashes: A random parameters approach with heterogeneity in means and variances. Analytic Methods in Accident Research 16, 35-47.

Behnood, A., Mannering, F., 2019. Time-of-day variations and temporal instability of factors affecting injury severities in large-truck crashes. Analytic Methods in Accident Research 23, 100102, 1-17.

Behnood, A., Roshandeh, A., Mannering, F., 2014. Latent class analysis of the effects of age, gender, and alcohol consumption on driver-injury severities. Analytic Methods in Accident Research 3-4, 56-91.

Belk, R., 1988. Possessions and the extended self. Journal of Consumer Research 15(2), 149–168.

Belk, R., 2010. Sharing. Journal of Consumer Research 36(5), 715-734.

Belk, R., 2014. You are what you can access: sharing and collaborative consumption on-line. Journal of Business Research 67(8), 1595–1600.

Bellotti, V., Ambard, A., Turner, D., Gossmann, C., Demkova, K., Carroll, J., 2015. A muddle of models of motivation for using peer-to-peer economy systems. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, ACM, 1085-1094.

Ben-Akiva, M., Lerman, S., 1985. Discrete choice analysis: Theory and application to travel demand. MIT Press, Cambridge, Mass.

Benjaafar, S., Kong, G., Li, X., Courcoubetis, C., 2018. Peer-to-Peer product sharing: Implications for ownership, usage and social welfare in the sharing economy. Management Science 26(9), 951-956.

Benkler, Y., 2004. Sharing nicely: On shareable goods and the emergence of sharing as a modality of economic production. Yale Law Journal 114(2), 273-358.

Berdoulat, E., Vavassori, D., Sastre, M. 2013. Driving anger, emotional and instrumental aggressiveness, and impulsiveness in the prediction of aggressive and transgressive driving, Accident Analysis and Prevention 50. 758-767.

Bhat, C. 2001. Quasi-Random Maximum Simulated Likelihood Estimation of the Mixed Multinomial Logit Model. Transportation Research Part B 17(1), 677-693.

Bhat, C., 1996. A generalized multiple durations proportional hazard model with an application to activity behavior during the work-to-home commute. Transportation Research Part B 30(6), 465-480.

Bhat, C., 1998. A model of post-home arrival activity participation behavior. Transportation Research Part B 32(6), 387-400.

Bhat, C., 2003. Simulation estimation of mixed discrete choice models using randomized and scrambled Halton sequences. Transportation Research Part B 37(9), 837-855.

Bhat, C., 2015. A new generalized heterogeneous data model (GHDM) to jointly model mixed types of dependent variables. Transportation Research Part B 79, 50-77.

Bhat, C., 2018. New matrix-based methods for the analytic evaluation of the multivariate cumulative normal distribution function. Transportation Research Part B 109, 238-256.

Bhat, C., Born, K., Sidharthan, R., Bhat, P., 2014. A count data model with endogenous covariates: formulation and application to roadway crash frequency at intersections. Analytic Methods in Accident Research 1, 53-71.

Bhat, C., Eluru, N., 2009. A copula-based approach to accommodate residential self-selection effects in travel behavior modeling. Transportation Research Part B 43, 749-765.

Björklund, G. 2008. Driver irritation and aggressive behavior, Accident Analysis and Prevention 40(3), 1069-1077.

Bluebikes. System Data. 2020 Motivate International Inc., 2020. https://www.bluebikes.com/system-data. Accessed January 5, 2020.

Bocker, L., Meelen, T., 2017. Sharing for people, planet or profit? Analysing motivations for intended sharing economy participation. Environmental Innovation and Societal Transitions 23, 28-39.

Boersch-Supan, A., Hajivassiliou, V., 1993. Smooth unbiased multivariate probability simulators for maximum likelihood estimation of limited dependent variable models. Journal of Econometrics 58, 347–368.

Brathwaite, T., Walker, J., 2018. Causal inference in travel demand modeling (and the lack thereof). Journal of Choice Modelling 26, 1-18.

Brownstone, D., Train, K., 1999. Forecasting new product penetration with flexible substitution patterns. Journal of Econometrics 89, 109–129

Burns, P., Wilde, G., 1995. Risk taking in male taxi drivers: relationships among personality, observational data and driver records. Personality and Individual Differences 18(2), 267-278.

Burt, R.S. Network items and the general social survey. Social networks, 1984. 6:293-339.

Calastri C., S. Hess, D. Palma, R.C. dit Sourd, 2020. Capturing relationship strength: a choice model for leisure time, frequency of interaction and ranking in name generators. Travel Behaviour and Society, 20, 290-299.

Calvi, A., Benedetto, A., De Blasiis, M., 2012. A driving simulator study of driver performance on deceleration lanes. Accident Analysis and Prevention 45, 195-203.

Capital Bikeshare. System Data. 2020 Lyft, Inc., 2020. https://www.capitalbikeshare.com/system-data. Accessed January 5, 2020.

Carlson, R., Wishart, J., Stutenberg, K., 2016. On-road and dynamometer evaluation of vehicle auxiliary loads. SAE International Journal of Fuels and Lubricants, 9(1), 260-268.

Carrasco, J., Cid-Aguayo, B.. 2012. Network capital, social networks, and travel: an empirical illustration from Concepción, Chile. Environment and Planning A, 44, 1066-1084.

Carrasco, J., Miller, E, 2006. Exploring the propensity to perform social activities: a social network approach. Transportation, 33, 463-480.

Carrasco, J., Miller, E., 2009. The social dimension in action: a multilevel, personal networks model of social activity frequency between individuals. Transportation Research Part A: Policy and Practice, 43, 90-104.

Carrasco, J., Hogan, B., Wellman, B., Miller, E., 2008. Agency in social activity interactions: the role of social networks in time and space. Journal of Economic and Social Geography, 99:562-583.

Cervero, R., Golub, A., Nee, B., 2007. City carshare: longer-term travel demand and car ownership impacts. Transportation Research Record: Journal of the Transportation Research Board 1992, 70–80.

Cerwick, D., Gkritza, K., Shaheed, M., Hans, Z., 2014. A comparison of the mixed logit and latent class methods for crash severity analysis. Analytic Methods in Accident Research 3-4, 11-27.

Chakraborty, D., Bunch, D., Lee, J., Tal, G., 2019. Demand drivers for charging infrastructure-charging behavior of plug-in electric vehicle commuters. Transportation Research Part D: Transport and Environment, 76, 255-272.

Chang, L.-Y., 2005. Analysis of freeway accident frequencies: Negative binomial regression versus artificial neural network. Safety Science 43(8), 2005, 541-557.

Clewlow, R., 2016. Carsharing and sustainable travel behavior: results from the San Francisco bay area. Transport Policy 51, 158–164.

Clewlow, R., Mishra, G., 2017. Disruptive transportation: The adoption, utilization, and impacts of ride-hailing in the United States. Report. Davis: Institute of Transportation Studies, University of California (UCD-ITS-RR-17-07).

Coll, M., Vandersmissen, M., Theriault, M., 2014. Modeling spatio-temporal diffusion of carsharing membership in Quebec City. Journal of Transport Geography 38, 22–37.

Corcoran, J., Li, T., Rohde, D., Charles-Edwards, E., and Mateo-Babiano, D., 2014. Spatio-temporal patterns of a Public Bicycle Sharing Program: the effect of weather and calendar events. Journal of Transport Geography, 292-305.

Cornwell, B., Laumann, E., Schumm, L., 2008. The social connectedness of older adults: A national profile. American Sociological Review 73(2), 185-203.

Correia, G., Viegas, J., 2011. Carpooling and carpool clubs: clarifying concepts and assessing value enhancement possibilities through a stated preference web survey in Lisbon, Portugal. Transportation Research Part A 45(2), 81-90.

Costain, C., Ardron, C., Habib, K., 2012. Synopsis of users' behaviour of a carsharing program: a case study in Toronto. Transportation Research Part A 46(3), 421–434.

Crossley, N., Bellotti, E., Edwards, G., Everett, M., Koskinen, J., Tranmer, M., 2015. Social network analysis for ego-nets: Social network analysis for actor-centred networks. Sage.

Daina, N., 2014. Modelling electric vehicle use and charging behavior. (Doctoral dissertation, Imperial College London).

Daina, N., Polak, J., Sivakumar, A., 2015. Patent and latent predictors of electric vehicle charging behavior. Transportation Research Record: Journal of the Transportation Research Board, 116-123.

Daina, N., Sivakumar, J., Polak, W., 2017. Electric vehicle charging choices: Modelling and implications for smart charging services. Transportation Research Part C: Emerging Technologies 81, 36-56.

Dale, S., Krueger, A., 2002. Estimating the payoff to attending a more selective college: An application of selection on observables and unobservables. Quarterly Journal of Econometrics 117(4), 1491-1527.

Darke, P., Chung, C., 2005. Effects of pricing and promotion on consumer perceptions: it depends on how you frame it. Journal of Retailing, 81(1), 35-47.

Deffenbacher, J., Filetti, L., Lynch, R., Dahlen, E., Oetting, E., 2003. Anger, aggression, risky behavior and crash-related outcomes in three groups of drivers. Behavior Research and Therapy 41(3), 333-349.

Deffenbacher, J., Huff, M., Lynch, R., Oetting, E., Salvatore, N., 2000. Characteristics and treatment of high anger drivers. Journal of Counseling Psychology 47(1), 5-17.

Deka, D., Fei, D., 2019. A comparison of the personal and neighborhood characteristics associated with ridehailing, transit use, and driving with NHTS data. Journal of Transport Geography 76, 24–33.

Delen, D., Sharda, R., Bessonov, M., 2006. Identifying significant predictors of injury severity in traffic accidents using a series of artificial neural networks. Accident Analysis and Prevention 38(3), 434-444.

Diamantopoulos, A., Schlegelmilch, B., Sinkovics, R., Bohlen, G., 2003. Can socio-demographics still play a role in profiling green consumers? A review of the evidence and an empirical investigation. Journal of Business Research 56(6), 465-480.

Diamond, W., Sanyal, A.. 1990. The effect of framing on the choice of supermarket coupons. NA-Advances in Consumer Research Volume 17: Eds. Goldberg, M.E., Gorn, G., and Pollay, R.W. Provo, UT: Association for Consumer Research, 488-493.

Dias, F., Lavieri, P., Garikapati, V., Astroza, S., Pendyala, R., Bhat, C., 2017. A behavioral choice model of the use of car-sharing and ride-sourcing services. Transportation 44(6), 1307–1323.

Diekmann, A., Jungbauer-Gans, M., Krassnig, H., Lorenz, S., 1996. Social status and aggression: a field study analyzed by survival analysis. Journal of Social Psychology 136, 761-768.

Dill, J., McNeil, N., Howland, S., 2019. Effects of peer-to-peer carsharing on vehicle owners' travel behavior. Transportation Research Part C 101, 70-78.

Divvy. Divvy Data. 2020 Divvy Bikes, 2020. https://www.divvybikes.com/system-data. Accessed January 5, 2020.

Dowling, R., Maalsen, S., Kent, J., 2018. Sharing as sociomaterial practice: Car sharing and the material reconstitution of automobility. Geoforum 88, 10-16.

Dunckel-Graglia, A., 2013. Women-only transportation: How “pink” public transportation changes public perception of women’s mobility. Journal of Public Transportation, 16, 85-105.

El-Assi, W., Mahmoud, M., and Habib, K., 2018. Effects of built environment and weather on bike sharing demand: a station level analysis of commercial bike sharing in Toronto.

Ellison-Potter, P., Deffenbacher, P., Deffenbacher, J. 2001. The effects of trait driving anger, anonymity, and aggressive stimuli on aggressive driving behavior. Journal of Applied Social Psychology 31(2), 431-443.

Eluru, N., Bhat, C., 2007. A joint econometric analysis of seat belt use and crash-related injury severity. Accident Analysis and Prevention 39(5), 1037-1049.

Eluru, N., Bhat, C., Hensher, D., 2008. A mixed generalized ordered response model for examining pedestrian and bicyclist injury severity level in traffic crashes. Accident Analysis and Prevention 40(3), 1033-1054.

Erdogan, S., Cirillo, C., Tremblay, J., 2015. Ridesharing as a green commute alternative: A campus case study, International Journal of Sustainable Transportation 9(5), 377-388.

Eren, E., Uz, E, 2020. A review on bike-sharing: The factors affecting bike-sharing demand. Sustainable Cities and Society, 2020, Sustainable Cities and Society 54 101882.

Ettema, D., Schwanen, T., 2012. A relational approach to analysing leisure travel. Journal of Transport Geography, 24, 173-181.

FDOT, Florida Department of Transportation, 2019. Speed or aggressive Driving, State Safety Office, Available at: https://www.fdot.gov/safety/2a-programs/aggressive-driving.shtm.

Faghih-Imani, A., Eluru, N., El-Geneidy, A., Rabbat, M., Haq, U, 2014. How land-use and urban form impact bicycle flows: evidence from the bicycle-sharing system (BIXI) in Montreal. Journal of Transport Geography, 306-314.

Fletcher, C., Garasky, S., Nielsen, R., 2005. Transportation hardship: Are you better off with a car? Journal of Family and Economic Issues, 26(3), 323–343

Fountas G., Anastasopoulos, P., Mannering, F., 2018a. Analysis of vehicle accident-injury severities: A comparison of segment- versus accident-based latent-class ordered probit models with class-probability functions. Analytic Methods in Accident Research 18, 15-32.

Fountas, G., Pantangi, S., Hulme, K., Anastasopoulos, P., 2019. The effects of driver fatigue, gender, and distracted driving on perceived and observed aggressive driving behavior: A correlated grouped random parameters bivariate probit approach. Analytic Methods in Accident Research 22, 100091.

Fountas, G., Sarwar, M., Anastasopoulos, P., Blatt, A., Majka, K., 2018b. Analysis of stationary and dynamic factors affecting highway accident occurrence: A dynamic correlated random parameters binary logit approach. Accident Analysis and Prevention 113, 330–340.

Frank, L., Andresen, M., Schmid, T., 2004. Obesity relation- ships with community design, physical activity, and time spent in cars. American Journal of Preventive Medicine 27(2), 87-96.

Franke, T. Krems, J., 2013. Understanding charging behaviour of electric vehicle users. Transportation Research Part F: Psychology and Behaviour, 21, 5–89.

Frei, A., Axhausen, K, 2008. Modelling the frequency of contacts in a shrunken world. Arbeitsberichte Verkehrs-und Raumplanung Volume 532.

Frenken, K., Schor, J., 2017. Putting the sharing economy into perspective. Environmental Innovation and Societal Transitions 23, 3-10.

Galovski, T., Blanchard, E., 2002.The effectiveness of a brief psychological intervention on court-referred and self-referred drivers. Behavior Research and Therapy 40 (12), 1385-1402.

Ge, Y., MacKenzie, D., 2018. Modeling vehicle choices and charging behavior of plug-in electric vehicle owners jointly using dynamic discrete choice model. Transportation Research Board 97th Annual Meeting. Washington, DC.

Ge, Y., MacKenzie, D., Keith, D., 2018. Gas anxiety and the charging choices of plug-in hybrid electric vehicle drivers. Transportation Research Part D: Transport and Environment, 64, 111-121.

Gebhart, K., Noland, R., 2014 The impact of weather conditions on bikeshare trips in Washington, D.C.. Springer Science+Business Media New York, 2014. 10.1007/s11116-014-9540-7.

Geweke, J., Keane, M., Runkle, D.. 1994. Alternative computational approaches to inference in the multinomial probit model. The Review of Economics and Statistics 76, 609–632

Glöss, M., McGregor, M., Brown, B., 2016. Designing for labour: Uber and the on- demand mobile workforce. CHI conference on human factors in computing systems, Santa Clara, California, USA, 7-12 May 2016, 1632–1643.

Gosling, S., Rentfrow, P., Swann, W., 2003. A very brief measure of the Big-Five personality domains. Journal of Research in Personality, 3, 504-528.

Greene, D. L., Kontou, E., Borlaug, B., Brooker, A., Muratori, M., 2020. Public charging infrastructure for plug-in electric vehicles: What is it worth?. Transportation Research Part D: Transport and Environment, 78, 102182.

Greene, W., Hensher, D., 2003. A latent class model for discrete choice analysis: Contrasts with mixed logit. Transportation Research Part B 37, 681– 698.

Hamari, J., Sjöklint, M., Ukkonen, A., 2015. The sharing economy: Why people participate in collaborative consumption. Association for Information Science and Technology 67(9), 2047-2059.

Hamed, M., Mannering, F., 1993. Modeling travelers' postwork activity involvement: toward a new methodology. Transportation Science 27(4), 381-394.

Harb, R., Radwan, E., Yan, X., Pande, A., Abdel-Aty, M., 2008. Freeway work-zone crash analysis and risk identification using multiple and conditional logistic regression. Journal Transportation Engineering 134 (5), 203-214.

Harder, K., Kinney, T., Bloomfield, J., 2008. Psychological and Roadway Correlates of Aggressive Driving Final Report. Center for Human Factors Systems Research and Design, University of Minnesota.

Hawlitschek, F., Teubner, T., Gimpel, H., 2018. Consumer motives for peer-to-peer sharing. Journal of Cleaner Production 204, 144-157.

Heckman, J., 1974. Shadow prices, market wages and labor supply. Econometrica 42(4), 679-694.

Heckman, J., 1979. Sample selection bias as a specification error. Econometrica 47(1), 153-161.

Hellwig, K., Morhart, F., Girardin, F., Hauser, M., 2015. Exploring different types of sharing: A proposed segmentation of the market for sharing businesses. Psychology and Marketing 32(9), 891-906.

Hennessy, D., Wiesenthal, D., 1999. Traffic congestion, driver stress, and driver aggression. Aggressive Behavior 25(6), 409-423.

Hensher, D.., Rose, J., Greene, W., 2015. Applied choice analysis: a primer (2nd ed.). Cambridge university press.

Heyman, J., Ariely, D., 2004. Effort for payment: A tale of two markets. Psychological science, 15(11), 787-793.

Hole, A., 2011. A discrete choice model with endogenous attribute attendance. Economics Letters, 110(3), 203-205.

Hossain, M., Muromachi, Y., 2012. A Bayesian network based framework for real-time crash prediction on the basic freeway segments of urban expressways. Accident Analysis and Prevention 45, 373-381.

Hughes, R., MacKenzie, D., 2016. Transportation network company wait times in Greater Seattle, and relationship to socioeconomic indicators. Journal of Transport Geography 56, 36–44.

Hyland, M., Hong, Z., Pinto, H., Chen, Y, 2018. Hybrid cluster-regression approach to model bikeshare station usage. Transportation Research Part A: Policy and Practice, 71-89.

Häuberer, J. 2011. Social capital theory. Towards a methodological foundation. Wiesbaden: VS Verlag.

Islam M., Alnawmasi, N., Mannering, F., 2019. Unobserved heterogeneity and temporal instability in the analysis work-zone crash-injury severities. Working paper, College of Engineering, University of South Florida.

Islam, M., Hernandez, S., 2013. Modeling injury outcomes of crashes involving heavy vehicles on Texas highways. Transportation Research Record: Journal of the Transportation Research Board 2338, 28-36.

Islam, M., Mannering, F. 2020. A temporal analysis of driver-injury severities in crashes involving aggressive and non-aggressive driving. Analytic Methods in Accident Research 100128.

Iversen, H., Rundmo, T., 2002. Personality, risky driving and accident involvement among Norwegian Drivers. Personality and Individual Differences 33(8), 1251-1263.

Jacoby, J., Berning, C., Dietvorst, T., 1977. What about disposition? Journal of Marketing, 41(2), 22–28.

James, L., Nahl, D., 2000. Road rage and aggressive driving: steeering clear of highway warfare. Prometheus Books, Amherst, NY.

Jin, S., Kong, H., Wu, R., Sui, D., 2018. Ridehailing, the sharing economy, and the future of cities. Cities 76, 96–104.

Joanisse, M., Gagnon, S., Voloaca, M., 2013. The impact of stereotype threat on the simulated driving performance of older drivers. Accident Analysis and Prevention 50, 530-538.

Jonah, B., 1997. Sensation seeking and risky driving: a review and synthesis of the literature. Accident Analysis and Prevention 29(5), 651-665.

Jovanovic, D., Lipovac, K., Stanojevic, P., Stanojevic, D. 2011. The effects of personality traits on driving-related anger and aggressive behavior in traffic among Serbian drivers. Transportation Research Part F 14(1), 43-53.

Joye, D., Sapin, M., Wolf, C., 2019. Measuring social networks and social resources: an exploratory ISSP survey around the world. (GESIS-Schriftenreihe, 22). Köln: GESIS - Leibniz-Institut für Sozialwissenschaften.

Kadushin, C. 2012, Understanding social networks: Theories, concepts, and findings. OUP USA.

Kaiser, S., Furian, G., Schlembach, C. 2016. Aggressive behavior in road traffic – findings from Austria, Transportation Research Procedia 14, 4384-4392.

Kaltenbrunner, A., Meza, R., Grivolla, J., Codina, J., Banchs, R, 2010. Urban cycles and mobility patterns: Exploring and predicting trends in a bicycle-based public transport system. Pervasive and Mobile Computing, 455-466.

Karlaftis, M., Golias, I., 2002. Effects of road geometry and traffic volumes on rural roadway accident rates. Accident Analysis and Prevention 34(3), 357-365.

Kaysi, I., Abbany, A., 2007. Modeling aggressive driver behavior at unsignalized intersections. Accident Analysis and Prevention 39(4), 671-678.

Kim, J., Rasouli, S., Timmermans, H., 2018. Social networks, social influence and activity-travel behaviour: a review of models and empirical evidence. Transport Reviews 38, 499-523.

Kim, J.-K., Ulfarsson, G., Kim, S., Shankar, V., 2013. Driver-injury severity in single-vehicle crashes in California: A mixed logit analysis of heterogeneity due to age and gender. Accident Analysis and Prevention 50, 1073-1081.

Kim, K., 2018. Investigation on the effects of weather and calendar events on bike-sharing according to the trip patterns of bike rentals of stations. Journal of Transport Geography, 309-320.

Kim, K., 2015. Can carsharing meet the mobility needs for the low-income neighborhoods? Lessons from carsharing usage patterns in New York City. Transportation Research Part A 77, 249–260.

Kim, S., Yang, D., Rasouli, S., Timmermans, H., 2017. Heterogeneous hazard model of PEV users charging intervals: Analysis of four year charging transactions data. Transportation Research Part C: Emerging Technologies, 82, 248-260.

Kodransky, M., Lewenstein M., 2014. Connecting low-income people to opportunity with shared mobility. Institute for Transportation and Development Policy, New York, NY.

Lang, P., Bradley, M., 2013. Appetitive and defensive motivation: Goal-directed or goal-determined? Emotion Review 5(3), 230-234.

Latinopoulos, C., Sivakumar, A. Polak, J., 2017. Modeling electric vehicle charging behavior: What is the relationship between charging location, driving distance, and range anxiety? Presented at 96th Annual Meeting of the Transportation Research Board, Washington, D.C. https://spiral.imperial.ac.uk/bitstream/10044/1/44474/2/17-05273.pdf.

Latinopoulos, C., Sivakumar, A., Polak, J. 2017. Response of electric vehicle drivers to dynamic pricing of parking and charging services: risky choice in early reservations. Transportation Research Part C: Emerging Technologies, 80, 175–189.

Latoski, S., Dunn, W., Wagenblast, B., Randall, J., Walker, M., 2003. Managing Travel for Planned Special Events. No. FHWA-OP-04-010. United States. Federal Highway Administration. Office of Transportation Management, 2003.

Lavieri, P., Bhat, C., Pendyala, R., Garikapati, V., 2016. Introducing latent psychological constructs in injury severity modeling: multivehicle and multioccupant approach. Transportation Research Record 2601, 110-118.

Lee, C., Coughlin, J., 2015. Older adults' adoption of technology: An integrated approach to identifying determinants and barriers. The Journal of Product Innovation Management 32(5), 747-759.

Lee, L., 1983. Generalized econometric models with selectivity. Econometrica 51(2), 507-512.

Li, X., Lord, D., Zhang, Y., Xie, Y., 2008. Predicting motor vehicle crashes using support vector machine models. Accident Analysis and Prevention 40(4), 1611-1618.

Li, Y., Bai, Y., 2008. Development of crash-severity-index models for the measurement of work zone risk levels. Accident Analysis and Prevention 40 (5), 1724-1731.

Li, Y., Bai, Y., 2009. Highway work zone risk factors and their impact on crash severity. Journal of Transportation Engineering 135 (10), 694-701.

Li, Y., Pickles, A., Savage, M., 2005. Social capital and social trust in Britain. European Sociological Review 21(2), 109-123.

Li, Z., Liu, P., Wang, W., Xu, C., 2011. Using support vector machine models for crash injury severity analysis. Accident Analysis and Prevention 45, 478-486.

Liaw, A., Wiener, M., 2002. Classification and regression by randomForest. R news 2(3), 18-22.

Lin, N., 2001. Building a network theory of social capital. In Social capital: Theory and Research (N. Lin, K. Cook, R. Burt, eds.), Transaction Publishers, 3-29.

Lin, N. and D. Ao, 2008. The invisible hand of social capital: an exploratory study. Social capital: An international research program, 107-132.

Lin, N., Fu, Y, Chen, C., 2013. eds. Social capital and its institutional contingency: A study of the United States, China and Taiwan. Routledge, 2013.

Lloyd, K., Auld, C., 2002. The role of leisure in determining quality of life: Issues of content and measurement. Social Indicators Research, 57, 43-71.

Lohse-Busch, H., Duoba, M., Rask, E., M. Meyer, 2012. Advanced powertrain research facility AVTA Nissan Leaf testing and analysis. Argonne National Laboratory.

Lohse-Busch, H., Duoba, M., Rask, E., Stutenberg, K., Gowri, V., Slezak, L., Anderson, D., 2013. Ambient temperature (20 F, 72 F and 95 F) impact on fuel and energy consumption for several conventional vehicles, hybrid and plug-in hybrid electric vehicles and battery electric vehicle (No. 2013-01-1462). SAE Technical Paper.

Loo, R., 1979. Role of primary personality factors in the perception of traffic signs and driver violations and accidents. Accident Analysis and Prevention 11(2), 125-127.

Lord, D., Mannering, F., 2010. The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives. Transportation Research Part A 44(5), 291-305.

Lyubomirsky, S., King, L., Diener, E., 2005. The benefits of frequent positive affect: Does happiness lead to success?. Psychological bulletin, 131, 803.

Mackett, R., 2014. The health implications of inequalities in travel. Journal of Transport and Health, 1(3), 202–209.

Maizlish, N., Linesch, N., Woodcock, J., 2017. Health and greenhouse gas mitigation benefits of ambitious expansion of cycling, walking, and transit in California. Journal of Transport and Health 6, 490-500.

Malyshkina, N., Mannering, F., 2009. Markov switching multinomial logit model: An application to accident-injury severities. Accident Analysis and Prevention 41(4), 829-838.

Maness, M., 2017. A theory of strong ties, weak ties, and activity behavior: leisure activity variety and frequency. Transportation Research Record: Journal of the Transportation Research Board, 2665:30-39.

Maness, M., 2017. Comparison of social capital measures from position generators and name generators in predicting activity selection. Transportation Research Part A: Policy and Practice, 2017. 106:374-395

Maness, M., Lin, Z., 2019. Free charging: Exploratory study of its impact on electric vehicle sales and energy. Transportation Research Record, 0361198119844966.

Maness, M., 2017. A theory of strong ties, weak ties, and activity behavior: Leisure activity variety and frequency. Transportation Research Record: Journal of the Transportation Research Board 2665, 30-39.

Mannering, F., 1986a. A note on endogenous variables in household vehicle utilization equations. Transportation Research Part B 20(1), 1-6.

Mannering, F., 1986b. Selectivity bias in models of discrete/continuous choice: An empirical analysis. Transportation Research Record 1085, 58-62.

Mannering, F., 2018. Temporal instability and the analysis of highway accident data. Analytic Methods in Accident Research 17, 1-13.

Mannering, F., Abu-Eisheh, S., Arnadottir, A., 1990. Dynamic traffic equilibrium with discrete/continuous econometric models. Transportation Science 24(2), 105-116.

Mannering, F., Bhat, C., 2014. Analytic methods in accident research: Methodological frontier and future directions. Analytic Methods in Accident Research 1, 1-22.

Mannering, F., Bhat, C., Shankar, V., Abdel-Aty, M., 2020. Big data, traditional data and the tradeoffs between prediction and causality in highway-safety analysis. Analytic Methods in Accident Research 25, 100113, 1-9.

Mannering, F., Hensher, D., 1987. Discrete/continuous econometric models and their application to transport analysis. Transport Reviews 7(3), 227-244.

Mannering, F., Shankar, V., Bhat, C., 2016. Unobserved heterogeneity and the statistical analysis of highway accident data. Analytic Methods in Accident Research 11, 1-16.

Mannering, F., Winston, C., 1985. A dynamic empirical analysis of household vehicle ownership and utilization. Rand Journal of Economics 16(2), 215-236.

Manski, C., 1995. Identification Problems in the Social Sciences, Harvard University Press, 194 pages.

Manski, C., 2009. Identification for Prediction and Decision. Harvard University Press, 368 pages.

Martin, C., 2016. The sharing economy: a pathway to sustainability or a nightmarish form of neoliberal capitalism? Ecological Economics 121(1), 149–159.

Martin, E., Shaheen, S., Lidicker, J., 2010. Impact of carsharing on household vehicle holdings. Transportation Research Record: Journal of the Transportation Research Board 2143, 150–158.

Maslow, A., 1943. A theory of human motivation. Psychological Review 50(4), 370-396.

Mattson, J., Godavarthy, R., 2017. Bike-share in Fargo, North Dakota: Keys to success and factors affecting ridership. Sustainable Cities and Society, 174-182.

Mazumder, H., Pokhare, B., 2019. Sexual violence on public transportation: A Threat to women's mobility in Bangladesh. Journal of Aggression, Maltreatment, and Trauma, 28(8), 1017-1019.

McFadden, D., 1981. Econometric Models for Probabilistic Choice. Structural Analysis of Discrete Data Using Econometric Applicatios. MIT Press, Cambridge, MA.

McFadden, D., Ruud, P., 1994. Estimation by simulation. Review of Economics and Statistics, 76, 591– 608.

McFadden, D., Train, K., 2000. Mixed MNL models for discrete response. Journal of Applied Econometrics 15(5), 447–470.

Menon, N., Barbour, N., Zhang, Y., Pinjari, A., Mannering, F., 2019. Shared autonomous vehicles and their potential impacts on household vehicle ownership: An exploratory empirical assessment. International Journal of Sustainable Transportation 13(2), 111-122.

Metro Bikeshare. Data. 2018 Metro Bikeshare, 2018. https://bikeshare.metro.net/about/data/. Accessed January 5, 2020.

Milton, J., Shankar, V., Mannering, F., 2008. Highway accident severities and the mixed logit model: An exploratory empirical analysis. Accident Analysis and Prevention 40(1), 260-266.

Morgan, A., Mannering, F., 2011. The effects of road-surface conditions, age, and gender on driver-injury severities. Accident Analysis and Prevention 43(5), 1852-1863.

Motoaki, Y., Shirk, M., 2017. Consumer behavioral adaption in EV fast charging through pricing. Energy Policy, 108, 178-183.

National Academies, 2013. Frontiers in massive data analysis. The National Academies Press, Washington, DC.

National Association of City Transportation Officials. Shared Micromobility in the U.S.:2018. NACTO, 2019. https://nacto.org/wp-content/uploads/2019/04/NACTO_Shared-Micromobility-in-2018_Web.pdf. Accessed May 3, 2020.

National Oceanic and Atmospheric Administration. Climate Data Online. Department of Commerce, 2020. https://www.ncdc.noaa.gov/cdo-web/. Accessed February 7, 2020.

Nesbit, S., Conger, J., 2012. Predicting aggressive driving behavior from anger and negative conditions, Transportation Research Part F 15, 710-718.

Neuman, T., Pfefer, R., Slack, K., Hardy, K., Raub, R., Lucke, R., Wark, R. 2003. National Cooperative Highway Research Program (NCHRP) Report 500 -- Guidance for Implementation of the AASHTO Strategic Highway Safety Plan Volume 1: A Guide for Addressing Aggressive-Driving Collisions provides strategies that can be employed to reduce the number of crashes due to aggressive driving behavior. Available at: http://onlinepubs.trb.org/onlinepubs/nchrp/nchrp_rpt_500v1.pdf

Nice Ride. System Data. 2020 Lyft, Inc, 2020. https://www.niceridemn.com/system-data. Accessed January 5, 2020.

Nicholas, M., Hall, D., Lutsey, N., 2019. Quantifying the electric vehicle charging infrastructure gap across U.S. markets. International Council on Clean Transportation. White paper.

Nicolau, J., Sellers, R., 2012. The free breakfast effect: an experimental approach to the zero price model in tourism. Journal of Travel Research, 51(3), 243-249.

Oh J., Shankar, V., 2011. Corridor safety modeling via segmentation: A selectivity bias perspective to crash count data. Lambert Academic Publishing, 84 pages.

Osman, M., Paleti, R., Mishra, S., 2018. Analysis of passenger car crash injury severity in different work zone configurations. Accident Analysis and Prevention 111, 161-172.

Ouimet, M., Pradhan, A., Simons-Morton, B., Divekar, G., Mehranian, H., Fisher, D., 2013. The effect of male teenage passengers on male teenage drivers: findings from a driving simulator study. Accident Analysis and Prevention 58, 132-139.

Oz, B., Ozkan, T., Lajunen, T., 2010. Professional and non-professional drivers' stress reactions and risky driving. Transportation Research Part F 13(1), 32-40.

Paleti, R., Eluru, N., Bhat, C., 2010. Examining the influence of aggressive driving behavior on driver injury severity in traffic crashes, Accident Analysis and Prevention 42(6), 1839-1854.

Pan, G., Liping, F., Thakali, L., 2017. Development of a global road safety performance function using deep neural networks. International Journal of Transportation Science and Technology 6(3), 159-173.

Pande A., Abdel-Aty M., 2006b. A comprehensive analysis of the relationship between real-time traffic surveillance data and rear-end crashes on freeways. Transportation Research Record 1953, 31-40.

Pande A., Abdel-Aty, M., 2006a. Assessment of freeway traffic parameters leading to lane-change related collisions. Accident Analysis and Prevention 38(5), 936-948.

Pannone, G., Betz, B, Reale, M., Thomas, J., 2017. Decomposing fuel economy and greenhouse gas regulatory standards in the energy conversion efficiency and tractive energy domain. SAE International Journal of Fuels and Lubricants 10(1), 202-216. https://doi.org/10.4271/2017-01-0897.

Parady, G., Katayama, G., Yamazaki, H., Yamanami,T., Takami, K., Harata, N., 2019. Analysis of social networks, social interactions, and out-of-home leisure activity generation: Evidence from Japan. Transportation, 46, 1-26.

Pereira, R., Schwanen, T., Banister, D., 2017. Distributive justice and equity in transportation, Transport Reviews 37 (2), 170-191.

Philip, H., Ozanne, L., Ballantine, P., 2015. Examining temporary disposition and acquisition in peer-to-peer renting. Journal of Marketing Management 31(11-12), 1310–1332.

Philippe, F., Vallerand, R., Richer, I., Vallières, E., Bergeron, J., 2009. Passion for driving and aggressive driving behavior: a look at their relationship. Journal of Applied Social Psychology 39(12), 3020-3043.

Pinjari, A., Bhat, C., Hensher, D., 2009. Residential self-selection effects in an activity time-use behavior model. Transportation Research Part B 43(7), 729-748.

Prieto, M., Baltas, G., Stan. V., 2017. Car sharing adoption intention in urban areas: What are the key sociodemographic drivers? Transportation Research Part A 101, 218-227.

Rayle, L., Dai, D., Chan, N., Cervero, R., Shaheen, S., 2016. Just a better taxi? A survey-based comparison of taxis, transit, and ridehailing services in San Francisco. Transport Policy 45, 168–178.

Rhodes, K., Kok, D., Sohoni, P., Perry, E., Kraska, M., Wallace, M., 2017. Estimation of the effects of auxiliary electrical loads on hybrid electric vehicle fuel economy (No. 2017-01-1155). SAE Technical Paper.

Riviere C., Lauret P., Manicom, J., Page Y., 2006. A Bayesian neural network approach to estimating the energy equivalent speed. Accident Analysis and Prevention 32(2), 248-259.

Rong, J., Mao, K., Ma, J., 2011.Effects of individual differences on driving behavior and traffic flow characteristics. Transportation Research Record 2248, 1-9.

Saelens, B., Sallis, J., Black, J., Chen, D., 2003. Neighborhood-based differences in physical activity: An environment scale evaluation. American Journal of Public Health 93(9), 1552-1558.

Sarwar, M., Anastasopoulos, P., Golshani, N., Hulme, K., 2017. Grouped random parameters bivariate probit analysis of perceived and observed aggressive driving behavior: a driving simulation study. Analytic Methods in Accident Research 13, 52-64.

Savolainen, P., Mannering, F., Lord, D., Quddus, M., 2011. The statistical analysis of crash-injury severities: A review and assessment of methodological alternatives. Accident Analysis and Prevention 43(5), 1666-1676.

Sánchez, T., Stolz, R., Ma, J., 2003. Moving to equity: Addressing inequitable effects of transportation policies on minorities. Cambridge, MA: The Civil Rights Project at Harvard University.

Seraneeprakarn, P., Huang, S., Shankar, V., Mannering, F., Venkataraman, N., Milton, J., 2017. Occupant injury severities in hybrid-vehicle involved crashes: A random parameters approach with heterogeneity in means and variances. Analytic Methods in Accident Research 15, 41-55.

Shaheed, M., Gkritza, K., 2014. A latent class analysis of single-vehicle motorcycle crash severity outcomes. Analytic Methods in Accident Research 2, 30-38.

Shaheed, M., Gkritza, K., 2014. A latent class analysis of single-vehicle motorcycle crash severity outcomes. Analytic Methods in Accident Research 2, 30-38.

Shaheen S., Cohen A., Zohdy, I., 2016. Shared mobility: Current practices and guiding principles. U.S. Department of Transportation. Report No. FHWA-HOP-16-022

Shaheen, S., Cohen, A., 2008. Worldwide carsharing growth: an international comparison. Transportation Research Record: Journal of the Transportation Research Board 1992, 81-89.

Shaheen, S., Mallery, M., Kingsley, K., 2012. Personal vehicle sharing services in North America. Research in Transportation Business and Management 3, 71-81.

Shaheen, S., Martin, E., Bansal, A., 2018. Peer-to-peer (P2P) carsharing: understanding early markets, social dynamics, and behavioral impacts. UC Berkley, Research Reports. DOI: 10.7922/G2FN14BD

Shampanier, K., Mazar, N., Ariely, D., 2007. Zero as a special price: The true value of free products. Marketing science, 26(6), 742-757.

Sharmeen, F., Arentze, T., Timmermans, H., 2014. Dynamics of face-to-face social interaction frequency: role of accessibility, urbanization, changes in geographical distance and path dependence. Journal of Transport Geography 34, 211-220.

Sheela, P., Mannering, F., 2020. The effect of information on changing opinions toward autonomous vehicle adoption: An exploratory analysis. International Journal of Sustainable Transportation, 14(6), 475-487.

Shen, J., Saijo, T., 2008. Reexamining the relations between socio-demographic characteristics and individual environmental concern: evidence from Shanghai data. Journal of Environmental Psychology 28(1), 42-50.

Shin, S., Shankar, V., 2013. Selection bias and heterogeneity in severity models: Some insights from an interstate analysis. Lambert Academic Publishing, 132 pages.

Shinar, D. 1998. Aggressive driving: the contribution of the drivers and the situation. Transportation Research, Part F 1(2), 137-160.

Shinar, D., Compton, R. 2004. Aggressive driving: an observational study of driver, vehicle and situational variables. Accident Analysis and Prevention 36(3), 429-437.

Society for Human Resource Management, 2020. SHRM Survey Findings: 2015 Holiday Schedules. SHRM, November 2014. https://www.shrm.org/hr-today/trends-and-forecasting/research-and-surveys/Pages/shrm-2015-holiday-schedules.aspx. Accessed June 15, 2020.

Spissu, E., Pinjari, A., Pendyala, R., Bhat, C., 2009. A copula-based joint multinomial discrete-continuous model of vehicle type choice and miles of travel. Transportation 36(4), 403-422.

Stanford, M., Greve, K., Boudreaux, J., Mathias, C., Brumbelow, J., 1996. Impulsiveness and risk-taking behavior: comparison of high-school and college students using the Barratt impulsiveness Scale. Personality and Individual Differences 21(6), 1075-1973.

Stauffacher, M., Schlich, R., Axhausen, K., Scholz, R., 2015. The diversity of travel behaviour: motives and social interactions in leisure time activities. Arbeitsberichte Verkehrs-und Raumplanung, Volume 328.

Stern, S., 1997. Simulation-based estimation. Journal of Economic Literature 35, 2006–2039.

Stuster, J. 2004. Aggressive driving enforcement: Evaluation of Two Demonstration Programs. NHTSA. DOT HS 809 707.

Sun, F., Chen, P., Jiao, J., 2018. Promoting public bike-sharing: A lesson from the unsuccessful Pronto system. Transportation Research Part D: Transport and Environment, 533-547.

Sun, J., Sun, J., 2015. A dynamic Bayesian network model for real-time crash prediction using traffic speed conditions data. Transportation Research Part C 54, 176-186.

Tarko, A., Anastasopoulos, P., Pérez-Zuriaga, A., 2011. Can education and enforcement affect behavior of car and truck drivers on urban freeways? 3rd International Conference on Road Safety and Simulation, Indianapolis, IN.

Tarko, A., Islam, M., Thomaz, J., 2011. Improving safety in high-speed work zones: A Super 70 study. Joint Transportation Research Porgram. Indiana Department of Transportation and Purdue University, West Lafayette, IN.

Tasca, L., 2000. A review of the literature on aggressive driving research. Aggressive Driving Issues Conference, Sponsored by the Ministry of Transportation of Ontario, Canada, in Cooperation with the Transportation Research Board. Available at: http://www.aggressive.drivers.com/papers/tasca/tasca.pdf.

Taylor, A., Dorn, L. 2006. Stress, fatigue, health, and risk of road traffic accidents among professional drivers: The contribution of physical inactivity. Public Health 27, 371-391.

Teunissen, T., Sarmiento, O., Zuidgeest, M., Brussel, M., 2015. Mapping equality in access: The case of Bogotá's sustainable transportation initiatives, International Journal of Sustainable Transportation 9 (7), 457-467.

Thomas, J., 2014. Drive cycle powertrain efficiencies and trends derived from EPA vehicle dynamometer results. SAE International Journal of Passenger Cars-Mechanical Systems, 7(2014-01-2562), 1374-1384.

Tinsley, H., Eldredge, B., 1995. Psychological benefits of leisure participation: A taxonomy of leisure activities based on their need-gratifying properties. Journal of Counseling Psychology, 42, 123-132.

Train, K., 2009. Discrete choice methods with simulation. Second edition. Cambridge University Press, New York, NY.

Turner, C., Layton, J., Simons, L., 1975. Naturalistic studies of aggressive behavior: aggressive stimuli, victim visibility, and horn honking. Journal of Personality and Social Psychology 31(6), 1098-1107.

Tussyadiah, I., 2016. Factors of satisfaction and intention to use peer-to-peer accommodation. Journal of Hospital Administration 55, 70–80.

Van den Berg, P., Arentze, T., Timmermans, H., 2012. A multilevel path analysis of contact frequency between social network members. Journal of Geographical Systems 14, 125-141.

Van den Berg, P., Arentze, T., Timmermans, H., 2015. A multilevel analysis of factors influencing local social interaction. Transportation, 42, 807-826.

Van den Berg, P., Arentze, T., Timmermans, H., 2010. Location-type choice for face-to-face social activities and its effect on travel behavior. Environment and Planning B: Planning and Design 37, 1057-1075.

Van den Berg, P., Arentze, T., Timmermans, H., 2009. Size and composition of ego-centered social networks and their effect on geographic distance and contact frequency. Transportation Research Record: Journal of the Transportation Research Board, 2135, 1-9.

Van der Gaag, M., Snijders. T., 2005. The Resource Generator: social capital quantification with concrete items. Social Networks, 27, 1-29.

Vanlaar, W., Simpson, H., Mayhew, D., Robertson, R., 2008. Aggressive driving: a survey of attitudes, opinions and behaviors. Journal of Safety Research 39(4), 375-381.

Vavrik, J. 1997. Personality and risk-taking: a brief report on adolescent male drivers. Journal of Adolescence 20(4), 461-465.

Venkataraman, N., Ulfarsson, G., Shankar, V., 2013. Random parameter models of interstate crash frequencies by severity, number of vehicles involved, collision and location type. Accident Analysis and Prevention 59, 309-318.

Venkataraman., N., Shankar., V., Ulfarsson., G., Depauch., D., 2014. A heterogeneity-in-means count model for evaluating the effects of interchange type on heterogeneous influences of interstate geometrics on crash frequencies. Accident Analysis and Prevention 2, 12-20.

Wang, J., Lindsey, G, 2019. Neighborhood socio-demographic characteristics and bike share member patterns of use. Journal of Transport Geography, 102475.

Wang, M., Mu, L., 2018. Spatial disparities of Uber accessibility: an exploratory analysis in Atlanta, USA. Computers, Environment and Urban Systems 67, 169–175.

Waseem, M., Ahmed, A., Saeed, T., 2019. Factors affecting motorcyclists' injury severities: An empirical assessment using random parameters logit model with heterogeneity in means and variances. Accident Analysis and Prevention 123, 12-19.

Washington, S., Karlaftis, M., Mannering, F., 2011. Statistical and econometric methods for transportation data analysis. Second Edition, Chapman and Hall/CRC, Boca Raton, FL.

Washington, S., Karlaftis, M., Mannering, F., Anastasopoulos, P., 2020. Statistical and econometric methods for transportation data analysis. Third Edition, CRC Press, Taylor and Francis Group, New York, NY.

Weather Underground. History. TWC Product and Technology LLC., 2014, 2020. https://www.wunderground.com/. Accessed February 7, 2020.

Wells-Parker, E., Ceminsky, J., Hallberg, V., Snow, R., Dunaway, G., Guiling, S., Williams, M., Anderson, B. 2002. An exploratory study of the relationship between road rage and crash experience in a representative sample of US drivers. Accident Analysis and Prevention 34(2), 271-278.

Wen, Y., MacKenzie, D., Keith, D., 2016. Modeling the charging choices of battery electric vehicle drivers by using stated preference data. Transportation Research Record, 2572, 47-55.

Wilhelms M., Henkel, S., Falk, T., 2017. To earn is not enough: A means-end analysis to uncover peer-providers' participation motives in peer-to-peer carsharing. Technological Forecasting and Social Change 125, 38–47.

Wolbertus, R., Gerzon, B., 2018). Improving electric vehicle charging station efficiency through pricing. Journal of Advanced Transportation, 4831951.

Wolbertus, R., Kroesen, M., van den Hoed, R., Chorus, C., 2018. Fully charged: An empirical study into the factors that influence connection times at EV-charging stations. Energy Policy 123, 1-7.

Wu, Z., Sharma, A., Mannering, F., Wang, S., 2013. Safety impacts of signal-warning flashers and speed control at high-speed signalized intersections. Accident Analysis and Prevention 54, 90-98.

Xie, Y., Lord, D., Zhang, Y., 2007. Predicting motor vehicle collisions using Bayesian neural networks: An empirical analysis. Accident Analysis and Prevention 39(5), 922-933.

Xiong, Y., Mannering, F., 2013. The heterogeneous effects of guardian supervision on adolescent driver-injury severities: A finite-mixture random-parameters approach. Transportation Research Part B 49, 39 -54.

Xiong, Y., Mannering, F., 2013. The heterogeneous effects of guardian supervision on adolescent driver-injury severities: A finite-mixture random-parameters approach. Transportation Research Part B 49, 39-54.

Xiong, Y., Tobias, J., Mannering, F., 2014. The analysis of vehicle crash injury-severity data: A Markov switching approach with road-segment heterogeneity. Transportation Research Part B 67, 109–128.

Xue, M., Biying, Y., Du, Y., Wang, B., Tang, B., Wei, Y., 2018. Possible emission reductions from ride-sourcing travel in a global megacity: The case of Beijing. Journal of Environment and Development 27(2), 156-185.

Yang, Y., Yao, E., Yang, Z., Zhang, R., 2016. Modeling the charging and route choice behavior of BEV drivers. Transportation Research Part C: Emerging Technologies, 65, 190-204.

Yasmin, S., Eluru, N., Bhat, C., Tay, R., 2014. A latent segmentation based generalized ordered logit model to examine factors influencing driver injury severity. Analytic Methods in Accident Research 1, 23-38.

Yee, A., Fairchild, H., Weizmann, F., Wyatt, G., 1993. Addressing psychology's problem with race. American Psychologist 48(11), 1132-1140.

Younes, H., Zou, Z., Wu, J., Baiocchi, G., 2020. Comparing the Temporal Determinants of Dockless Scooter-share and Station-based Bike-share in Washington, D.C. Transportation Research Part A: Policy and Practice, 2020, 208-320.

Yu, R., Abdel-Aty, M., 2013. Utilizing support vector machine in real-time crash evaluation. Accident Analysis and Prevention 51, 252-259.

Zhang, Y., Thomas, T., Brussel, M., and Maarseveen, M., 2017. Exploring the impact of built environment factors on the use of public bikes at bike stations: Case study in Zhongshan, China. Journal of Transport Geography 59-70.

Zhou, X., 2015. Understanding Spatiotemporal Patterns of Biking Behavior by Analyzing Massive Bike Sharing Data in Chicago. PLoS ONE, 10(1), e0137922.

Zoepf, S., MacKenzie, D., Keith, D., Chernicoff, W., 2013. Charging choices and fuel displacement in a large-scale demonstration of plug-in hybrid electric vehicles. Transportation Research Record: Journal of the Transportation Research Board, 2385, 1-10.