Boise State University
ScholarWorks

Economics Faculty Publications and Presentations   Department of Economics

12-1-2015

Determinants of Seat Belt Use: Regression Analysis with FARS Data Corrected for Self-Selection

Frank Goetzke
University of Louisville

Samia Islam
Boise State University

Post Refereed Manuscript:

Determinants of Seat Belt Use: Regression Analysis with FARS Data
Corrected for Self-selection

Frank Goetzke[a] and Samia Islam[b]

a: [Department of Urban and Public Affairs, University of Louisville, Louisville, KY 40208],

b: [Department of Economics, Boise State University, Boise, ID 83725-1620]

Abstract

We develop a methodology to use FARS data as an alternative to NOPUS in estimating seat belt usage. The advantages of using FARS over NOPUS are that (i) FARS is broader because it contains more variables relevant for policy analysis, (ii) FARS allows for easy multivariate regression analysis, and, finally, (iii) FARS data is more cost-effective. *Methodology*: We apply a binary logit model in our analysis to determine the likelihood of seat belt usage given various occupant, vehicle, and built environment characteristics. Using FARS data, we derive coefficient estimates for categories such as vehicle occupants' age and night time seat belt use that observational surveys like NOPUS cannot easily provide. *Results*: Our results indicate that policies should focus on passengers (as opposed to drivers), male and young vehicle occupants, and that law enforcement should focus on pick-up trucks, rural roads, and nights. We find evidence that primary seat belt laws are effective. *Conclusions*: Although this is primarily a methodological paper, we present and discuss our results in the context of public policy so that our findings are relevant for road safety practitioners, researchers, and policymakers.

Keywords: Seatbealt use, FARS, Peltzman effect, safety behavior

1   Introduction

While the effectiveness of seat belts is well-established in research (Blincoe, 1994; Carpenter & Stehr, 2008; Crandall, Olson & Sklar, 2001; Evans, 1986; Kahane, 2000; Klein & Walz, 1995; Levitt & Porter, 2001; McCartt & Northrup, 2004; Partyka, 1988; Partyka & Womble, 1989), U.S. seat belt usage rates have been relatively low compared to other developed nations (NHTSA, 2007)[1]. In 1997, the federal government set targets to increase seat belt usage from 68 percent in 1996 to 85 percent by 2000, and then 90 percent in 2005, both of which went unmet according to statistics from the annual National Occupant Protection Use Survey (NOPUS)[2]. According to the National Highway Traffic Safety Administration (NHTSA) of the U.S. Department of Transportation, over half of passenger vehicle occupants killed in traffic accidents in 2006 died unbuckled (NHTSA, 2007).

With the rise in federal funding for highway safety initiatives and awareness programs, and more stringent primary and secondary law enforcement at the state level, the failure to meet targeted usage rates is confounding. But more importantly, it points to a need for targeted policies to incentivize usage. Before we can design such policies it is critical to determine the factors that affect vehicle occupants' decision to wear a seat belt. In the past, seat belt effectiveness studies that used NOPUS data could not address many of these factors because of the lack of certain details in the NOPUS data.

---

[1] It must be noted here, however, that safety devices such as seat belts and air bags can cause injury, and even fatalities at times (Hollands, et al., 1996; Morris & Borja, 1998).

[2] NOPUS is an annual survey providing probability-based observed data on seat belt use in the United States conducted by the National Center for Statistics and Analysis of the NHTSA.

Research on seat belt usage typically utilizes one of two publicly available data sources: (a) National Occupant Protection Use Surveys (NOPUS), and (b) Fatality Analysis Reporting System (FARS). The National Center for Statistics and Analysis of the NHTSA consider NOPUS to be their most reliable data set tracking the trends in seat belt usage by motorists. However, the observational nature of NOPUS data not only subjects it to some limitations due to the probability of human error in the data collection, but also do lack reliable data on vehicle occupants' personal characteristics or night time travel behavior.

Fatality Analysis Reporting System (FARS) is the other available database for evaluating the usage rates of occupant's protection devices[3]. An advantage of using the FARS database over NOPUS is that it is more comprehensive in the reported variables, providing, for example, additional data for vehicle occupant characteristics, as well as night time data. However, one critical problem with FARS data is that it underestimates seat belts usage when compared to estimates obtained from observational data such as NOPUS[4] (Salzberg, et al., 2002) due to the nature of the reporting system. To be included in FARS, a crash must result in the death of a person (occupant of a vehicle or a non-occupant) within 30 days of the crash. Since it lists only those crashes where there is at least one fatality, there is a potential issue of sample selection given that an individual's seat belt use affects his or her probability of death, which in turn influences whether the crash is included in the data because of the correlation between seat belt use and fatality. It has been

---

[3] FARS is a national census of motor vehicle fatalities which has been collected since 1975 and contains information on over 989,451 motor vehicle fatalities within the 50 States, the District of Columbia, and Puerto Rico as well as over 100 different coded data elements that characterize the crash, the vehicle, and the people involved.

[4] NOPUS is not the only source of probability-based observational data. States conduct their own observational surveys annually. However, several states only survey passenger cars and omit light trucks (Kahane, 2000).

shown that such sample selection leads to biased regression coefficient estimates (Angrist & Krueger, 1999; Heckman, 1979; Heckman, et al., 1996). The extent of this sample selection bias becomes even more significant when we consider that only about 0.5 percent of motor vehicle crashes involve a fatality and in 90 percent of the incidents, there is just a single death (NHTSA, 1998). Had that death not occurred, the crash would not be included in the FARS data base. Therefore, empirically the impact of sample selection can be substantial, and failing to account for it leads to estimates that systematically understate seat belt usage. Previously, Salzberg, et al. (2002) investigated seat belt usage rates by comparing FARS data with observation surveys and concluded that unbelted occupants are over-represented in fatal collisions for two reasons: (a) because of a greater chance of involvement in potentially fatal collisions in the first place, and (b) because they are not afforded the protection of seat belts when a collision does occur. Their model focused on risk but they did not address the sample selection bias.

In our 2009 paper (Islam & Goetzke, 2009), we used an identification method developed by Levitt and Porter (2001)[5] to correct for the sample selection bias in FARS data in order to obtain a credible estimate of seat belt usage in the U.S. Curiously but rather insightfully, the sample selection problem that arises because of the exclusion of a majority of non-fatal crash statistics from the FARS data set is countered by further limiting the data. We used two different specifications to correct the sample selection bias in FARS data: (a) correction based on strict independence of seat belt choice (Model 1, details of which are described in Section 3.1 below), and (b) correction based on strict

---

5 Levitt and Porter (2001) use a simple but ingenious identification strategy that allows them to directly estimate the impact of seat belts and air bags on crash survival rates, despite sample selection in the data.

dependence of seat belt choice (Model 2, described in Section 3.2 below). By applying these corrections, we showed that the FARS database can be established as an acceptable and comparable alternative to the observational NOPUS data (Islam & Goetzke, 2009). In this paper, we extend on this previous work on correcting the sample selection bias evident in the FARS dataset, and using a multivariate regression analysis, address the following critical question: *What factors affect the decision to use seat belts*? This question is critical in designing effective policies. For example, are there particular socio-demographic groups that on which policy should focus? Do specific times of day or regions require heightened enforcement? These specific policy-relevant questions are precisely what the corrected FARS dataset allows us to answer due to the rich cross-sectional detail that it contains on each reported event.

While in Islam & Goetzke (2009) we presented a method to correct sample selection bias in FARS data, and then proceeded to conduct a univariate statistical analysis to show its applicability as an alternative to NOPUS data, in this paper we go further by presenting results from a multivariate regression analysis of different categories of vehicle occupants, times of day, regions and vehicle types on seat belt use. We begin our analysis with model specifications similar to our 2009 paper (described above). For consistency of methodology, we use the same 2006 data that we used in our earlier paper. We produce coefficient estimates for categories such as night time seat belt use that NOPUS cannot provide since road-side surveys are conducted during daylight hours. We develop a methodology to use FARS data as an alternative to NOPUS. The advantages of using FARS over NOPUS are that FARS data is (a) broader because it contains more variables relevant for policy analysis (e.g. night-time observations), (b)

5

better in quality (once corrected for sample selection bias) because multivariate regression analysis can be easily conducted, and (c) more cost-effective. Although this is primarily a methodological paper, we present and discuss our results in the context of public policy so that our findings become relevant for road safety practitioners, researchers and policy makers.

The remainder of this paper is organized as follows: Section 2 describes the methodology and data used for our analysis. Section 3 describes our model specifications and variables. Section 4 presents the results and a discussion of their policy implications. Section 5 presents a concluding summary of our findings and outlines possible future extensions of this paper. Tables and figures are included at the end of the paper.

2   Methodology and Data

For our analysis we use 2006 data from the Fatality Analysis Reporting system (FARS) compiled by the National Highway Traffic Safety Administration (NHTSA)[6]. We focus on the 2006 data to remain compatible with our previous analysis of NOPUS (Islam & Goetzke 2009). We define the following categories of vehicles and passenger characteristics for easy correspondence with the NOPUS data: Occupants wearing seat belts (shoulder belts or both lap and shoulder belts) and occupants without seat belts (only lap belts or no belts); front seat passengers (seating position code 11 and 13); Passenger cars (body type codes 1 – 12), vans & SUVs (body type codes 14 – 21. 'Day' is considered to be the hours between 8 a.m. to 6 p.m., following NHTSA survey times. The categories for expressways and surface streets are derived from the reported roadway

---

6   Source: http://www-fars.nhtsa.dot.gov/Main/index.aspx

functional class. We follow the classification for urban, suburban and rural roadways in combination with the city code.

We apply a binary logit model in our analysis to determine the likelihood of seat belt usage given various occupant, vehicle and built environment characteristics. In general, a logistic regression model with multiple predictor variables can be characterized as follows (Hosmer and Lemeshow, 2010):

$$logit(p) = log(p/(1\text{-}p)) = \beta_0 + \beta_1 X_1 + ... + \beta_k X_k$$

Each estimated coefficient is the expected change in the log odds, $log(p/(1\text{-}p))$, of the dependent binary choice variable, i.e., using seat belts as in our case, for a unit increase in the corresponding independent variable, holding the other predictor variables constant. When a binary outcome variable is modeled using logistic regression, the interpretation of the logistic regression coefficients becomes easier after transforming of the logistic regression coefficient into odds ratios. The odds ratio is the odds of an event occurring in one group to the odds of it happening in another group, or a sample based estimate of that ratio.[7] An odds ratio of 1 indicates that the condition or event under study is equally likely in both groups. Using the odds ratio, it is also possible to compare the regression results of the FARS data with the NOPUS survey findings.

There are several advantages to this method that allows researchers to use FARS data instead of NOPUS in their estimations. As mentioned earlier, FARS data contains more variables relevant for policies, and, with corrections for sample section bias, allows for multivariate regression analysis to be conducted, not to mention that it is relatively inexpensive to collect and maintain.

---

[7] If the probabilities of an event happening is p_1 (first group) and p_2 (second group), then the odds ratio is:
$(p_1/(1\text{-}p_1)) / (p_2/(1\text{-}p_2)) = p_1/q_1 / p_2/q_2 = p_1 q_2 / p_2 q_1$ where $q_1 = 1 - p_1$ and $q_2 = 1\text{-} p_2$.

3 Model

3.1 Model 1: Assuming independence of seat belt choice

Levitt and Porter (2001) established the fact that FARS data set oversamples people without a seat belt because the failure to use seat belts will result in a higher probability of death. The single selection criterion that a crash is reported in the FARS data base is the occurrence of at least one fatality. Since the selection criteria for fatality is correlated with seat belt use, the data set cannot be considered a random sample for the purpose of effectively analyzing seat belt use. However, if we assume independence of individual observations, which means that there are no 'social spillovers' in seat belt usage within the same vehicle, we can remove all records of the passengers/occupants who succumbed to a crash in all one-fatality accidents. The remaining observations then become a random sample, because all non-random observations, based on the selection criteria, are excluded. In other words, suppose, we have two vehicles involved in an accident: vehicle *a* and vehicle *b*. If there is a fatality in vehicle *a*, it is self-selected, but vehicle *b* being in the accident is random in the case that there is no fatality in vehicle *b*. Therefore, in Model 1, to eliminate sample selection bias, we remove just the fatality observation, not the entire vehicle. Crashes with more than one fatality can be seen as random, because if seat belt choice is independent, then from the perspective of one fatality, the other fatality(s) is/are random (not self-selected). With this specification, we have a sample of 36,584 observations. A detailed discussion of the coefficient estimates obtained from a logistic regression using this model specification is provided in Section 5.

3.2 Model 2: Assuming dependence of seat belt choice

What if we cannot assume independence of observations in a car? This is relevant because research shows that seat belt choice is not strictly independent in vehicles with more than one occupant (Cohen & Einav, 2003; Saibel & Salzberg, 1996; Nambisan & Vasudevan, 2007) and there exist what we refer to as 'social spillovers' of seat belt usage within a vehicle. In order to remove sample selection bias under this assumption, we now need to remove additional observations from the dataset to ensure a random sample. Since seat belt use is assumed to be dependent, the reported fatality is not only correlated with seat belt use, but also with the seat belt use of all the other occupants of the same vehicle. Therefore, if there is a single fatality in vehicle *a*, all the observations of the entire vehicle is removed. But if there is one fatality in each car in a collision between vehicle *a* and vehicle *b*, each fatality is random from the perspective of the other(s), and therefore, all the observations of both vehicles remain in the sample used for Model 2. We refer to this second specification as "dependence of seat belt choice". What remains in the final data set - with a sample size of 23,815 - are only the vehicle observations without fatalities as well as all vehicles from crashes where the fatalities occur in more than one vehicle[8]. The results for this specification are shown in the last column of Table 1. We must note here, however, that assuming dependence of seat belt choice may remove too many unbelted vehicle occupants from the sample due to their association with a fatality - an event which has a higher probability of occurring precisely when vehicle occupants are unrestrained.

4 Variables

---

[8] As already explained in section 2.1, if there is a fatality in each vehicle involved in a two-car crash, from the point of view of either vehicle, the fatalities in the other vehicle are random, i.e., not self-selected.

A coefficient estimate (Table 2) or an odds ratio (Table 3) corresponding to a particular variable is interpreted as the likelihood of wearing a seat belt compared to the omitted category. For example, for 'Passengers', the omitted category is 'Driver with passenger'. For 'Male', the omitted category is 'Female'. For 'Van', 'Minivan', 'SUV', and 'Pickup truck', the omitted category is 'Passenger Cars'. For 'Night', omitted category is 'Day', and 'Weekend night' isolates an additional effect above and beyond nights. For 'Expressway', the omitted category is all other 'Surface streets'. For 'Rural' and 'Suburban', omitted category is 'Urban'. For 'Primary seat belt law' and 'No seat belt law (NH)', omitted category is 'Secondary seat belt law'. For 'Northeast', 'Midwest' and 'West', the omitted category is 'South' combined. In Table 2, a positive coefficient indicates a higher likelihood, while a negative coefficient reflects a lower likelihood. In Table 3, the odds ratios are all positive and range between zero and positive infinity.

We have described the assumptions and specifications for our two models in section 3.1 and 3.2 above. Reality is likely to be somewhere between Model 1 and Model 2 because in Model 1 we don't remove enough observations (therefore, non-seat belt users are over-sampled), while in Model 2 we remove too many observations. The results of our two models, therefore, can be interpreted as the lower and upper bound estimates.

5   Results and Discussion

Table 2 presents the results of our binary logistic regressions. Overall, the analysis produces quite robust results. In both models, the regression coefficients have the same signs, except two coefficients ('Midwest' and 'Weekend night', which are insignificant in both models 1 and 2). Only two regression coefficients, 'Suburban' and 'Weekend night

& Age 16 - 25' change from statistically significant in one model to not significant in the other but this is primarily an artifact of our model specification (explained below).

According to Table 2, we find that the 'Passenger' coefficient estimate is negative in both models, i.e., passengers are less likely to wear seat belts than drivers. According to Table 3, the chance or likelihood of passengers wearing seat belts is between 57 percent (model 1) to 76.5 percent (model 2) of the chance of drivers wearing a seat belt, i.e., drivers are almost twice as likely to be wearing a seat belt compared to their passengers (reciprocal of the odds ratio. The odds ratio from the two models using FARS data is comparable to the odds ratio derived from the 2006 NOPUS data which reports the likelihood of passengers wearing a seat belt at 77.8 percent. Gender and age are both significant determinants of seat belt usage according to our findings. The coefficient estimates are negative for males, as well as 'Age 16 - 25', which is in line with our expectations and NOPUS estimates. Younger vehicle occupants are less risk averse and, therefore, less likely to use seat belts. According to the estimated odds ratios reported in Table 3, males are only between 59 (model 2) to 62 percent (model 1) as likely to wear a seat belt as females, and 'Age 16 - 25' are between 62 percent (model 1) to 75 percent (model 2) as likely to wear seat belts compared to the (omitted) age 25 – 70 group. These results are in line with the observational estimates from NOPUS as well as with our previous statistical estimates from corrected FARS data (Islam & Goetzke, 2009).

For people aged 70 or older, the coefficient is positive and significant implying that older drivers are more cautious and law-abiding and, therefore, between 45 percent (model 2) to 51 percent (model 1) more likely to wear a seat belt (Table 3) than the omitted reference group of middle-aged occupants (Age 25 – 70). This finding supports

11

several NHTSA studies reporting that vehicle occupants aged 70 or older are more likely to use seat belts more than any other age groups (NHTSA, 2009).

The coefficient estimates for 'Vans and SUVs' categories are insignificant in comparison to passenger cars. Two offsetting effects may be at work here: (a) heavier vehicles are perceived to be safer and, therefore, occupants are less likely to wear a restraint. This is referred to as the Peltzman effect[9], and, (b) SUV's and vans are typically family vehicles, and thus, a similar explanation as that for minivans would apply, counteracting the Peltzman effect.

As for the final type of vehicle, 'Pickup trucks', we find a negative and highly significant coefficient estimate. Pickup truck occupants are between 76 percent (model 2) to 77 percent (model 1) as likely to use a seat belt as occupants in other types of vehicles (Table 3). This result which supports both the NOPUS estimates as well as the previous analysis conducted by Islam & Goetzke (2009) can also be explained by the Peltzman effect. The reason why the coefficients in both models do not differ much is that the dependence assumption in Model 2 does not apply as strongly for this category of vehicle, since pick-up truck drivers are often without passengers.

In comparison to states with secondary seat belt laws, people in states with primary seat belt laws are 1.66 to 1.69 times more likely to wear seat belts compared to secondary law states. The coefficients are both positive and significant. The odds ratio from NOPUS data suggests that the likelihood of seat belt use in primary seat belt law states is 1.99 times more than states with secondary seat belt laws.

---

[9] Peltzman effect (Peltzman, 1975) "arises when people adjust their behavior to a regulation in ways that counteract the intended effect of the regulation. So, for example, when the government passes a seat belt law, some drivers may respond by driving less safely." (Peltzman, 2004).

The coefficients are highly positive and significant for expressways. Our estimates support the odds ratios derived from NOPUS data and suggest the possibility that being more conscious of the dangers of risky driving on high-speed, high- volume expressways, vehicle occupants are 1.64 to 1.66 times more likely to use seat belts on expressways (see Table 3) than on surface streets.

For 'Suburban' roads, the regression coefficient is positive, but only statistically significant in Model 1 which can be used to determine the odds ratios reported in Table 3. Vehicle occupants in suburban roads are is 1.1 times as likely to be using a seat belt compared to the omitted group, Urban. The explanation is revealed by our assumptions for the models. Because of the assumption of dependence, more people without seat belts have been removed from the data set in model 2. Thus, if suburban vehicle occupants are more likely to wear seat belts than their urban counterparts, more observations of occupants without seat belts have been removed from the sample of observations in model 2, reducing the relative difference between the two categories being compared. Next, we look at rural versus urban vehicle occupants. Coefficient estimates from both models are negative and significant, implying that rural occupants are less likely to use seat belts. In determining the magnitude (within a range) of this likelihood, we turn to the odds ratios as reported in Table 3. Vehicle occupants are between 71.5 percent (model 1) to 80 percent (model 2) as likely to use seat belts in rural areas compared to urban areas. Our odds ratios from the FARS dataset are comparable to the NOPUS odds ratio.

Turning our attention to regional differences next, we find that seat belt usage among Northeasterners are about 83 percent (model 2) to 90 percent (model 1) compared to the South, but Westerners are around 1.6 times as likely to use seat belts in comparison

to the South, regardless of which model we consider (see Table 3). The NOPUS odds ratio for seat belt use in the Northeast and Midwest are somewhat lower than our findings, however, the odds ratio for the West are comparable between NOPUS and our results.

Finally, we look at night time data, which NOPUS does not provide since observational seat belt surveys such as NOPUS are conducted during daylight hours. We are able to analyze night time behavior, however, using corrected FARS data. We find that the coefficients estimates from both models are negative and significant for 'Night'. Compared to the omitted reference category, 'Day', motor vehicle occupants are 72 percent (model 1) to 87 percent (model 2) as likely to use seat belts at night. The coefficient for 'Weekend' and the interaction term 'Weekend night' is not significant in either model. We also isolated the night time behavior of occupants between Age 16 – 25 as an interaction term and found that the coefficient estimate is negative in both models and significant in model 2, which assumes dependence in seat belt choice. We find that young drivers and vehicle occupants are less likely to use seat belts at night and their choice is correlated with the choice of other occupants in the vehicle. Low seat belt usage at night makes it a critical consideration for policy, especially for the youth (16 – 25) cohort.

All the results of the road functional class, geographic regions, type of seat belt laws and population density are in line with odds ratios from NOPUS data, our earlier study (Islam & Goetzke 2009), as well as Strine, et al. (2010) and Nelson, Bolen & Kresnow (1998), and furthermore, we could analyze night-time behavior using the corrected FARS data.

It is important to note that the differences in the estimates for some of the variables from model 1 and 2 are a result of the assumptions we used in our model specification. Although we did our analysis with two separate model specifications, we hypothesized that dependence of seat belt use choice (model 2) is a more robust assumption and closer to the true model. Results presented here confirm this hypothesis, given that odds ratios from model 2 are closer to NOPUS estimates than model 1, except, most notably for "Males" who may act more as independent agents. Now, in testing for dependency if seat belt use is dependent (model 2), the true model would use the model 1 dataset but include a 'social spillover' term to account for dependency, which would be the seat belt use of the vehicle companion (passenger for the driver and vice versa). However, this 'social spillover' variable will be endogenous which then needs to be adequately addressed. The focus of the current paper is to establish corrected FARS data as a feasible alternative to NOPUS. In a subsequent extension of this paper we will explore the issue of dependency in greater detail.

6   Conclusion

In Islam & Goetzke (2009) we presented a univariate statistical analysis of corrected FARS data to show that the dataset is overall comparable to the observational survey dataset, NOPUS. In this paper, we go further by presenting results from a multivariate regression analysis of different categories of vehicle occupants, times of day, regions and vehicle types. We produce coefficient estimates for additional categories such as night time seat belt use that NOPUS cannot provide since road-side surveys can only be conducted during daylight hours and interactions of nighttime with personal

characteristics. Thus, results presented here are very relevant for road safety practitioners, researchers and policy makers.

Although in this paper we develop a methodology to only determine lower and higher bound coefficient estimates with the corrected FARS data, we successfully replicate the findings of prior research that, for different sub-categories of vehicles, location, or occupant type, seat belt usage is correlated with the degree of risk-taking (Salzberg, et al. 2002), as well as the dependence of seat belt choice (Cohen & Einav, 2003; Saibel & Salzberg, 1996; Nambisan & Vasudevan, 2007, or the NOPUS survey results). In light of the interpretation based on the sampling process, it seems that sometimes Model 1 (e.g., for 'Suburban') and sometimes Model 2 (e.g., 'Weekend night') estimates are more realistic, which shows the trade-off between both approaches.

The main contribution of this methodological paper is to show that seat belt usage analysis using correct FARS data is a cost-effective alternative to the NOPUS survey. We consider our results especially important given the rich cross-sectional detail available in the disaggregated FARS data base, which extends its potential for applicability in addressing and answering a myriad of policy questions in safety research. Our results indicate that policies should focus on passengers, male and young (Age 16 - 25) vehicle occupants. Law enforcement should focus on pick-up trucks, rural roads, and nights, especially for the young on weekend nights. Our results also strongly corroborate earlier research findings supporting the effectiveness of primary seat belt laws[10] (Table 3).

Any discussion of possible policy measures to address a social issue has to consider that there is significant variation in the effectiveness of different policies as well

---

[10] For a detailed summary of existing research on the effectiveness of primary seat belt laws, see Cohen & Einav (2003).

as the obstacles to their successful implementation (see Dinh-Zarr, et al., 2001). We acknowledge that implementing policies that could successfully target specific vehicle types or drivers could be challenging, for example on grounds of privacy protection and civil rights, but the same insights may still be important for targeting specific/high-risk groups for educational purposes. Moreover, heightened enforcement could just prove prohibitively costly, especially on rural roads. Under those circumstances, our results can act as a guide for policy makers in finding alternative measures. For instance, one possible alternative to costly heightened enforcement on rural roads could be increased advertising. According to an NHTSA seat belt survey report published in 2000, awareness messages and campaigns have been found to have lasting impact on the targeted audience, especially younger audience: "[The NHTSA 2000] *survey found widespread public exposure to the crash dummy ads. More than eight-in-ten persons (83%) recalled seeing or hearing ads that used crash dummies. Among those who had seen the ads, 70% recalled that the message was to wear seat belts.* […] *The crash dummy advertisements appeared to be most successful in reaching younger target audiences. Recall of both the ads and their seat belt message was highest among the 21-to-24-year-olds (73%), and declined with each subsequent age group.*" (NHTSA, 2000)

Although predicted seat belt usage rates from the FARS data set, corrected for self-selection, compares very positively with NOPUS survey observations overall and therefore can be used as a credible data alternative in seat belt and road safety studies (Islam & Goetzke, 2009), it is important to note that FARS has some limitations. FARS data reporting is fairly accurate when there is a fatality in a vehicle from which data is being collected so that the officer can verify whether the occupant was belted or not. But

17

in the absence of fatality, it is difficult for the recording officer to confirm seat belt use since by the time they arrive on the scene the occupants may have exited the vehicle. In these cases the FARS seat belt data may biases, and, thus, be less reliable. Furthermore, fatal crashes are more likely to occur on highways with higher speeds, where it is reasonable to assume that more vehicle occupants choose to buckle up. These same people could be less likely to wear a seat belt when they are on urban or suburban roads. While we ignored this issue here, it could be potentially addressed by applying geographic weights based on road classification. In addition, the quality of the FARS data is only as good as the accuracy of the reporting, however, similar data quality issue would also apply to the current alternative, the NOPUS data. Future extensions of this paper will explore the question of dependency in seat belt choice, as well as these other issues.

Table 1: Descriptive statistics of model variables

	Model 1		Model 2	
	Mean	**STD**	**Mean**	**STD**
Male	0.650	0.477	0.660	0.474
Passenger	0.308	0.462	0.215	0.411
Age < 16	0.032	0.176	0.024	0.154
Age 16 – 25	0.283	0.451	0.230	0.421
Age ≥ 70	0.061	0.239	0.048	0.214
Vans & SUVs	0.239	0.426	0.226	0.418
Pickup truck	0.205	0.403	0.208	0.406
Primary seat belt law	0.644	0.479	0.647	0.478
Expressway	0.185	0.388	0.171	0.377
Rural	0.424	0.494	0.376	0.484
Suburban	0.113	0.317	0.124	0.329
Northeast	0.102	0.302	0.107	0.309
Midwest	0.180	0.384	0.178	0.383
West	0.234	0.424	0.230	0.421
Weekend	0.355	0.478	0.330	0.470
Night	0.537	0.499	0.517	0.500
Weekend night	0.209	0.407	0.185	0.389
Weekend night & age 16 – 25	0.079	0.270	0.056	0.230
Seat belt use assumption	independent		dependent	
Sample size	36,584		23,815	

Table 2: Binary logit models for seat belt use

	Model 1		Model 2	
Seat belt use assumption	independent		dependent	
Sample size	36,584		23,815	
Pseudo-R^2	0.129		0.059	
-2 Log Likelihood	35,833.24		18,567.98	
Constant term	1.811***	(0.047)	2.098***	(0.067)
Male	-0.480***	(0.029)	-0.534***	(0.044)
Passenger	-0.556***	(0.028)	-0.267***	(0.048)
Age < 16	0.070	(0.074)	0.135	(0.130)
Age 16 – 25	-0.474***	(0.033)	-0.291***	(0.049)
Age ≥ 70	0.412***	(0.066)	0.373**	(0.103)
Vans & SUVs	-0.031	(0.032)	-0.014	(0.048)
Pickup truck	-0.262***	(0.033)	-0.273***	(0.047)
Primary seat belt law	0.524***	(0.027)	0.508***	(0.039)
Expressway	0.468***	(0.037)	0.517***	(0.059)
Suburban	0.095**	(0.046)	0.044	(0.063)
Rural	-0.335***	(0.028)	-0.216***	(0.042)
Northeast	-0.113**	(0.045)	-0.186**	(0.063)
Midwest	0.033	(0.036)	-0.056	(0.051)
West	0.477***	(0.035)	0.496***	(0.053)
Weekend	-0.037	(0.043)	-0.064	(0.286)
Night	-0.325***	(0.033)	-0.144**	(0.047)
Weekend night	-0.091	(0.061)	0.052	(0.088)
Weekend night & age 16 – 25	-0.101	(0.062)	-0.213**	(0.100)

standard errors in parenthesis

*** $p < 0.01$

** $p < 0.05$

* $p < 0.1$

Table 3: Odds ratio for seat belt use models

	NOPUS	Model 1		Model 2	
Male	0.626	0.619	(0.585 – 0.655)	0.586	(0.537 – 0.639)
Passenger	0.778	0.574	(0.542 – 0.607)	0.765	(0.697 – 0.841)
Age < 16	1.152	1.074	(0.929 – 1.241)	1.145	(0.887 – 1.478)
Age 16 – 25	0.695	0.622	(0.583 – 0.664)	0.747	(0.678 – 0.824)
Age ≥ 70	1.244	1.510	(1.327 – 1.718)	1.452	(1.187 – 1.778)
Vans & SUVs	1.152	0.969	(0.909 – 1.033)	0.986	(0.897 – 1.084)
Pickup truck	0.625	0.770	(0.722 – 0.821)	0.761	(0.694 – 0.834)
Primary seat belt law	1.991	1.690	(1.602 – 1.783)	1.662	(1.538 – 1.795)
Expressway	1.898	1.597	(1.485 – 1.717)	1.678	(1.496 – 1.882)
Suburban	1.396	1.100	(1.005 – 1.204)	1.045	(0.923 – 1.184)
Rural	0.942	0.715	(0.676 – 0.757)	0.806	(0.742 – 0.875)
Northeast	0.583	0.893	(0.817 – 0.976)	0.830	(0.734 – 0.939)
Midwest	0.686	1.034	(0.964 – 1.109)	0.945	(0.855 – 1.045)
West	1.843	1.611	(1.504 – 1.726)	1.643	(1.479 – 1.825)
Weekend	1.069	0.964	(0.885 – 1.050)	0.937	(0.832 – 1.056)
Night	n/a	0.723	(0.677 – 0.772)	0.866	(0.789 – 0.949)
Weekend night	n/a	0.913	(0.810 – 1.029)	1.053	(0.885 – 1.253)
Weekend night & age 16 – 25	n/a	0.904	(0.800 – 1.021)	0.808	(0.664 – 0.984)

95% confidence interval in parenthesis.

References

Angrist, J. and Krueger, A. (1999) Empirical strategies in labor economics in Orley Ashenfelter and David Card (eds.), *Handbook of Labor Economics*, 3(23), Amsterdam: Elsevier.

Blincoe, L.J., (1994) Estimating the benefits from increased safety belt use. DOT HS 808-133, National Highway Traffic Safety Administration, Washington, DC.

Carpenter, C.S. & Stehr, M. (2008) The effects of mandatory seat belt laws on seat belt use, motor vehicle fatalities, and crash-related injuries among youths, *Journal of Health Economics*, May, 27(3), 642-662.

Cohen, A., and Einav, L., (2003) The effects of mandatory seat belt laws on driving behavior and traffic fatalities. Review of Economics and Statistics, 85 (4), 828–843.

Crandall, C.S., Olson, L.M., and Sklar, D.P., (2001) Mortality reduction with air bag and seat belt use in head-on passenger car collisions, *American Journal of Epidemiology*, 153(3), 219-24.

Dinh-Zarr, T.B., Sleet, D.A., Shults, R.A., Zaza, S., Elder, R.W., Nichols, J.L., et al., (2001) Reviews of evidence regarding interventions to increase the use of safety belts, *American Journal of Preventive Medicine*, 21 (4), 48–65.

Evans, L. (1986) The effectiveness of safety belts in preventing fatalities, *Accident Analysis & Prevention*, 18, 229–241.

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

Heckman, J., Ichimura, H., Smith, J. and Todd, P. (1996) Sources of selection bias in evaluating social problems: an interpretation of conventional measures and evidence on the effectiveness of matching as a program evaluation method, *Proceedings of the National Academy of Science of the Unites Stated of America,* 93, 13416-13420.

Hollands, C. M., Winston, F. K., Stafford, P. W., and Shochat, S. J., (1996) Severe head injury caused by airbag deployment, *Journal of Trauma,* 41(5) , 920–922.

Hosmer, D.W., Jr. and Lameshow, S. (2010), *Applied Logistic Regression*. Wiley Series in Probability and Statistics, New York: Wiley.

Islam, S. and Goetzke, F. (2009) Correcting Sample Selection in FARS data to Estimate Seat belt Use, *Journal of Safety Research*, 40 (5): 389-393.

Kahane, C.J., (2000) Fatality reduction by safety belts for front seat occupants of cars and light trucks: Updated and expanded estimates based on 1986–1999 FARS data, NHTSA Report 809 199, USDOT, Washington, DC.

Klein, T.M. and Walz, M.C., (1995) Research note: estimating lives saved by restraint use in potentially fatal crashes. National Highway Traffic Safety Administration, Washington, DC.

Levitt, S. and Porter, J. (2001) Sample Selection in the Estimation of Air Bag and Seat belt Effectiveness, *The Review of Economics and Statistics,* 83(4), 603-615.

McCartt, A.T. and Northrup, V.S., (2004) Factors related to seat belt use among fatally injured teenage drivers, *Journal of Safety Research*, 35 (1), 29-38.

Morris, M. S., and Borja, L., (1998) "Letter to the Editor," *Journal of Trauma: Injury, Infection, and Critical Care,* 44, 238–239.

Nambisan, S. S. and Vasudevan, V. (2007) Is seat belt usage by front seat passengers related to seat belt usage by their drivers? *Journal of Safety Research*, 38(5), 545-555.

National Highway Traffic Safety Administration, *Traffic Safety Facts 1997*, Washington DC: NHTSA, 1998.

_____, *Traffic Safety Facts 2006*, Washington DC: NHTSA, 2007.

_____, *Traffic Safety Facts 2008*, Washington DC: NHTSA, 2009.

National Highway Traffic Safety Administration, *Motor Vehicle Occupant Safety Survey: Volume 2, Seat Belt Report*, March, DOT HS 308 061, Traffic Safety Facts 20021997,

Washington DC: NHTSA, 2000. Accessed on May 31, 2012 at: http://www.nhtsa.gov/people/injury/research/safetysurvey/Chapter2.html

Nelson, D. E., Bolen, J., and Kresnow, M. (1998) Trends in safety belt use by demographic type of state safety belt law, 1987–1993. *American Journal of Public Health*, 88, 245–249.

Partyka, S.C. (1988) Lives saved by seat belts from 1983 through 1987. DOT HS 807 324. National Highway Traffic Safety Administration, Washington, DC.

Partyka S.C., and Womble K.B., (1989) Research notes: projected lives savings from greater belt use. National Highway Traffic Safety Administration, Washington, DC.

Peltzman, S. (1975) The Effects of automobile safety regulation. *The Journal of Political Economy*, 83(4): 677-726.

Peltzman, S. (2004) "Regulation and the Natural Progress of Opulence", a lecture by S. Peltzman at the American Enterprise Institute.

Saibel, C. and Salzberg P. (1996) The effect of seat belt use on injury severity in motor vehicle collisions: a paired-comparison study, Washington Traffic Safety Commission, Olympia, WA.

Salzberg, P., Yamada, A., Saibel, C. and Moffat. J. (2002) Predicting seat belt use in fatal motor vehicle crashes from observation surveys of belt use, *Accident Analysis & Prevention*, 34(2), 139-148.

Strine, T.W., Beck, L.F., Bolen, J., Okoro, C., Dhingre, S., and Balluz, L. (2010) Geographic and sociodemographic variation in self-reported seat belt use in the United States, *Accident Analysis & Prevention*, 42, 1066–1071.

Frank Goetzke is an associate professor in the Department of Urban and Public Affairs at the University of Louisville. He has published in the fields of transportation, specifically in social interactions in travel demand models

and traffic safety, as well as in urban and regional economics. Frank Goetzke received his Ph.D. in 2006 from West Virginia University.

Samia Islam is an associate professor at the Department of Economics at Boise State University. She has published in the fields of entrepreneurship, transportation, air quality, and transportation safety. Samia Islam received her Ph.D. in 2004 from West Virginia University.