NBER WORKING PAPER SERIES

SAMPLE SELECTION IN THE ESTIMATION OF AIR BAGS AND SEAT BELT EFFECTIVENESS

Steven D. Levitt
Jack Porter

Working Paper 7210
http://www.nber.org/papers/w7210

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
July 1999

We would like to thank Gary Chamberlain, Austan Goolsbee, and Lawrence Katz for helpful comments. All remaining errors are our own. Financial support of the National Science Foundation and the John Lee Fund are gratefully acknowledged. Previn Mankodi provided outstanding research assistance. All opinions expressed are those of the authors and not those of the National Bureau of Economic Research.

Sample Selection in the Estimation of

Air Bag and Seat Belt Effectiveness
Steven D. Levitt and Jack Porter
NBER Working Paper No. 7210
July 1999
JEL No. C20, R40

ABSTRACT

Measurement of seat belt and air bag effectiveness is complicated by the fact that systematic data are collected only for crashes in which a fatality occurs. These data suffer from sample selection since seat belt and air bag usage influences survival rates which in turn determine whether a crash is included in the sample. Past researchers either ignored sample selection or adopted indirect estimation methods subject to other important biases. We propose a simple, but novel, solution to the selection problem: limiting the sample to crashes in which someone in a *different* vehicle dies. Under relatively weak conditions, consistent estimates can be obtained from this restricted sample. Empirically, we find seat belts to be more effective in saving lives than most previous estimates. Air bags, however, appear to be less effective than generally thought. If our coefficients can be generalized to all crashes, the cost per life saved with seat belts is approximately $30,000, compared to $1.6 million for air bags.

Steven D. Levitt
Department of Economics
University of Chicago
1126 E. 59th St.
Chicago, IL 60637
and NBER
slevitt@midway.uchicago.edu

Jack Porter
Department of Economics
Harvard University
Cambridge, MA 02138
jporter@harvard.edu

I. Introduction

Over 40,000 Americans were killed in traffic accidents in 1997, making road fatalities the leading cause of death among those aged 6-34 in the United States. As large as this death toll is, Figure 1 demonstrates that the number of traffic deaths has declined dramatically over the last two decades from a peak of over 54,000 in 1972 to roughly 40,000 in recent years. The decline in fatalities per vehicle mile traveled is even more dramatic: the 1994 rate (1.7 fatalities per 100 million vehicle miles traveled) is less than one-third as high as the level in the late 1960's.

Increased seat belt usage (e.g., Graham et al. 1997, NHTSA 1984, Orsay et al. 1988) and proliferation of airbags (e.g., Graham et al. 1997, Lund and Ferguson 1995, NHTSA 1996) are two factors that have been identified as contributing to the declining death toll on the roads.[1] As of December 1995, every state except New Hampshire had some form of mandatory seat belt use law on the books. NHTSA (1996) estimates seat belt usage rates of 58-68 percent in recent years, up from 11 percent in 1980. Air bags, first available on passenger vehicles in the 1987, diffused quickly. By 1992, over half of all new cars sold were equipped with at least one air bag (NHTSA 1996). Dual air bags are mandatory under federal law in all 1998 model year passenger cars and all new 1999 model year light trucks. Even in advance of federal requirements, however, air bags have been standard equipment on almost all vehicles in recent years. Consumer demand for air bags appears strong, despite their cost ($410 for dual airbags according to Graham et al. (1997)). The annual cost of equipping all new vehicles with dual air bags is roughly $4 billion.

---

[1]Other factors that have been identified as contributing to reduced fatalities are reductions in drunk driving due to increased enforcement, changing attitudes, and changes in the minimum legal drinking age (Grossman et al 1993, Ruhm 1996, Saffer and Grossman 1987, Zobeck et al. 1987), lower speed limits (NHTSA 1989, although Lave and Elias 1997 disagree), child-safety seats (NHTSA 1998), and mandatory helmet laws for motorcycle riders (NHTSA 1996).

While there is little question that seat belts save lives, past estimates of the magnitude of the effect span a wide range. Robertson (1976), for instance, cites nineteen studies on belt effectiveness with estimates ranging from 8-86 percent.[2] NHTSA (1984) concludes that seat belts are 45-55 percent effective in reducing fatalities. Although different methodologies are sometimes used, the standard approach to measuring seat belt effectiveness is to identify a sample of crashes and compare outcomes among those with and without seat belts (e.g. Kaplan and Cowley 1991, Marine et al. 1994, Orsay et al. 1988, Robertson 1976). One important shortcoming of this approach, which we discuss at length below, is the possibility of sample selection. Typically, only particular kinds of crashes are included in a sample (for example crashes with fatalities or crashes involving injuries that require an ambulance). If seat belt usage reduces injury severity, then such sample selection will tend to bias downward the measured benefits of seat belts.

Although it is recognized that air bags pose risks of both injury (Hollands et al. 1996, Morris et al. 1998) and death (NHTSA 1996), recent research and government evaluations continue to support the effectiveness of air bags. NHTSA estimates that air bags saved over 3,000 lives between their introduction and September 1998. Graham et al. (1997) concludes that the cost-effectiveness of air bags is on par with other medical and public-health interventions.

Studies of air bags have generally used one of two empirical approaches. The first approach analyzes fatality rates per registered vehicle for vehicles with and without airbags (e.g. Lund and Ferguson 1995; Ferguson, Lund, and Greene 1995). An important potential weakness

[2] Not all of these studies focus on reductions in fatalities. It is generally agreed that seat belts are more effective in preventing death than in preventing injury.

of this approach is the endogeneity of vehicle choice. Faced with an option, those consumers who value safety most highly may disproportionately purchase vehicles with air bags.[3] This form of selection emerges clearly in the data. For example, in our data set, comparing drivers with and without air bags who are involved in fatal two-vehicle crashes, those with air bags are 19 percent less likely to be reported to be drinking by the police, 11 percent less likely to have a recent speeding ticket or accident, 14 percent less likely to have been convicted of driving while intoxicated or to have had a license suspension, 12 percent less likely to be male, and 13 percent less likely to be under the age of 25.[4] All of those characteristics (driving while intoxicated, bad past driving record, being young or male) are associated with greater likelihoods of fatal crash involvement (e.g. Fell and Nash 1989, Levitt and Porter 1999, NHTSA 1998). Thus, this approach is likely to exaggerate any causal impact of air bags on fatality reductions.

The second approach that is frequently used to measure air bag effectiveness involves comparing ratios of the number of deaths in frontal versus non-frontal collisions for vehicles with and without air bags. This approach is known as a "double pair comparison" and is a form of difference-in-difference estimator. Air bags are designed to protect occupants only in frontal crashes. If the ratio of deaths in frontal versus non-frontal crashes is lower for vehicles with air bags relative to those without air bags, this difference is interpreted as lives saved due to air bags.

---

[3] In theory, the opposite type of selection could also occur. The most dangerous drivers might disproportionately choose cars with air bags because they will obtain the greatest reduction in death rates. Empirically, however, this does not appear to be the case.

[4] In addition to selection of drivers based on observable characteristics such as sex, there is likely to be selection on unobservables as well. For instance, among men of a given driving age and similar previous driving records, those valuing safety most highly may be more likely to choose vehicles with air bags.

Using variations on this approach, researchers have found air bags to be 20-35 percent effective in reducing death rates in frontal crashes (Braver et al. 1997, Kahane 1996, NHTSA 1996, Zador and Ciccone 1993).

Because this analysis is based on ratios of *numbers* of deaths rather than crash survival rates, it relies critically on the assumption that the proportion of frontal crashes is constant for drivers with and without air bags. This assumption, however, appears to be false. As noted above, "safe" drivers (i.e. sober, female, over 25 and having a good past driving record) are disproportionately represented among those with air bags. "Safe" drivers, however, are more likely to "be hit" than to hit another vehicle, and thus are under-represented in frontal crashes. For instance, in two-car crashes in our data set, being sober is associated with a 72 percent reduction in the proportion of frontal to non-frontal crashes, wearing a seat belt, having a previous drunk driving conviction or license suspension, and being female are associated with 25, 33, and 28 percent reductions respectively. These observable driver characteristics which are related *not* to differential survival rates in frontal and non-frontal crashes, but rather to the frequency of such crashes, yield estimates similar to or larger than those for air bags.[5] Thus, it appears that the "double pairs" approach confounds differences in crash survival rates with the frequency with which frontal crashes occur, leading to exaggeratedly large estimates of the benefits of air bags.[6] Previous researchers have employed these indirect approaches to

[5] These lower instances of death in frontal crashes for safe drivers is not attributable to the fact that more of the safe drivers have air bags. Similar results are obtained when controlling for air bag status.

[6] On the other hand, if air bags provide benefits in non-frontal crashes (or there is classification error so that some frontal crashes are reported as non-frontal), this approach will underestimate the true effectiveness of air bags.

measuring the impact of seat belts and air bags due to an important data limitation. The most comprehensive data set is the Fatality Analysis Reporting System (FARS) which provides extensive information on all passengers in virtually all U.S. crashes, *but only in accidents where a death occurs*. The obvious problem in using these data to analyze the benefits of air bags and seat belts is sample selection. Seat belts and air bags influence the probability of death which in turn determines whether or not a crash is included in the data set.[7] As economists have long understood in other contexts, (e.g. Heckman 1979, Heckman et al. 1996, Heckman 1999), sample selection leads to biased estimation. Empirically, the extent of sample selection in the fatal crash data is enormous. Crashes with a fatality account for only 0.5 percent of all reported crashes and less than two percent of reported crashes with injuries (NHTSA 1998). Moreover, in almost 90 percent of fatal crashes, there is a single fatality. If that individual had not died, the crash would be excluded from the data set.

In this paper we use a simple, but novel, identification strategy that allows us to directly estimate the impact of seat belts and air bags on crash survival rates despite sample selection in the data. Sample selection arises because a given individual's seat belt usage affects his probability of death, which in turn influences whether the crash is included in the data. The key insight is that as long as *anyone else* dies in the crash, it is included in the fatal accident data regardless of what happens to others in the crash. We propose to restrict the sample to cases

[7] There are other data sets available, but the problems posed by these data sets are even more severe. The crashes included in these data are a representative sample of crashes reported to the police, not of all crashes. Crashes in which injuries occur are more likely to be reported, leading to sample selection if the probability and extent of injury is a function of seat belts and airbags. In addition, much less information is available for these accidents.

where anyone dies in another vehicle in the crash. By focusing on this subset of crashes, sample selection is eliminated under the conditions derived in Section II.[8] Ironically, the sample selection problem, which arises because observations are excluded from the data set, is overcome by further restricting the data that are used.

Table 1 demonstrates how sample selection distorts the data. The top row of the table presents the probabilities of death in the raw data as a function of the total number of individuals occupying all vehicles in the crash.[9] By definition, when there is only one vehicle occupant the probability of death is equal to one; if the individual did not die, the crash would be excluded from the data set. As the number of occupants increases, sample selection becomes decreasingly severe as evidenced by the steady decline in the probability of death. When five motorists are involved in a crash, each individual's probability of death is 26.9 percent. The falling death rates in the top row are *not* a consequence of decreasing crash severity as demonstrated by the bottom two rows of the table. Restricting the sample to cases in which *anyone else* dies in a crash results in a probability of death that is nearly constant at 15 percent as the number of occupants increases. Conditional on anyone dying in a different vehicle, individual death rates hover around 7 percent regardless of the number of total motorists.

Failing to account for sample selection leads to estimates that systematically understate the

---

[8] One could include any individual who is involved in a crash in which someone else dies, but the required identifying assumptions are difficult to justify within a vehicle due to the correlation of fatality outcomes even after controlling for observables and crash severity.

[9] Unlike later results which are restricted to two-vehicle crashes, these results are for all crashes regardless of the number of vehicles. Only crashes in which some vehicle occupant dies are included (there are some crashes in the data where the only fatality is a pedestrian, bicyclist, or motorcyclist).

benefits of effective life-saving devices. Part of the intuition for this result is that on average it takes a more severe crash to kill someone who is protected by a safety device. Since crash severity is not directly observed by the econometrician, the positive correlation between crash severity and the safety device is likely to lead to an underestimate of the device's value. Unobserved crash severity, however, is not the whole story. A downward bias persists even when crash severity is held constant.[10] Thus, the sample-selection problem is more pernicious than it might initially appear to be.

Empirically, using data on fatal crashes in the United States from 1994-97, we find that correcting for sample-selection bias as much as doubles the measured effectiveness of seat belts. We find seat belts to be more effective than has generally been the case in past research. Wearing a seat belt reduces the risk of death in our sample of crashes by well over fifty percent. Air bags are 16 percent effective in lowering death rates in direct frontal crashes; averaging over all types of crash impacts, air bags are roughly 9 percent effective. Our air bag estimates are lower than those typically obtained via the indirect approaches used by previous researchers.

The outline of this paper is as follows. Section II develops the theoretical model and identification strategy. Section III uses data on fatal crashes to estimate the impact of seat belts and airbags on probabilities of death and injury, correcting for sample selection. Section IV offers interpretation of the coefficients in terms of lives saved and cost-benefit ratios. Section V concludes.

---

[10] Evans (1986) attempts to correct for unobserved crash severity in seat belt effectiveness estimation.

II. Theoretical Model and Identification Strategy

Our goal is to estimate the effectiveness of seat belts and air bags in preventing death in motor vehicle crashes. In this section we summarize the intuition underlying our estimation approach. The appendix provides a rigorous derivation of our results in a fully general setting.

For ease of notation and exposition in presenting the intuition, we adopt four simplifying assumptions. First, we limit our attention to a single safety device (for concreteness we call this device a seat belt). Incorporating multiple safety devices (e.g. seat belts and air bags) is straightforward. Second, we focus on two-car crashes where only the two drivers and no passengers are involved.[11] Third, we impose linearity on the problem in what follows. Ad demonstrated in the appendix, linearity is in no way integral to identification of the effectiveness Finally, we assume that after conditioning on both observable factors of a particular driver, as well as both observable and unobservable characteristics of the crash, the other driver's characteristics do not help to explain the first driver's outcome. The model can be readily expanded to handle the more complicated case, but at a substantial notational burden.

Given these simplifying assumptions, the relevant model of crash survival is as follows:

[11] This simplification is purely for notational convenience, allowing us to avoid an additional layer of subscripts.

$$y_{jc} = \alpha + \beta s_{jc} + \gamma_{x'} x_{jc} + \gamma_h \eta_{jc} + \varepsilon_{jc} \quad \text{where } E[\varepsilon_{jc} | s_{jc}, x_{jc}, \eta_{jc}, s_{kc}, x_{kc}, \eta_{kc}] = 0 \tag{1}$$

and y_{jc} is an indicator variable as to whether individual *j* dies in crash *c*, s_{jc} is an indicator for seatbelt usage, x_{jc} is a vector of observed individual, vehicle, and crash characteristics, η_{jc} is *unobserved* crash severity, and ε_{jc} is a residual. Note that all of these characteristics, including unobserved crash severity, are allowed to vary across different individuals (denoted *j* and *k*) in the same crash.

It is useful to make a distinction between the unobservables η and ε. In our framework η captures all of the unobserved crash-related characteristics (e.g. the precise angle of impact, velocity at impact, distance from a hospital, etc.) and ε incorporates unobserved characteristics of the driver or vehicle (e.g. whether the driver has a heart condition, a defect in the way the steering column is attached, etc.). Another way of thinking about the distinction between η and ε is that ε reflects those factors that could in theory be characterized when a driver pulls out of the garage, whereas η is only determined when the crash actually occurs. Given these definitions, it is likely that for drivers *j* and *k* in a particular crash, η_{jc} and η_{kc} will be correlated, whereas it is plausible that $\mathring{a}_{jc}$ and $\mathring{a}_{kc}$ will be uncorrelated.[12]

The parameter $\hat{a}$ from (1) represents the decrease in the probability of dying associated with wearing a seatbelt. In what follows, we present conditions for interpreting this parameter causally. Ultimately, our focus will be on $\hat{a}$ normalized by the probability of dying without a seat belt (denoted p_0):

$$r \equiv -\, b \;/\; p_0 \tag{2}$$

This ratio r is the fraction of deaths that would have occurred without seat belts that are saved through their usage.

In practice, there are two important data limitations that preclude estimating equation (1) above. First, only crashes involving a fatality are observed. Second, crash severity h is unobservable. A naive attempt at estimating (1) in the available data would take the following form:

[12] In practice, we require only that the $\mathring{a}$ terms be uncorrelated conditional on all observed characteristics and crash severity.

$$y_{jc} = a + bS_{jc} + g_{x}' x_{jc} + l_{jc} \quad \text{for all crashes such that } y_{jc} + y_{kc} > 0$$

(3)

$$l_{jc} = g_{h} h_{jc} + e_{jc}$$

---

then sample selection will tend to understate the device's effectiveness.[13] The intuition for this result is that all of the cases in which a seat belt fails to prevent a death are included in the data set, but only a subset of the cases where a life is saved are recorded. If exactly one person would have died in a crash, but because of seat belt use the life is saved, then the crash is not included in the data set. By excluding this saved life from the sample, seat belts appear less effective than they actually are.

Sample selection bias can be eliminated by limiting the sample to the set of individuals who are involved in two-car crashes *in which someone in the other car dies.* Dependence between driver *j*'s outcome and inclusion in the sample must be eliminated. As long as driver *k* dies (or more generally someone in driver *k*'s vehicle), driver *j*'s outcome has no impact on whether the observation is included in the sample. Formally, we require conditional independence across the *å*'s for different drivers in the same crash:

$$(A1) \qquad e_{jc} \perp e_{jkc} \,|\, s_{jc}, s_{kc}, x_{jc}, x_{kc}, h_{jc}, h_{kc}$$

If after conditioning on the observed characteristics and the unobserved crash severity, the other driver's survival outcome is independent of own survival, then this sub-sampling method will eliminate the sample selection bias. By conditioning on the whole vector *(s,x,h)* we allow for many types of non-random mixing of drivers in crashes. For example, it does not pose a problem to our identification if trucks tend to travel on the same roads, drunks tend to hit safe drivers, or teenagers are more likely to drive at night. Here the fact that unobserved crash severity is

[13] The direction of the sample selection bias in *r* can be derived more formally by considering a set of crashes with *h* and *x* fixed. For this set of crashes, *r* has a probability limit of 1-C(*á+â)/á* where it can be shown that if *â<0* then C<1. Thus, the safety device effectiveness will be biased downward.

included in the conditioning set is crucial. If only observed crash characteristics were included, then it would be likely that survival outcomes would still be dependent, i.e. the other driver dying would be correlated with higher unobserved crash severity which would be correlated with a higher probability of death for the first driver.

Omitted variable bias

The lack of observability of *h* leads to a standard omitted-variable problem. If those who wear seat belts in the sample are on average involved in more (less) severe crashes, then the impact of seat belts will be understated (overstated). We require, therefore, that the following condition hold

$$(A2)\ \ h_j \perp s_{jc} | x_{jc}, y_{kc} = 1$$

Assumption A2 states that for the sub-sample of crashes with characteristics *x* and someone in the other vehicle dying, crash severity and seat belt use are independent.[14,15] Thus, although crash severity is unobserved, the survival probability is being estimated at the same average crash severity level for both seatbelt and non-seatbelt wearers. Consistent seatbelt effectiveness estimates are the result.

Conditional independence in *A2* may be violated in one of two ways. First, if drivers feel

---

[14] Limiting the sample to crashes where someone in the other vehicle dies is critical here. In the full sample of fatal crashes, assumption *A2* is unlikely to hold because it takes a harder crash on average to kill someone wearing a seat belt. In other words, there is an interaction between sample selection and omitted variable bias that distorts naive estimates even more than might initially be apparent. If the sample selection correction implied by *A1* is effective, this interaction is eliminated.

[15] In fact, the condition implies that for the subsample with characteristics *x* the *distribution* of crash severity is the same for both seatbelt wearing and non-seatbelt wearing drivers.

safer when wearing seat belts, this may induce them to drive more recklessly. This may lead to an increased number of crashes for those wearing seat belts (Peltzman 1975), as well as more severe accidents on average. While the "Peltzman effect" itself (more crashes) does not pose a problem for consistent estimation, if seat belt wearers have more severe crashes on average then the estimation of the causal impact of seat belts conditional on a crash will be biased downward.[16] On the other hand, one might expect safety conscious people to both wear a seatbelt and to drive more cautiously, resulting in a negative correlation between seatbelt status and crash severity. However, note that it is *conditional* independence that is assumed in *A2*. For example if we condition on safety consciousness, then the assumption becomes more believable. Of course, safety consciousness is not directly observable. But age, gender, and past driving record are among the variables observed, as are vehicle type, time of day, urban vs. rural, and speed limit. These characteristics might reasonably control for such factors. The more such characteristics are observed, the more believable the conditional independence assumption becomes, justifying the "kitchen sink"-type approach we employ.[17]

It must be noted, however, that there is an important limitation to the approach that we

---

[16] In the presence of a Peltzman-type moral hazard problem, seat belt wearing influences the number of crashes, so from a public policy perspective, we might be interested in more than the impact of a seat belt conditional on a crash.

[17] If assumptions *A1* and *A2* hold, then we obtain consistent estimates of the impact of safety devices on survival rates. In order to obtain a causal interpretation of the parameters, safety device usage must be "as good as randomly assigned" (see Heckman 1999, Imbens and Angrist 1994):

$$(A3)\quad s_{jc} \perp e_{jc} | x_{jc}, h_{jc}$$

i.e. conditional on the other covariates, seat belt usage must be independent of the error term in equation (1). This independence is a stronger version of the standard orthogonality between the error term and regressor.

adopt. The best we can do, given that *h* is unobserved, is to estimate the mean seat-belt effectiveness conditional on a given vector of characteristics *x* and averaged over *h* in the sample. Note, however, that the subset of crashes in which someone in the other vehicle dies may have a distribution of crash severity that is very different from the universe of crashes. How well safety devices work is likely to be a function of crash severity. For example, tn the most gentle crashes, even unbelted occupants will not die, so seat belts are completely ineffective (with respect to preventing fatalities); in the most severe crashes (e.g. a convertible plunging off a cliff), seat belts may also be useless. Thus, the estimate we obtain on seat-belt effectiveness may not be readily generalizable outside our sample.

III. Estimating the Impact of Air bags and Seat belts

We estimate the model of the previous section using FARS data for fatal crashes occurring in the years 1994-1997. These data contain detailed information on virtually every crash in the United States in which a fatality occurs among either vehicle occupants or pedestrians. The sample of vehicle occupants actually used in the estimation that follows is limited in a number of ways. First, we exclude back-seat passengers, both because of low death rates for such passengers and the absence of air bags. Second, we exclude children in car seats. The risks posed by air bag deployment when car seats are placed in the front seat are well established (NHTSA 1996). As this fact became public knowledge, the prevalence of car seats placed in the front seat declined dramatically. In terms of predicting future effectiveness of air bags, excluding car seats is likely to provide a more accurate measure. Third, the sample is limited to two-vehicle crashes. One-vehicle crashes are not included because the sample-selection correction requires a fatality in another vehicle. Crashes with three or more vehicles (seven percent of all fatal crashes) are not included because of concerns that crash severity may vary dramatically across vehicles in a manner that is difficult to control for. Fourth, crashes in which the only fatalities are motorcyclists, bicyclists, or pedestrians are also dropped since such accidents are likely to pose little risk to vehicle occupants. Fifth, we exclude occupants of large trucks from the sample as well as vehicles of model years prior to 1991. In both cases, air bag installation is extremely low. Note, however, that if a newer vehicle crashes into a large truck or an older vehicle, the occupants of the newer vehicle are included in the sample. None of the basic results are sensitive to these exclusions.

There are serious data quality concerns in the FARS data, especially for the variables measuring seat belt and air bag status. Eight percent of the individual-level observations are missing information as to whether a seat belt was worn. Those individuals are excluded from the sample. As we later report, the results are not particularly sensitive to their inclusion under different sets of assumptions. There is also substantial measurement error in reporting the presence of air bags in the data set. For makes and model years where air bags are standard equipment, air bags are often reported as present less than 50 percent of the time. On the other hand, when air bags are not available, false positives are occasionally reported. In order to address this problem, we code all vehicles of a given make, model, year, and seating position (i.e. driver or front-seat passenger) as having air bags if more than ten percent of the observations report the presence of an air bag. In order to increase the reliability of this calculation, we limit the sample to makes and model years for which there are at least 20 vehicles involved in fatal crashes in our data (this excludes about 10 percent of the observations in the sample).[18] After we correct the air bag data in this way, the fraction of vehicles equipped with air bags by model year and seat position correspond closely to those reported in NHTSA (1996).

Our baseline sample after these exclusions includes about 40,000 individuals. Means and standard deviations for this sample are reported in columns 1 and 2 of Table 2. A more complete description of variables as well as information about the way they were constructed is available from the authors.

As the model in the preceding section demonstrates, it is necessary to correct for sample selection. We do so by further restricting the data set to occupants of vehicles in which anyone in the other vehicle dies in the crash. Summary statistics for the selection-corrected samples are presented in the remaining columns of Table 2. This restricted sample has roughly 20,000 observations. There are a few important differences between the full sample and the selection-corrected sample. First, death rates in the full sample are four times higher. This is primarily due to the artificial inflation of death rates due to sample selection. Second, passengers in vehicles with frontal impact and in larger vehicles are over-represented in the restricted sample. Vehicles that strike other vehicles head-on are more likely to inflict fatalities, as are larger vehicles. Since the effectiveness of seat belts and air bags may be a function both of vehicle size/type and the manner of impact (e.g. air bags are designed to protect occupants only in frontal crashes), we present separate estimates for automobiles and other vehicles, and divide crash impacts into

---

[18] Our results are not sensitive to either the choice of twenty or more crashes or ten percent air bag presence. Nearly identical results are obtained when the crash cutoff is raised to 100 or an air bag cutoff of thirty percent is used. For the air bag data, there is a clear demarcation in the data, with many makes and models very close to zero and most others at or above forty percent.

direct-frontal, partial-frontal, and non-frontal collisions. A direct-frontal impact is one in which the principal vehicle impact is a twelve on the clock face. A partial-frontal impact is one in which either the initial or principal vehicle impact is between ten and two on the clock face, excluding those crashes that qualify as directly frontal.

Before turning to the regression analysis, we first present results based on the raw data without any other control variables in Table 3. The values in the table are the fraction of lives saved/lost through the use of seat belts and air bags relative to passengers who are not wearing seat belts and do not have air bags, i.e. the parameter r defined in equation (2). The first three columns are estimates based on the full sample; the last three columns correct for sample selection by restricting the sample to cases where someone dies in the other vehicle. The table also reports the probability of death for occupants with neither seat belts nor air bags.

Focusing first on automobiles in the top panel of Table 3, in the first three columns that ignore sample selection, death rates for those with neither safety device are between sixty and seventy percent. Seat belts reduce death rates by 20-35 percent. As expected, air bags are most effective in head on collisions, reducing death rates by 22 percent in direct frontal crashes, compared to 10 percent in partial frontal crashes, and a small *negative* impact in non-frontal crashes. Controlling for sample selection (columns 4-6) roughly doubles the estimated effectiveness of seat belts and slightly reduces the impact of air bags.

For larger vehicles (the second panel of Table 3), seat belts appear even more effective. Correcting for sample selection once again increases the impact of seat belts. Air bags, however, look less effective in larger vehicles relative to autos, particularly in frontal crashes. Seat belts appear to be ten times more effective than air bags in saving lives in larger vehicles.

There are many reasons why the simple analysis of Table 3 may yield misleading results. For instance, seat belts usage varies across individuals along dimensions that also affect the likelihood of surviving crashes (e.g. by sex or age). Since air bags are not randomly distributed across vehicles (they are more common in newer and more expensive vehicles), the raw tabulations may not accurately reflect the causal impact of air bags. To explore these possibilities, we run regressions of the following form:

$$y_{jvc} = a + b_1 Seat\_belt_{jvc} + b_2 Air\_bag_{jvc} + X_{jvc}\Gamma + V_{vc}\Theta + Z_c\Lambda + e_{jvc} \qquad (4)$$

where *j* indexes individual vehicle occupants, *v* corresponds to a given vehicle, and *c* reflects a particular crash. The dependent variable *y* is an indicator variable equal to one if the occupant is killed and zero otherwise. In addition to the seat belt and air bag measures, we include a vector of individual-level characteristics *X* (age, sex, seat position), vehicle-level controls *V* for the driver's vehicle and/or the other vehicle in the crash (the type of vehicle that is crashed into, vehicle weight, measures of the weight differential between the vehicles and the squared weight differential, model year, driver's past driving record, other driver's past driving record), and crash-level factors *Z* (year of crash, a speed limit indicator, urban vs. rural, time of day indicators).[19] All of the results presented are based on linear probability models with robust standard error corrections for heteroskedasticity and within-vehicle correlation. Probit and logit estimation yield similar results.

[19] We have also experimented with including an interaction between seat belts and air bags that would allow a differential impact of air bags depending on whether a seat belt is worn. No systematic patterns emerged, although the interaction term was frequently statistically significant. The estimated overall impact of seat belts and air bags is not affected by the inclusion of the interaction, therefore we have omitted the interaction term for the sake of simplicity of exposition.

Table 4 presents the regression results. The structure of the table mirrors that of the earlier table. Only the seat belt and air bag coefficients are reported (the other variables in the regression are discussed below). In spite of the extensive list of controls included in the estimation, the measured effectiveness of seat belts and air bags is largely unchanged from the simple summary statistics presented in Table 3. Seat belts again appear between 50 and 75 percent effective in saving lives once sample-selection bias is eliminated. Air bags are 16-17 percent effective in direct frontal crashes, and 6-13 percent effective (though not statistically different than zero) in partial frontal crashes. Although air bags appear to have a large negative impact in non-frontal crashes in the corrected sample, the coefficients are imprecisely estimated due to the small number of non-frontal crashes and cannot be distinguished from zero.

The other covariates in the regressions underlying Table 4 are plausibly estimated. Full regression results corresponding to column 4 of Table 4 are reported in Table A-1 in the appendix. Full results for all specifications are available on request from the authors.[20] Drivers are two to three percentage points more likely to die than right-front-seat passengers. Gender is not an important predictor, but the probability of death is an increasing function of age. Even controlling for the weight differential of the vehicles involved in the crash (occupants of heavier vehicles fare better), survival probabilities are greater when crashing into an automobile than a utility vehicle, van, or small truck. The chance of death is higher during the night, in rural areas, and on roads with posted speed limits of 55 miles per hour or more. Bad previous driving records, both for your own driver and for the driver of the other vehicle are associated with

[20]The signs and magnitudes of the covariates are generally similar across different types of crash impacts. The air bag coefficient is the only one that is highly sensitive to crash type.

slightly higher death rates. There does not appear to be a systematic trend in death rates for later model-year vehicles once seat belts and air bags are controlled for, suggesting that other vehicle safety design innovations between 1990 and 1997 have not had a dramatic impact on crash survival.

Returning to the effect of seat belts and air bags, Table 5 presents sensitivity analysis from a range of alternative specifications. Only selection-corrected specifications are included in Table 5. Non-frontal crashes, which are imprecisely estimated, are not shown. Results are once again broken down by automobiles versus other vehicles and by direct and partial frontal crashes. The top row of the table shows the baseline estimates taken from columns 4 and 5 of Table 4. The next two rows add indicator variables to soak up potential unobserved heterogeneity. Adding make and model dummies to the regressions has little impact on the results. Including vehicle-fixed effects reduces the measured effectiveness of both seat belts and air bags somewhat in most specifications. The precision of the estimates decreases substantially since over seventy percent of the vehicles in the sample have only a single passenger and thus provide no information when vehicle-fixed effects are included.

The next seven rows of Table 5 focus on subsets of the data in order to isolate potential differences in safety device effectiveness across population sub-groups or types of crashes. Comparing the results for drivers and front-seat passengers, seat belts look equally effective, but air bags appear more beneficial to passengers than drivers in three of the four specifications. This result appears plausible in light of the fact that the air bag deploys from the dashboard rather than the steering wheel for front-seat passengers. The greater distance between the air bag and the occupant both increases the likelihood that the air bag will have inflated in time to be of use, and lessens the chance that the occupant will be close to the explosion that triggers the air bag with potential negative consequences. Our results also provide some suggestive evidence that air bags are more effective in protecting men than women. This is true in three of the four specifications, but in the remaining specification the pattern is reversed. This result would be consistent with existing evidence that air bags pose the greatest danger to people of small stature. Excluding those under the age of sixteen does not substantially change the results, but it is important to bear in mind that children in car seats are excluded from our baseline sample. There does not appear to be a strong relationship between safety device effectiveness and the posted speed limit or as a function of the weight differential between the vehicles involved in the crash.

As noted earlier, eight percent of the individuals have missing values for seat belt usage. In the results presented thus far, these observations were dropped. The bottom two rows of the table present results if the missing values are included under the assumption that all of these individuals were wearing seat belts, or that none were wearing seat belts. The estimated seat-belt effectiveness falls only slightly under each of these assumptions. The air bag coefficients are not greatly affected.

It is useful to compare the magnitude of our estimates for seat belts and air bags to previous values in the literature. The commonly accepted range of seat belt effectiveness in reducing fatalities is 45-50 percent (Evans 1986, Graham et al. 1997, Kahane 1996, NHTSA 1996). Our estimates correcting for sample selection range from 50-76 percent. Thus, we find seat belts to be substantially more effective than previous estimates. This result does not appear to be an artifact of the sample we use. Ignoring sample selection, we actually obtain lower benefits of seat belts than previous studies.

Our estimates for air bags, however, are lower than previous values. We find air bags to be roughly 16-17 percent effective in direct frontal crashes. In comparison, NHTSA (1996) reports a value of 31 percent and Zador and Ciccone (1993) finds a 28 percent reduction. Braver et al. (1997) find a 20 percent effectiveness for right-front passengers when excluding those under the age of 10 (which makes their sample comparable to ours). As with seat belts, the difference between our results and previous research is not due to a difference in samples. When we apply the standard "double-pair comparison" methodology to our sample of crashes, we obtain estimates of air bag effectiveness of 42 percent in direct-frontal crashes – *higher* than any of the studies cited above. Extrapolating our coefficients to all fatal crashes, taking into account the relative frequencies of frontal and non-frontal crashes, we obtain an overall air bag effectiveness of 9 percent. Graham et al. (1997) surveys the literature and concludes that 13 percent is the best available estimate for adults based on the previous literature – an estimate that is forty percent higher than ours.

IV. Calculating the number of lives saved by seat belts and air bags

In terms of understanding the value of seat belts and air bags, it is useful to translate the effectiveness estimates into numbers of lives saved. Calculations of lives saved do not require estimates of the fraction of drivers wearing seat belts (or driving vehicles equipped with air bags), but rather depend solely on r reported in the tables and the number of actual deaths with and without the safety device.[21] Let D_S and D_O equal respectively the number of individuals who die

[21] As a sidelight, we note that it is possible given r and the number of deaths with and without seat belts to back out an estimate of the fraction of front-seat passengers wearing a seat belt. That fraction is equal to $(D_S/(1-r))/(D_S/(1-r))+D_0)$. Based on *all* front-seat occupants dying

with and without a given safety device. The number of lives saved as a result of the current level of usage of that safety device is equal to $(r/(1-r))D_S$. The number of cases where seat belts failed to save lives provides the key to how many lives they actually saved, for a given *r*. Using 0.60 as an overall estimate of seat belt effectiveness and extrapolating our estimates out of sample to the set of all fatal crashes, we calculate that slightly more than 15,000 lives of front-seat occupants were saved by seat belts in 1997 alone. Given that there were almost 50,000 motor-vehicle related fatalities (including pedestrians, back-seat passengers, etc.), our estimates suggest that without seat belts, total fatalities would have been at least thirty percent higher. As important as seat belts are in saving lives, however, increased usage can explain less than half of the observed decline in fatality rates per vehicle mile traveled between 1980 and 1997.

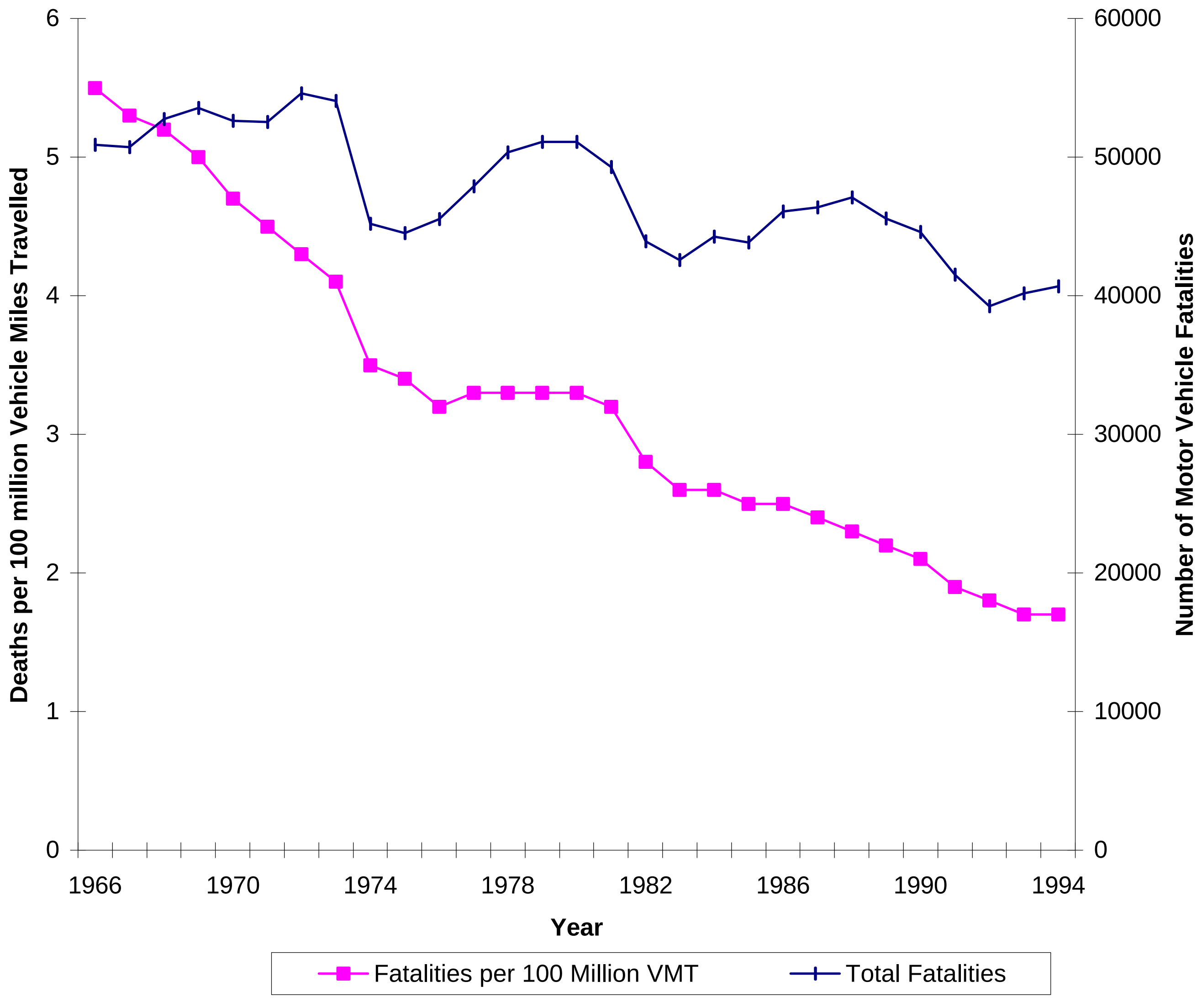
Lives saved due to air bags are substantially lower. Using an effectiveness of .16 for air bags in direct frontal crashes and .09 in partial frontal crashes yields an estimate of 623 lives saved in 1997, or less than five percent of the number of lives saved by seat belts.

Given the comparatively limited effectiveness of air bags, the possibility that seat belt usage declines with the availability of an air bag becomes an important public policy concern. If one in six seat belt wearers decided not to buckle up because of a mistaken notion that the air bag would provide complete protection, then the net impact of air bags on saving lives would be negative. There does not appear to be any evidence for this kind of behavioral response, either in survey data (Williams et al. 1990) or in our sample. In the raw data, seat belt usage is higher

---

in fatal crashes in 1997 (not just our limited sample of two-vehicle crashes), we estimate a seat belt usage rate of 58 percent. This number is lower than estimates based on survey responses, but virtually identical to values obtained in observational studies in which researchers actually tallied seat belt usage by the roadside (NHTSA 1996).

Figure 1: Motor Vehicle Fatalities 1966-1994



among those with air bags. After controlling for an extensive set of observable variables, including whether other front-seat occupants in the vehicle are wearing seat belts, there does not appear to be any systematic relationship between the availability of air bags and seat belt usage.

In addition to calculating the number of lives saved, one might also be interested in the additional number of lives that would be saved if occupants who currently do not use seat belts and air bags were to use them. This value is given simply by rD_0. If all front-seat occupants had worn seat belts in 1997, almost 11,000 further deaths would have been averted. If all vehicles were equipped with dual air bags, an extra 1,900 lives would have been saved.

Given the estimates above, it is clear that seat belts are an extremely good investment from the perspective of cost-benefit analysis. The annual expenditure on equipping vehicles with seat belts is roughly $500 million, yielding a crude estimate of the cost per life saved of roughly $30,000.[22] In comparison over $4 billion dollars are spent annually on air bag installation and maintenance. If all vehicles had dual air bags, approximately 2,500 lives would be saved annually, for an average cost per life saved of $1.6 million. Note that the above estimates understate the true benefits of both seat belts and air bags because they focus solely on deaths prevented, ignoring any impact the safety devices have in reducing injury severity. Nonetheless, at least based on these rough calculations, air bags do not appear to be a particularly effective investment in public health.[23]

---

[22] This calculation assumes a steady state in the number of vehicles on the road, so that the flow of new vehicles offsets the number of vehicles retired annually. If this is the case, than the annual investment in seat belts is constant over time, allowing one to calculate the cost per life saved by simply dividing the annual expenditure on seat belts by annual lives saved.

[23] Graham et al. (1997) performs a much more careful cost-benefit analysis that includes reductions in injuries. Their conclusion with respect to air bags is somewhat more favorable than

V. Conclusion

This paper presents estimates of the effectiveness of seat belts and air bags in saving lives that overcome sample-selection bias inherent to fatal-crash data. We find that wearing a seat belt reduces the likelihood of death by roughly 60 percent. Air bags reduce the probability of death by approximately 16 percent in direct-frontal impacts and 9 percent in partial-frontal impacts. Based on our estimates, seat belts are more effective than is generally thought, whereas air bags are less effective. If our estimates are correct, roughly 15,000 lives were saved by seat belt usage in 1997, along with over six hundred lives saved by air bags. The benefit-cost ratio of seat belts is more than fifty times greater than that of air bags.

More generally, this paper presents an example of how creative use of data can provide useful estimates even when naive estimation provides coefficients that are clearly biased. The key to our identification is the ability to selectively keep or discard observations based on other agents' outcomes. This allows us to separate "good" and "bad" (i.e. biased by sample selection) variation in the data, similar to the way that a valid instrument identifies a parameter using only that part of the variation that is uncorrelated with the error term.

---

ours, both because injuries are considered and because they assume a higher effectiveness rate of air bags. In the sample of crashes we examine, seat belts and air bags are not nearly as successful in eliminating injuries or reducing their severity than they are in averting death. It may be the case, however, that in less severe crashes injury reduction is more effective.

References

Braver, E.R., S. A. Ferguson, M..A. Greene, and A.K. Lund, 1997, "Reductions in Deaths in Frontal Crashes among Right Front Passengers in Vehicles Equipped with Passenger Air Bags," *JAMA* 278:(17, Nov. 5):1437-1439.

Evans, L., 1986, "The Effectiveness of Safety Belts in Preventing Fatalities," *Accident Analysis and Prevention* 18:229-241.

Fell, James C., and Carl E. Nash, 1989, "The Nature of the Alcohol Problem in U.S. Fatal Crashes," *Health Education Quarterly* 16(3) Fall: 335-343.

Ferguson, S. A., A. K. Lund, and M. A. Greene, 1995, Driver Fatalities in 1985-1994 Air Bag Cars, Arlington, VA: Insurance Institute for Highway Safety.

Graham, John D., Kimberly Thompson, Sue Goldie, Maria Segui-gomez, and Milton Weinstein, 1997, "The Cost-effectiveness of Air Bags by Seating Position," *JAMA* (November 5) 278(17): 1418-1425.

Grossman, Michael, Frank J. Chaloupka, Henry Saffer, Adit Laixuthai, 1993, "Effects of Alcohol Price Policy on Youth," NBER Working Paper No. 4385.

Heckman, James, 1979, "Sample Selection Bias as a Specification Error," *Econometrica* 47(1):153-161.

Heckman, James, 1999, "The Economics and Econometrics of Active Labor Market Programs," in *Handbook of Labor Economics*, Volume 3, edited by Orley Ashenfelter and David Card, Chapter 31, Elsevier.

Heckman, James, Hidehiko Ichimura, Jeffrey Smith, and Petra Todd, 1996, "Sources of Selection Bias in Evaluating Social Programs: An Interpretation of Conventional Measures and Evidence on the Effectiveness of Matching as a Program Evaluation Method," *Proc. Natl. Acad. Sci. USA* 93(November): 13416-13420.

Hollands, C. M., F. K. Winston, P. W. Stafford, and S. J. Shochat, 1996, "Severe head Injury Caused by Airbag Deployment," *Journal of Trauma* 41(5; Nov): 920-922.

Imbens, Guido, and Joshua Angrist, 1994,"Identification and Estimation of Local Average Treatment Effects," *Econometrica* 62 (March): 467-75.

Kahane, C., 1996, Fatality Reduction by Air Bags: Analyses of Accident Data Through Early 1996, Washington, DC: National Highway Traffic Safety Administration.

Kaplan, Burton, and R. Adams Cowley, 1991, "Seatbelt Effectiveness and Cost of Noncompliance among Drivers Admitted to a Trauma Center," *American Journal of Emergency Medicine* 9(1; January):4-10.

Larkin, Gregory L., and James E. Weber, 1998, "Letter to the Editor," *JAMA* 279(7;February 18): 506.

Lave, Charles, and Patrick Elias, 1997, "Resource Allocation in Public Policy: The Effects of the 65-MPH Speed Limit," *Economic Inquiry* 35(July): 614-620.

Levitt, Steven, and Jack Porter, 1999, "Estimating the Effect of Alcohol on Driver Risk using only Fatal Accident Statistics," NBER Working Paper No. 6944.

Lund, A. K., and S. A. Ferguson, 1995, "Driver Fatalities in 1985-1994 Air Bag Cars," *Journal of Trauma* 38:49-475.

Marine, William, Edward Kerwin, Ernest Moore, Dennis Lezotte, Anna Baron, and Michael Grosso, 1994, "Mandatory Seatbelts: Epidemiologic, Financial, and Medical Rationale from the Colorado Matched Pairs Study," *Journal of Trauma* 35(1; January): 96-100.

Morris, Michael S. , and Leticia Borja, 1998, "Letter to the Editor," *Journal of Trauma: Injury,*

Infection, and Critical Care," 44(January): 238-239.
National Highway Traffic Safety Administration, 1984, Final Regulatory Analysis: Amendment to FMVSS No. 208, Passenger Car Front Seat Occupant Protection, Washington, DC, July.
National Highway Traffic Safety Administration, 1989, The Effects of the 65 MPH Speed Limit through 1988: A Report to Congress, Washington, DC, October.
National Highway Traffic Safety Administration, 1996, Effectiveness of Occupant Protection Systems and Their Use (Third Report to Congress), Washington, DC, December.
National Highway Traffic Safety Administration, 1998, Traffic Safety Facts 1997, Washington DC.
Orsay, Elizabeth, Timothy Turnbull, Mary Dunne, John Barrett, Patricia Langenberg, and Charles Orsay, 1988, "Prospective Study of the Effect of Safety Belts on Morbidity and Health Care Costs in Motor-Vehicle Accidents, *JAMA* 260 (Dec23/30): 3598-3603.
Peltzman, Sam, 1975, "The Effects of Automobile Safety Regulation," *Journal of Political Economy* 83 (August): 677-725.
Robertson, L.S., 1976, "Estimates of motor vehicle seat belt effectiveness and use: implications for occupant crash protection," *American Journal of Public Health* 66(September):859-64.
Ruhm, Christopher, 1995, "Alcohol Policies and Highway Vehicle Fatalities," *Journal of Health Economics* 15(August): 435-454.
Saffer, Henry, and Michael Grossman, 1987, "Drinking Age Laws and Highway Mortality Rates: Cause and Effect," *Economic Inquiry* 25(July):403-417.
Williams, A. F., J. K. Wells, and A. K. Lund, 1990, "Seat Belt Use in Cars with Air Bags," *American Journal of Public Health* 80(12; Dec): 1514-1516.
Zador, Paul L. and Michael A. Ciccone, 1993, "Automobile Driver Fatalities in Frontal Impacts: Air Bags Compared with Manual Belts," *American Journal of Public Health* 83(5; May):661-665.
Zobeck, Terry S., Gerald D. Williams. Bridget F.Grant, and Darryl Bertolucci, 1990, "Years of Potential Life Lost and Other Trends in Alcohol-Related Fatal Traffic Crashes: 1977-1987, *Alcohol Health and Research World* 14(1):63-68.

</ref_section>

Table 1: Probability of Death as a Function of Number of Occupants in Crash

	Total number of occupants in all vehicles involved in the crash:				
Sample	One occupant	Two occupants	Three occupants	Four occupants	Five occupants
All crashes	1.000	0.538	0.393	0.318	0.269
Anyone else dies in crash	----	0.142	0.157	0.150	0.138
Anyone dies in a different vehicle	----	0.068	0.074	0.070	0.070

Notes: Values in table are computed using all passengers in all vehicles for crashes in which at least one vehicle occupant dies. Crashes in which the only fatality is a pedestrian, motorcyclist, or other non-motorist are excluded. Unlike later tables, the sample used in this table are not limited to front-seat passengers in two-vehicle crashes. Data are for crashes over the period 1994-1997 as reported in FARS.

Table 2: Summary Statistics

	Full sample		With sample selection correction (anyone dies in other vehicle)	
Variable	Mean	Standard deviation	Mean	Standard deviation
Individual-level characteristics				
Die in crash	.386	.487	.090	.287
No seat belt; no air bag	.190	.392	.154	.361
Seat belt only	.385	.487	.412	.492
Air bag only	.108	.310	.082	.274
Both seat belt and air bag	.317	.465	.352	.477
Driver	.705	.456	.734	.472
Male	.577	.494	.635	.482
Age	41.1	20.7		
Vehicle-level Characteristics				
Model year	1992.8	1.9	1993.0	1.9
Vehicle weight	4480	2536	5208	2734
Difference in vehicle weights	2911	10712	1597	3628
Direct frontal impact	.540	.498	.729	.444
Rear impact	.025	.156	.023	.149
Right impact	.188	.391	.101	.302
Left impact	.232	.422	.137	.344
Other impact	.015	.122	.010	.099
Vehicle type: automobile	.591	.492	.434	.496
Driver previous minor violations	.358	.479	.380	.485
Driver previous major violations	.070	.256	.072	.258
Crash-level characteristics				
Year	1995.7	1.1	1995.7	1.1
Speed limit 55 mph or greater	.576	.494	.571	.495
Time of day: 6am to 8 pm	.716	.451	.707	.455
Time of day: 8pm to 1 am	.218	.413	.229	.420
Rural	.607	.488	.608	.488

Notes: Summary statistics are for front-seat occupants involved in two-vehicle crashes in which a motorist is fatally injured between 1994 and 1997. Occupants of large trucks and vehicles built prior to 1990 are excluded from the sample. For other sample restrictions and precise variable definitions, see the text or data appendix.

Table 3: Mean Probabilities of Death by Seat Belt and Air Bag Status
Percent reduction relative to no seat belt, no air bag case

	Full sample			With sample selection correction (anyone dies in other vehicle)		
	Direct frontal	Partial frontal	Non-frontal	Direct frontal	Partial frontal	Non-frontal
	(1)	(2)	(3)	(4)	(5)	(6)
Automobiles						
Seat belts	.35 (.01)	.28 (.02)	.20 (.01)	.49 (.03)	.60 (.05)	.66 (.06)
Air bags	.22 (.01)	.10 (.02)	-.03 (.02)	.20 (.03)	.07 (.06)	-.09 (.08)
Probability of death with no seat belt, no air bag	.615	.679	.698	.255	.291	.337
Utility vehicles, vans, and small trucks						
Seat belts	.65 (.01)	.58 (.02)	.54 (.02)	.69 (.03)	.75 (.04)	.78 (.08)
Air bags	.06 (.02	.06 (.03)	-.01 (.04)	.07 (.04)	.07 (.07)	-.11 (.12)
Probability of death with no seat belt, no air bag	.345	.506	.596	.117	.155	.228
Any covariates included?	No	No	No	No	No	No

Notes:Values in table are the fraction of lives saved relative to a vehicle occupant with no seat belt or air bag based on the raw data without any control variables. The sample is limited to front-seat occupants of vehicles involved in two-vehicle crashes in which a motorist dies. The values in the table are computed from linear probability regressions of an indicator variable for death on seat belt and air bag dummies. The raw probability of death for individuals with neither a seat belt nor an air bag are also reported. For precise variable definitions, see the data appendix. Table A-1 reports the full set of coefficients for columns 1 and 4 of this table.. The first three columns correspond to the whole sample; the last three provide sample-selection corrected estimates. Direct frontal crashes are those in which the principal point of impact is completely frontal (12 on the clock face);partial frontal crashes are those in which either the initial or principal point of impact is between 10 and 2 on the clock face (excluding crashes qualifying as direct frontal). White standard errors accounting for within-vehicle correlations in parentheses.

Table 4: Baseline Regression Estimates of the Effectiveness of Seat Belts and Air Bags in Reducing Fatalities

Percent reduction relative to no seat belt, no air bag case

	Full sample			With sample selection correction (anyone dies in other vehicle)		
	Direct frontal	Partial frontal	Non-frontal	Direct frontal	Partial frontal	Non-frontal
	(1)	(2)	(3)	(4)	(5)	(6)
Automobiles						
Seat belts	.33 (.01)	.28 (.02)	.21 (.01)	.50 (.03)	.59 (.06)	.70 (.07)
Air bags	.18 (.02)	.13 (.02)	-.01 (.02)	.16 (.04)	.06 (.08)	-.16 (.12)
Number of observations	11,049	4,785	7,585	6,123	1,592	869
Utility vehicles, vans, and small trucks						
Seat belts	.51 (.02)	.52 (.02)	.48 (.02)	.64 (.03)	.71 (.05)	.76 (.10)
Air bags	.13 (.03)	.18 (.04)	.13 (.04)	.17 (.06)	.13 (.10)	-.15 (.20)
Number of observations	10,149	3,249	2,416	8,156	1,938	879
Full set of covariates included?	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Values in table are the fraction of lives saved relative to a vehicle occupant with no seat belt or air bag. The sample is limited to front-seat occupants of vehicles involved in two-vehicle crashes in which a motorist dies. The values in the table are computed from linear probability regressions that include controls for the following: seat position, sex, age, principal point of impact, interactions between seat position and principal point of impact, the type of vehicle crashed into, speed limit, vehicle weight, weight differential between vehicles, squared weight differential, past driving record of both drivers, time of day, rural, model year of vehicle, and year of crash. For precise variable definitions, see the data appendix. Table A-1 reports the full set of coefficients for columns 1 and 4 of this table.. The first three columns correspond to the whole sample; the last three provide sample-selection corrected estimates. Direct frontal crashes are those in which the principal point of impact is completely frontal (12 on the clock face);partial frontal crashes are those in which either the initial or principal point of impact is between 10 and 2 on the clock face (excluding crashes qualifying as direct frontal). White standard errors accounting for within-vehicle correlations in parentheses.

Table 5: Sensitivity Analysis of the Reduction in Death Estimates for Seat Belts

	Automobiles				Utility vehicles, vans, and light trucks			
	Direct frontal		Partial Frontal		Direct frontal		Partial Frontal	
Sub-group analyzed	Seat belts	Air bags	Seat belts	Air bags	Seat belts	Air bags	Seat belts	Air bags
Baseline	.50 (.03)	.16 (.04)	.59 (.06)	.06 (.08)	.64 (.03)	.17 (.06)	.71 (.05)	.13 (.10)
Add make and model dummies	.51 (.03)	.15 (.05)	.52 (.06)	.19 (.09)	.64 (.03)	.10 (.08)	.73 (.06)	.02 (.14)
Add vehicle-fixed effects	.38 (.10)	.09 (.10)	.35 (.13)	.07 (.11)	.53 (.11)	.07 (.13)	.57 (.18)	.12 (.22)
Driver	.50 (.03)	.12 (.04)	.62 (.06)	.01 (.09)	.64 (.04)	.15 (.07)	.70 (.03)	.12 (.05)
Front-seat passenger	.52 (.05)	.28 (.08)	.59 (.10)	.10 (.19)	.60 (.08)	.23 (.21)	.66 (.06)	.12 (.15)
Males	.54 (.04)	.12 (.06)	.54 (.08)	.11 (.12)	.65 (.04)	.20 (.07)	.69 (.03)	.17 (.05)
Females	.48 (.04)	.20 (.05)	.65 (.06)	.02 (.10)	.61 (.07)	.01 (.13)	.64 (.06)	-.01 (.10)
Exclude age<16	.49 (.03)	.16 (.04)	.59 (.06)	.07 (.08)	.63 (.03)	.16 (.06)	.67 (.03)	.13 (.05)
Speed limit ≥55	.47 (.03)	.15 (.04)	.62 (.06)	.02 (.09)	.58 (.04)	.17 (.08)	.65 (.03)	.14 (.06)
Vehicles of similar weight	.48 (.03)	.19 (.05)	.60 (.07)	-.00 (.10)	.51 (.07)	.08 (.12)	.57 (.06)	.13 (.08)
Missing data coded as wearing seat belt	.48 (.03)	.15 (.04)	.59 (.06)	.07 (.08)	.60 (.03)	.19 (.06)	.68 (.05)	.11 (.10)
Missing data coded as not wearing seat belt	.46 (.03)	.17 (.04)	.50 (.06)	.10 (.09)	.60 (.03)	.21 (.06)	.70 (.05)	.14 (.10)

Notes: Specifications in this table are variations on the results reported in column 4 of Table 4. The values in the table are the fraction of lives saved by safety devices relative to occupants with no seat belt or air bag. For fuller description, see notes to Table 4 and the data appendix. Direct frontal crashes are those in which the principal point of impact is completely frontal (12 on the clock face);partial frontal crashes are those in which either the initial or principal point of impact is between 10 and 2 on the clock face (excluding crashes qualifying as direct frontal). White standard errors accounting for within-vehicle correlations in parentheses.

Table A-1: Full Estimates of the Raw Regression Coefficients for Direct Frontal Crashes

	Automobiles		Utility vehicles, vans, and light trucks	
Variable	Coefficient	Standard error	Coefficient	Standard error
Seat belt	-.131	.013	-.075	.007
Air bag	-.042	.011	-.020	.007
Driver	.027	.009	.022	.007
Male	-.013	.008	-.001	.006
Age 16-40	.008	.017	.013	.009
Age 41-65	.071	.019	.046	.010
Age 65+	.216	.026	.161	.021
Crash into utility, van or small truck	.058	.015	.048	.007
Speed limit less than 55 mph	-.095	.010	-.034	.006
Vehicle weight ($x10^3$)	-.031	.007	.001	.003
Weight differential * lighter ($x10^3$)	.012	.006	.015	.007
(Weight differential)^2*lighter ($x10^8$)	-.049	.017	-.035	.026
Weight differential * heavier ($x10^3$)	-.057	.010	-.013	.003
(Weight differential)^2*heavier ($x10^8$)	1.020	.017	.036	.010
Previous minor incidents (own driver)	.002	.009	.007	.005
Previous major incidents (own driver)	.007	.016	.002	.005
Previous minor incidents (other driver)	.014	.010	.001	.004
Previous major incidents (other driver)	.011	.017	.034	.012
Time of day: 6 am to 8 pm	-.094	.020	-.049	.016
Time of day:8 pm to 1 am	-.068	.021	-.026	.017
Rural	.044	.010	.018	.006
Model year=1996	-.020	.034	-.017	.016
Model year=1995	.008	.033	-.010	.016
Model year=1994	.003	.033	-.015	.017
Model year=1993	.000	.034	-.013	.018
Model year=1992	-.018	.033	-.017	.018
Model year=1991	-.007	,034	-.017	.018
Model year=1990	.006	,034	-.021	.018
Crash year: 1997	-.015	.014	.009	.008
Crash year: 1996	-.008	.014	-.002	.008
Crash year: 1995	-.009	.014	.000	.008
Constant	.400	.051	.138	.028
Number of obs.	6,123	----	8,156	----
R^2	.127	----	.111	----

Notes to Table A-1: Values in the table are the raw regression coefficients underlying the specifications reported in column 4 of Table 4, i.e. direct frontal crashes with the sample selection correction. See notes to Table 4 or data appendix for a fuller description of the sample. White standard errors accounting for within-vehicle correlations in parentheses.

DATA APPENDIX

(Not for inclusion in paper; just to be available to send out if necessary)

All data used in this study are taken from the Fatality Analysis Reporting System (FARS) collected by the National Highway Traffic Safety Administration (NHTSA). FARS provides detailed data on all occupants and non-motorists involved in motor-vehicle accidents resulting in at least one fatality. Crashes from calendar years 1994-1997 were included in the analysis. In the description below, we include the FARS variable names. Programs and log files are available on request from the authors.

The following exclusions were made from the data set:

1) All non-motorists and motorcyclists and other non-standard vehicles (body_typ>79).

2) Occupants of all vehicles in which only a non-motorist, motorcyclist, or occupant of a non-standard vehicle dies.

3) All large trucks (body_typ>49&body_typ<79).

4) All vehicles with model years prior to 1990 (mod_year<90).

5) All vehicle occupants not in the driver's seat (seat_pos=11) or in the right-front passenger's seat (seat_pos=13). Also, if there were multiple individuals in any one seating position, all were excluded. If there were more than two people in the front-seat, all were excluded (this was a small number of individuals).

6) All vehicles of a make and model-year with fewer than twenty front-seat occupants in the data set over the four-year period covered by our data. This excluded about 10 percent of the observations, but was done to provide a better match on the air bag imputation.

7) Observaations were dropped when the injury severity was unknown (inj_sev==9), when an occupant died prior to the accident (inj_sev==6), or when there was no value given for injury severity (inj_sev==.).

8) Observations where it was unknown whether a restraint was used (rest_use==99), and observations where a value was not coded (rest_use==.) were dropped. If a child safety seat (rest_use==04 or rest_use==14), a motorcycle helmet (rest_use==05 or rest_use15), or a bicycle helmet (rest_use==06) was used the observation was also dropped. Only if no restraint was used (rest_use==0) was the occupant viewed as not having used a seatbelt, otherwise he/she was counted as having worn a seatbelt. Thus, if the passenger was wearing a lap belt, shoulder belt, or a lap and shoulder belt they were considered to have been belted.

9) Air bag availability was assumed whenever more than ten percent of occupants of a given make, mode, year, and seat position were reported as having an air bag (either deployed or undeployed) at least ten percent of the time.

10) Only two-vehicle crashes were included in the data set.

11) In the regressions, any observation with missing data for any of the covariates were excluded.

Other variable definitions

If whether a driver had a previous accident, previous speeding violation, previous moving violation, previous license suspension or previous DWI was not known, it was assumed that the driver did not have that particular violation. A driver was said to have a previous minor violation (rec_min==1) if he/she had a previous accident (prev_acc>0), a previous speeding violation (pre_spd>0), or a previous moving violation (prev_oth>0). A driver was said to have a previous major violation (rec_maj==1) if he/she had a previous license suspension (prev_sus>0) or a previous DWI (prev_dwi>0). Dummy variables were generated for if the driver of my car had a previous minor or previous major offense (rec_min or rec_maj), and for if the driver of the other car involved in the accident had a previous major or previous minor offense (rec_omin or rec_omaj).

Principle point of impact variables were generated for direct frontal impact (cc2front=impact2==12), passenger's side impact (cc2left=(impact2>=7&impact2<=11)), driver's side impact (cc2right=(impact2>=1&impact2<=5)), and rear impact (cc2back=(impact2==6)). If information on the principle point of impact was missing (impact2==.) the observation was dropped. Furthermore, interaction variables were generated for observations where there was a passenger side principle point of impact and the observation was that of a passenger, and similarly when there was a driver's side principle point of impact and the observation was that of a driver. Partial frontal crashes were defined as those that were not direct frontal, but had some frontal component (impact1>9 or impact1<3 or (impact2>9& impact2<12) or impact2<3).

A discrete choice variable was generated for cases where the speed limit was less than 55 miles per hour (sp_limit>00&sp_limit<=55), unknown (sp_limit==99), or if there was no mandatory speed limit (sp_limit==00). Observations were dropped if the speed limit variable was not coded (sp_limit==.).

Two discrete choice variables were generated for the time of the crash: daytime and nightime. Daytime was considered to be after 5am and before 7pm (hour>=05&hour<19), while nighttime was considered to be later than 7pm and earlier than 1am (hour>=19 &hour<=24 |hour>=00&hour<01). Observations were dropped if the hour of the crash was missing (hour==.).

Three variables were generated for 3 of the 4 different years of data included in this study (year==95, year==96 and year==97).

Variables for the model year of the car were created for model years 1990-1996 (mdyr90. . .mdyr96). Variables for model year 1997 and 1998 were not created because of the high degree of collinearity these variables would have with the 1997 year variable and cars of model years before 1990 were dropped because very few had airbags.

A variable was generated for if the road was a rural road (road_fnc<10). A variable was generated for cases where all deaths occurred in one car.

Body type variables were divided into two categories: automobiles and all other (utility vehicles, vans, and light trucks). Automobiles included all types of standard automobiles (body_typ>00&body_typ<=11). All other is defined as body_typ>11&body_typ<50.

A control variable was created that was equal to one if the vehicle crashed into was not an auto (including the other body types above and large trucks).

Various variables taking account of weight were included in the regressions. Since specific weights for trucks were not given, the midpoint of the interval which they were within was used. For instance, if a truck's weight was given to be within the interval of 10,000 to 14,000 lbs, it was assumed that its weight was 12,000lbs. This was done for all intervals except the last weight class which specified only that the truck weighed over 33,001lbs. In this case, the weight was assumed to be 40,000 lbs.

Five variables were taken into account when considering weight: dif_wgtl, dif_wgth, dif_squl, dif_squh and dif_actu. Dif_wgtl is how much less in weight (lbs) your vehicle weighed compared to the other car in the wreck; if your vehicle is heavier, than dif_wgtl is coded as zero. Dif_wgth is the how much more your vehicle weighed in lbs compared to the other vehicle in the wreck; if your vehicle is lighter, this variable is coded as zero. Dif_squl is simply dif_wgtl squared, and similarly dif_squh is dif_wgth squared. The actual weight of the vehicle was also included in the regression.. Observations were dropped if the weight of the vehicle was not coded.

Three discrete choice variables for age were generated. Agee1640 (age>=16&age<=40) included persons aged 16 to 40 inclusive. Agee4165 (age>=41&age<=65) included persons 41 to 65 years of age. Finally, agee65 (age>65) included only those over the age of 65. Observations for which age was either not coded or unknown were dropped.

A variable for sex of the passenger was defined as (sex==1) and takes on a value of one if the passenger is male. If the sex variable was unknown or not coded the observations was dropped.