Preventing a Vehicle Arms Race: The Social Benefits of "Weightless" Safety Ratings

Thao Duong*

Jonathan B. Scott†

September 30, 2024

Abstract

The two prominent safety rating agencies evaluate their frontal crash tests unconditional on vehicle weight. Their in-lab, controlled crashes largely resemble collisions between two equal sized vehicles. However, it has been well-documented that a larger vehicle is objectively safer in a two-car collision, while imposing additional harm to drivers of smaller vehicles. This paper estimates the social benefits of a ratings regime which neglects the role of weight in its safety evaluations. We compare the current methodology to one incorporating relative weight by estimating the marginal effect of weight on vehicle fatalities and mapping those effects into a re-calibrated safety rating. For identification, we exploit underlying, in-lab crash metrics determining the rating to measure both the reduction in fatality risk, and to isolate a demand response to the rating. Counterfactual calculations demonstrate that incorporating the role of weight into safety ratings increase the demand for trucks and SUVs by 2.6%, potentially exacerbating an arms race in vehicle size.

*Texas A&M University; **thaoduong@tamu.edu**

†University of Texas, Dallas; **jscott@utdallas.edu**

Preliminary working draft. Any errors are our own.

1 Introduction

It has been well-established that consumers have a strong preference for large vehicles, holding other characteristics constant. These preferences generally arise from a demand for additional capacity or cargo space, but prior studies have also demonstrated the role safety plays in the demand for increased vehicle size (e.g., Li, 2012; Scott, 2022). Heavier vehicles have been empirically shown to mitigate much of the damages incurred in vehicle accidents. However, while safer to the drivers of these vehicles, larger cars have also been associated with higher emissions levels, degradation of public roads, as well as an increased risk to others when involved in a car accident.

There is an extensive literature on the external cost of heavier vehicles on the road and the behavior that produces these externalities. Scott (2022) provides evidence that consumers internalize vehicle size as a safety attribute by examining the purchasing behavior of consumers following information of a nearby neighbor's involvement in a fatal car accident. Anderson and Auffhammer (2014) document the external costs of pounds by estimating the increased fatality risk imposed by heavier vehicles in two-car collisions. Li (2012) examine consumer preferences for safety directly in the context of vehicle demand. Seminal work of White (2004) illustrates that consumer preferences for weight as a safety attribute can generate a vehicle "arms race" on the road; behavior that may generate an inefficiently heavy vehicle fleet.

The concept of a vehicle arms race describes a prisoners' dilemma wherein consumers purchase heavier and heavier vehicles to protect themselves against other car on the road. This behavior has clear implications for safety and produces an inefficient equilibrium, where the vehicle fleet is much heavier than it otherwise would be. How consumers best respond to the distribution of vehicle weights on the road is difficult to estimate due to the inherent endogeneity in choice. Thus, there is very limited evidence that this occurs in practice. In this paper, we explore an alternative channel to pin down how publicly available information on safety might contribute to (or mitigate) an arms race.

There are two prominent rating agencies charged with producing information on vehicle safety: the National Highway Traffic Safety Administration (NHTSA) and the Insurance Institute for Highway Safety (IIHS). Ratings are derived from various in-lab crash tests, estimating the damages incurred through a controlled vehicle accident. By design, both agencies derive a final safety rating that is independent of a vehicle's size. While empirical studies have shown that heavier vehicles are objectively safer, frontal car crash tests are conducted by driving a vehicle into a wall, an outcome which simulates a two-car collision between equal sized vehicles. Thus, as currently constructed, ratings provide information on the safety of a given vehicle, within weight class, and provide little information on the role of that vehicle's weight, or other vehicle weights, in real-world accidents. This approach has clear social efficiency implications.

In this paper, we explore the potential outcome of a counterfactual methodology which would incorporate information about vehicle weight into the ratings. This approach, arguably, produces a more accurate depiction of risk, but may generate an inefficient response by consumers who subsequently wish to purchase larger cars. By estimating the response to these "full-information" safety ratings, we are able to quantify the social benefit of hiding or, "conditioning out," the role of weight in reputable sources of information on vehicle safety.

Our empirical approach is two-fold. On the front end, we estimate the consumer response to safety ratings. We construct a simple vehicle demand model, incorporating IIHS top safety picks as an additional vehicle attribute. These picks are determined based on 8 unique crash tests, the crash outcomes translated to a Likert scale describing the vehicle's performance (i.e., from poor, marginal, acceptable, and good). While salient information of an IIHS top pick may be correlated with other vehicle characteristics, the criteria for choosing a top pick changes yearly, and should be orthogonal to other attributes. For example, in one year, a top pick classification may only require an "acceptable" rating for both crash test 1 and 2, while in the following year it may require a "good" classification for crash test 1. Our approach assumes that all confounding factors associated with unobserved vehicle characteristics are captured by the individual test outcomes, thus, leveraging the remaining variation in the criteria change to pin down the causal response to a top safety pick. Controlling for model year and dummies for the individual crash test ratings produces an empirical specification similar to a difference-in-differences design.

Of primary interest is the counterfactual demand for a vehicle under an alternative methodology which incorporates weight into the safety rating. Thus, following the identification of key preference parameters, we leverage empirical data on car crashes to recalibrate the rating. Ultimately, we explore the mechanical relationship between a top safety pick and fatality risk, then generate a mapping between fatality probabilities and the safety rating. As we are only interested in the mechanical relationship, the underlying empirical challenge is the potential selection of drivers of safer vehicles into more dangerous car crashes. To mitigate this concern, we leverage only information from the 8 individual car crash ratings. Our approach assumes that the IIHS top pick classification is the only information salient to the consumers, while the individual crash test outcomes determining the safety designation are less known. Using the unobserved (to the driver) crash test results in an instrumental variable design allows us to estimate how much safer the top picks are in traffic accidents.

The traffic fatality risk regression is both a function of the top pick designation and relative vehicle size—an additional component of risk not accounted for in the crash tests. To mitigate potential selection concerns related to vehicle size, we follow Anderson and Auffhammer (2014) and estimate the role of relative weight on the sample of two-car collisions, where the assignment of the "opposing" vehicle in the accident is assumed exogenous. Given the estimated effects of top picks and vehicle size on fatality risk, we are then able to perform a simple inversion to incorporate relative vehicle weight into a counterfactual safety rating.

Our results illustrate the manner in which demand for weight may be exogenously impacted through external information on vehicle safety. Under our recalibrated ratings, 13 percent of vehicles in our sample flip from a non-safety pick to a top safety pick, or vice versa, depending on vehicle weight. However, as our recalibration generates a continuous measure of a top pick assignment, all ratings differ to some extent depending on relative weight. This new ratings system significantly increases preferences for heavy vehicles. Given the estimated, baseline marginal demand for weight arising from other factors, our results show a 10 to 13 percent increase in implied preferences for weight after the ratings adjustment. This further translates into a 2 and 6 percent increase in demand for SUVs and light trucks, respectively. Given the well-known negative externalities generated from large vehicles, our findings highlight the potential for ratings agencies to mitigate a vehicle arms race by excluding relevant information on weight from its safety evaluations.

The conventional view is that heavier vehicles provide more safety when experiencing an accident (Crandall and Graham, 1989; Van Auken and Zellner, 2005; Anderson, 2008). However, consumers who invest in their safety by purchasing pickup trucks and SUVs create an increased risk on the road to other drivers. It is well-established that heavy vehicles create a higher risk of fatality to the opposing driver when involved in a two-car accident. For example, Jacobsen (2013) finds that a 1,000 pound increase in the weight of a vehicle involved in an accident increases the number of fatalities in other vehicles by about 46 percent. Anderson and Auffhammer (2014) find, conditional on the occurrence of an accident, a 1,000 pound increase in vehicle weight is associated with a 47 percent increase in the baseline fatality probability. Evans (2001) finds that the effects of adding mass in the form of a passenger adds to the increased risk of fatality in head-on collisions. He finds that adding a passenger to a car leads to a 7.5 percent reduction in driver's risk of fatality, while increasing the risk of fatality to the other driver by 8.1 percent. Bento, Gillingham, and Roth (2017) further examine the relationship between weight dispersion and fatalities in accidents, but in the context of the Corporate Average Fuel Efficiency (CAFE) standards. They find that, though CAFE increased the national fleet's weight dispersion, the resulting decrease in average weight has produced an overall reduction in fatalities.

To our knowledge, this is the first paper illustrating the role of full information in safety ratings on an arms race effect. Our counterfactual results hold the direct response to weight constant, while exploring the effects of weight through the safety ratings channel. When ratings are updated periodically, only the heaviest vehicles will be designated a top pick, holding other factors constant. This generates a feedback loop with inefficient outcomes. Our results suggest large efficiency gains from hiding information on weight, a current practice which potentially dampens the effect of a vehicle arms race.

The paper proceeds as follows. Section 2 provides an overview of data sources used in our analysis. Section 3 describes our empirical strategy. Section 4 presents our main findings. Section 5 concludes the paper with a brief discussion of the implications of our results.

2 Data

In this section, we review the primary data sources used in our analysis. These include safety ratings information, new vehicle sales, and traffic accidents. We describe these data sources in detail below.

2.1 Safety Ratings

There are two primary vehicle characteristics of interest in this paper: weight and safety ratings. These attributes are used both in a demand framework and to estimate their role in traffic fatalities. This section discusses the safety ratings used in the paper.

Historically, two prominent ratings agencies represent the dominant sources of information on vehicle safety; each agency deriving their safety ratings through in-lab crash tests. The National Highway Traffic Safety Administration (NHTSA) conducts periodic safety testing on various models and publishes their ratings on their website.[1] There are several limitations to these data, however. NHTSA crash tests are not universally conducted across all vehicle models due to excessive costs. Further, the vehicle models which do have ratings available are only tested on particular model years. This leaves many gaps in the data in which a safety rating for a given make-model-year may not be observed.

Given these limitations, we conduct our analysis using crash ratings published by the Insurance Institute for Highway Safety (IIHS), whose breadth and frequency of testing generally surpasses that of NHTSA. There are many similarities in the testing methodologies between IIHS and NHTSA. For example, each agency produces their ratings on a scale which may only be interpreted within a given weight class. On their website, NHTSA offers the disclaimer: "Overall Vehicle Scores can only be compared to other vehicles in the same class and whose weight is plus or minus 250 pounds of the vehicle being rated." The reason is due to the controlled design of the crash tests, which explicitly hold weight constant. IIHS

[1] nhtsa.gov/ratings

describes their frontal crash tests as one that simulates a collision between two vehicles of equal size: "The forces in the test are similar to those that would result from a frontal offset crash between two vehicles of the same weight, each going just under 40 mph." These "weightless" safety ratings motivate our research question.

The IIHS was established in 1959 by three leading insurance associations that collectively represented 80 percent of the U.S. auto insurance market. The nonprofit agency produces annual publications, awarding the best performing vehicles IIHS "top" and "top plus" safety picks. The information is widely publicized, often cited in manufacturer marketing campaigns when awarded top pick status. Still, we acknowledge that this is only one source of information on safety available to consumers and, thus, the results of this paper should be interpreted through the lens of IIHS ratings, rather than a combination of all available sources.

The IIHS awards are the result of several tests aimed at estimating a vehicle's performance in a car crash. Metrics from each of the 8 tests performed are translated into an individual score and are used to determine top pick designations. Tests 1-7 are evaluated on the four categories: "poor", "marginal", "acceptable", "good". The frontal crash prevention technology is evaluated in test 8 on, first, its availability and, second, whether the equipped technology is classified as "basic", "advanced", or "superior" according to IIHS metrics. The thresholds for each of the 8 tests in determining a top pick are revised each year, generating variation in criteria across model years.

We use the agency's public API to gather information for each crash test and merge the data by model to our sales and traffic accidents data. Both the individual crash tests and the ultimate top pick designations are essential components of our empirical design. The designation criteria in each model year in our data set are illustrated in Table 1. The identifying variation comes from the interaction of the results from the 8 primary tests, and the changes in the thresholds which determine a vehicle's top pick classification.

2.2 Vehicle Sales and Characteristics

New vehicle sales make up our main outcome of interest in estimating a vehicle choice model. The data derive from Texas vehicle registrations from the Department of Motor Vehicles (DMV), and are reported at the county-month level for the years 2014-2019. For our analysis, we aggregate sales to the metropolitan statistical area (MSA) level. The choice of MSA-level sales is chosen as a better representative sample of a market for vehicles; we observe multiple sales for the vast majority of vehicle models in our sample at this level of aggregation. Registrations of new vehicles are reported by their unique 17-digit vehicle identification number (VIN). To obtain the specific characteristics of each vehicle, we make use of a VIN decoder.

Vehicle characteristics come from DataOne Software, in which we directly join in unique information on each vehicle based on the 10-digit VIN stub (VIN10). A VIN10 defines a vehicle up to their make-model-year-trim level, in addition to particlar packages. Our analysis aggregates vehicles into a make-by-model-by-year-by-fuel type index, as occasionally, a model will be produced in multiple fuel types. The characteristics of interest are generally constant within this narrow vehicle description. The main characteristic of interest in this paper is a vehicle's weight, as defined by its curb weight.[2]

Once all characteristics are collected, we combine information on reported initial registration dates and model year in order to infer new vehicle sales. For this, we assume that a sale that takes place in a year in or preceding that of the model year is a new vehicle sale.

2.3 Traffic Accidents

A recalibration of the safety ratings to one which accounts for vehicle weight is of primary interest in this paper. Given the infeasibility of redesigning the crash tests to account for vehicle relative weights in two-car collisions, we gather this information directly from empirical data on traffic accidents.

Data on traffic accidents are collected from the Texas Department of Transportation's (TxDOT) Crash Records Information System (CRIS). CRIS is a comprehensive source of all

[2]The curb weight of a vehicle accounts for all standard components, including a full tank of gas and all necessary operating fluids.

reported accidents in Texas, with granular information describing the characteristics of an accidents at the crash, vehicle, and person level. The main analysis is conducted on the sample of two-car collisions, in which the VINs of the vehicles involved are the primary identifiers and allow us to merge together relevant vehicle characteristics.

The primary outcome of interest is a traffic fatality occuring, defined at the vehicle level. The VIN allows us to merge in any relevant characteristics (including safety ratings). Other characteristics of the crash such as county, date of crash, and speed limit offer additional controls.

3 Empirical Strategy

This paper seeks to measure the demand responses to a counterfactual ratings methodology which incorporates the role that weight plays in vehicle safety. We hypothesize that when weight is indirectly internalized through this additional channel, a feedback loop is generated whereby consumers purchase heavier and heavier vehicles; particularly when *relative* weight on the road is of significance.

Our empirical methodology consists of two primary components. First we estimate a simple vehicle demand model, identifying the causal response to IIHS safety ratings. Next, we recalibrate the historical ratings by incorporating vehicle weight. This requires us to create a mapping between the crash test data and empirical fatality risk, combining the estimated effects of weight on crash outcomes.

3.1 Demand Model

We begin with a standard logit model of vehicle choice which includes the IIHS top safety pick designation as an additional vehicle attribute. For a vehicle model j in model year t, define the variable top_{jt} as a binary indicator representing the top safety pick designation. As top picks are likely correlated with other unobserved vehicle characteristics, we control for a set of dummy variables representing the individual crash test outcomes that determine the top pick designations. Thus, for an individual crash test indexed by $r = 1, ..., 8$, we can describe the crash test outcomes on the set $\{Poor, Marginal, Acceptable, Good\}_{r=1,...,7}$, for tests 1-7, and the set $\{Not\,Available, Basic, Advanced, Superior\}_{r=8}$, for test 8. The test performance for each vehicle can then be described by the vector of dummy variables $R_{rjt} = (\mathbb{1}^{poor}_{rjt}, \mathbb{1}^{marg}_{rjt}, \mathbb{1}^{acc}_{rjt}, \mathbb{1}^{good}_{rjt})$, or $R_{8jt} = (\mathbb{1}^{na}_{8jt}, \mathbb{1}^{basic}_{8jt}, \mathbb{1}^{adv}_{8jt}, \mathbb{1}^{sup}_{8jt})$ for test 8, where $\mathbb{1}^{k}_{rjt}$ indicates whether test r produces an outcome of *at least* a k classification for model j in year t. Let the full vector of dummy variables across all 8 tests be defined as $R_{jt} = (R_{1jt}, ..., R_{8jt})$. This describes a vehicle's overall test performance and completely determines its top pick designation, conditional on the criteria in each model year.

Table 1 describes the relative outcome required in each test to be awarded a top safety pick. For example, in model year 2018, a top safety pick is defined as $top_{j,2018} = \mathbb{1}^{good}_{1j,2018} \times \mathbb{1}^{good}_{3j,2018} \times \mathbb{1}^{good}_{4j,2018} \times \mathbb{1}^{good}_{5j,2018} \times \mathbb{1}^{good}_{6j,2018} \times \mathbb{1}^{acc}_{7j,2018} \times \mathbb{1}^{adv}_{8j,2018}$. Notice how crash test 2 is not relevant in 2018 and, thus, only the minimum performance (i.e., poor) was required. In contrast to the 2018 criteria, there was no minimum requirement for crash test 7 in 2017 and, in 2019, a top safety pick required at least an "acceptable" rating for crash test 2. In addition to controlling for each test separately, to further isolate the interaction between each test, we control for a set of model year dummies to account for any systematic confounders across years, correlated with the criteria changes. Our main specification controls for model year-by-city fixed effects.

Our main specification is a standard logit which estimates the marginal effect of top safety picks on city-level vehicle demand. We estimate the following equation.

$$log(q_{cjt}) = \alpha top_{jt} + X_{jt}\beta + R_{jt}\Gamma + \phi_j + \lambda_{ct} + \varepsilon_{cjt} \tag{1}$$

where the outcome of interest is logged, aggregate quantities of model j in model year t, purchased in city c. Aggregate city-year purchases in the denominator of market shares are absorbed into city-by-time fixed effects, λ_{ct}, and we additionally control for model-specific fixed effects, j. Γ estimates the reduced-form contributions of each crash test outcome to aggregate demand and, jointly, $R_{jt}\Gamma$ absorb potential confounders correlated with safety performance.

We include various set of vehicle characteristics in X_{jt}, including vehicle weight. In some specifications we include vehicle price, identified through conventional instrumental variable approaches used in demand estimation (Berry, Levinsohn, and Pakes, 1995). However, prices are held out of our main specification in order to properly identify the overall, reduced form effect of the safety ratings. Since manufacturers likely respond to an increased demand for vehicles with top safety picks by increasing prices, without a full general equilibrium analysis, it is important to exclude prices to account for the net effects of safety ratings.

Finally, α estimates the causal effect of IIHS top pick designations on vehicle demand. Our identification strategy assumes conditional independence of the top pick designation, where unobserved vehicle characteristics correlated with the award are captured by a linear set of dummy variables describing crash performance. When interpreted as market shares, ε_{cjt}, is assumed to be an extreme value type I, logit shock, independent across models in the choice set.

3.2 Empirical Fatality Risk

The second component of the empirical strategy involves a recalibration of the top safety pick designations (i.e., top_{jt}) to incorporate additional information on vehicle weight. As frontal crash rating tests simulate a two-car collision of equal sized vehicles, the marginal effect of weight on the ratings is explicitly zero by design. Therefore, to recalibrate the ratings, we must leverage empirical data on traffic accidents.

We begin with a simple linear probability model depicting fatality risk. The model is estimated on the subset of two-car collisions, where vehicle weights are included in a manner similar to Anderson and Auffhammer (2014).

$$death_{ijt} = \delta top_{jt} + \gamma(weight_{ijt} - weight_{it}^{opp}) + \varepsilon_{ijt} \tag{2}$$

where the outcome indicates a death taking place in the collision. δ describe the mechanical relationship between the safety rating and fatality risk when drivers and cars are randomly assigned to accidents and γ is the marginal effect of relative weight, where $weight_{ijt}$ describes driver i's vehicle weight and $weight_{it}^{opp}$ is the opposing vehicle weight. For simplicity, j is collapsed to an index describing make, model, and year of the vehicle, while t is redefined as the year of the accident.

When δ is identified as the reduced fatality risk of a top pick, we can directly recalibrate the safety pick designations accounting for the role of vehicle weight. The procedure simply re-scales fatality risk to match top pick probability, implicitly assuming a linear mapping between the two metrics. The counterfactual safety ratings are then calculated as the following.

$$top_j^* = top_j + \frac{\partial top}{\partial w}(w_j - \bar{w}) = top_j + \frac{\gamma}{\delta}(w_j - \bar{w}) \tag{3}$$

where $\bar{w}$ defines a baseline vehicle weight and w_{jt} is the vehicle's weight. When $w_{jt} = \bar{w}$ — as is imposed by design in the controlled frontal crash tests — the counterfactual rating equals the true rating. Equation 3 illustrates how a feedback loop might occur when relative weight is incorporated into safety ratings, where average weights, $\bar{w}$, might be updated by the rating agency on a periodic basis.

Given that drivers and vehicles are not randomly assigned to accidents, the mechanical relationship describing the empirical safety of a top pick is not identified through direct estimation of Equation 2. In addition to the mechanical effect, there is likely a behavioral component captured; for example, if drivers of top safety picks drive more recklessly (e.g., Peltzman, 1975). Thus, the goal is to separate this behavioral response from the mechanical relationship.

To explain the intuition behind our strategy, let the following equation illustrate a linear projection of the top safety pick rating on the individual crash test dummies.

$$top_j = R_j \tilde{\Gamma} + u_j \tag{4}$$

u_j contains valuable information on the relevance of each element of R_j in the top pick designation, a classification which is modified by IIHS periodically. Thus, while the outcome of each individual crash test (i.e., R_j) may not be salient to the consumer, the top pick is, resulting in potential selection into fatal accidents based on u_j. Controlling for u separately in Equation 2 resembles a control function approach, explicitly accounting for the behavioral response, further isolating the mechanical effect of safety ratings. This is the intuition behind our strategy and, as the control function approach is equivalent to two-stage least squares, we implement an instrumental variables design, leveraging the individual crash test outcomes to pin down the mechanical relationship between IIHS safety ratings and empirical fatality risk.

Finally, given the potential for consumers to additionally select into fatal car crashes based on their vehicle size, we will only interpret our estimate of relative weight, γ, based on independent variation in opposing vehicle weights. As vehicles in two-car collisions are presumed to be matched together in a plausibly exogenous manner, this approach should credibly pin down the causal role of vehicle size in fatality risk. We additionally include varying sets of geographic and vehicle-specific controls and fixed effects to further isolate this variation.

3.3 Effect of Weight-Incorporated Safety Ratings

The primary estimate of interest in this paper is the effect of a counterfactual, weight-incorporated safety rating on vehicle demand. To calculate this, we substitute the recalibrated top pick rating in Equation 3 into the demand model from Equation 1. Deriving the treatment effect on demand, which compares the counterfactual rating to the current regime, gives the following.

$$\frac{\%\Delta q(w)}{\Delta regime} = \alpha \frac{\gamma}{\delta}(w - \bar{w}) \tag{5}$$

Thus, the combined parameter $\alpha\gamma/\delta$ is the key estimate of interest in this paper and illustrates the additional demand for relative vehicle weight exogenously produced by a ratings design which incorporates objective information on vehicle size into published safety evaluations.

4 Results

In this section, we present the main parameter estimates from both our demand and fatality risk models. Finally, the estimates of interest combine the two models and explores various counterfactual estimates of a new ratings regime.

4.1 Demand Estimates

The main estimates from our demand model are presented in Table 2. The coefficient of interest is that on the top safety pick indicator, α in Equation 1. Column 1 reports a standard fixed effects estimator, which controls for make-model and city-by-year effects. This specification does not control for other potential confounders related to the safety award. Column 2 introduces granular controls for the 8 individual safety test outcomes. Column 3 includes additional vehicle characteristics as controls and Column 4-5 control for MSRP.

Columns 2 and 3 report results from our primary specification which directly controls for the vehicles' performance in each of the crash test outcomes. These estimates suggest top pick awards generate 5 percent higher demand relative to non-top safety picks. A causal interpretation requires that all other confounders are absorbed into the crash test dummies, while exploiting the residual, plausibly exogenous variation in the criteria change for top pick classification. Under this conditional independence identification assumption, Columns 2 and 3 identify the *reduced form* effect of top pick designation on vehicle demand. This captures a direct response to the rating, but also a potential indirect response through vehicle prices. Without a comprehensive, general equilibrium analysis, this net effect is the one of interest.

To explore potential supply-side effects and isolate the direct effects of safety ratings, we control for vehicle prices directly. Column 4 includes MSRP in the regression, ignoring potential price endogeneities. Our primary estimate does not change significantly, which we can either attribute to a low correlation between prices and the top pick classification (e.g., no supply-side response to ratings) or to inelastic consumer demand (e.g., prices are not relevant in the demand equation). To correct for the likely upward biased estimates on MSRP arising from unobserved product characteristics, we estimate the equation by two-stage least squares, leveraging standard price instruments proposed by Berry, Levinsohn, and Pakes (1995). Specifically, these include the average level of each of the included characteristics across all other models in the choice set. Unsurprisingly, after instrumenting for price, elasticities increase and the estimate on top pick designation doubles in magnitude. Given the validity of these instruments, we interpret this finding as a possible supply-side response by the vehicle manufacturer, who sets higher prices following the top pick award, partially — but not completely — offsetting the demand response.

4.2 Fatality Risk Mapping

Table 3 reports the main estimates from the vehicle accidents model in Equation 2. Column 1 presents a naive specification which aims to directly measure the empirical safety of an IIHS top pick. While the negative sign indicating a drop in fatality risk is as expected, we also expect an upward biased estimate when top safety pick drivers are more likely to select into high risk accidents. Columns 2-6 aim to mitigate this bias by instrumenting for the safety ratings using the 8 individual crash test outcomes, assumed to be less salient to the driver. Column 2 presents the two-stage least squares results with no additional controls, while Column 3 includes the curb weights of both the driver's vehicle and the opposing vehicle. As in Anderson and Auffhammer (2014), we will only interpret the estimates of the opposing vehicle weight as causal given the potential endogeneity of one's own-vehicle weight. These estimates illustrate a relative increasing in fatality risk .01 percentage points per 1,000 pounds.

The results in Table 3 produce a mapping between safety ratings and fatality risk, which we next exploit in order to incorporate weight into a counterfactual safety rating. Of interest is the ratio of the weight coefficient to the top safety pick coefficient, or γ/δ in Equation 3. The counterfactual ratings are calculated for each vehicle in our data set, given the weight of that vehicle, evaluated relative to the overall mean weight for that model year. The use of model year-specific mean weights is performed strictly for illustrative purposes. An alternative regime might be one that uses mean fleet weight — the weight of a representative vehicle involved in an accident — or even one which varies over geographic location.

When the safety picks are reevaluated, we have a new rating system which accounts for the objective role that weight plays in vehicle accidents. For a heavy car, this may partially offset factors that otherwise weaken a vehicle's safety performance. To illustrate the degree to which the safety ratings might flip, in Table 4, we present the vehicle models with the largest relative change in rating after accounting for weight. Here, we define the relative change in rating of a vehicle j as: $\Delta_j = top_j \cdot (1 - top^*_j) + (1 - top_j) \cdot top^*_j$. Note that counterfactual ratings are not bounded between one and zero, given the linear mapping imposed in the derivation.

The distribution of the rating changes across all models in our data set is presented in Figure 1. By construction these changes are centered on zero, where no change is added when a vehicle's weight equals the average weight. Thirteen percent of observations are outside of the (-0.5,0.5) interval, indicating the vehicles with a probable switch in top pick status under the new methodology.

4.3 Combined Model

In a standard model of a vehicle arms race, consumers receive utility from weight for both the value placed on added capacity or cargo space and from the safety it provides (White, 2004; Li, 2012; Anderson and Auffhammer, 2014; Scott, 2022). However, these two mechanisms are difficult to disentangle from observational data on vehicle purchases. In this paper, we form a thought experiment, whereby publicly available information on safety of a vehicle is adapted to account for vehicle weight, suggesting an additional lens in which preferences for weight as a safety attribute may be isolated from other factors.

Under the new ratings regime, following similar notation from Equations 1 and 3, we might write demand for vehicles as a function of vehicle weight and safety ratings as follows.

$$\begin{aligned} log(q_{jt}) &= \alpha top^{*}_{jt} + \beta^{w} w_{jt} + \varepsilon_{jt} \\ &= \alpha top_{jt} + \alpha \frac{\gamma}{\delta}(w_{jt} - \bar{w}) + \beta^{w} w_{jt} + \varepsilon_{jt} \end{aligned} \tag{6}$$

This provides a framework for evaluating the extent to which preferences may exogenously be affected by an alternative ratings regime. Of interest is the additional demand for weight arising from the new methodology or, the combined parameter $\alpha\gamma/\delta$. Assuming constant demand for weight due to other factors, β^{w}, we can compare these two weight parameters in order to measure the extent to which the size of vehicles might grow due to this change in ratings methodology.

Estimates for the combined weight parameter are presented in Table 5. Columns 1-3 present estimates combining specification 2 of Table 2 (i.e., demand model D2) with Columns 3-5 of Table 3 (i.e., crash model C3-C5). Columns 4-6 combine Column 3 of Table 2 with Columns 3-5 of Table 3. Standard errors are derived using the delta method and the corresponding covariance matrices from these two models. The model suggests that the new ratings regime would increase the marginal demand of a 1,000 pound increase in weight by about 3 percent. When comparing this estimate to the weight parameters in Table 2, our estimates suggest that such a policy change would increase preferences for weight by about 10 to 13 percent.

Figure 2 plots the treatment effect of the methodology change from Equation 5 across the observable range of vehicle weights. A bar plot depicts the average effect for vehicles classified as a "Car", "SUV", or a "Truck". The width of each bar represents the interquartile range of relative weight for each vehicle type and weights are evaluated against the sales-weighted, average curb weight in our data set; approximately 3,800 pounds.

These estimates suggest that under this revised ratings system, cars are 1 percent less likely to be purchased, while SUVs and Trucks are 2 and 6 percent more likely to be bought, respectively. Jointly, SUVs and Trucks are 2.6 percent more likely to be purchased when vehicle size is accounted for in safety ratings.

5 Conclusion

A vehicle arms race occurs when consumers rationally internalize vehicle size as a safety attribute, thus, purchasing heavier and heavier vehicles to mitigate their risks against other large cars on the road. Estimating this best response mechanism is difficult given the inherent feedback loops defining an arms race, though prior studies have provided evidence that such preferences likely exist in practice (e.g., Li, 2012, Scott, 2022). The social costs associated with this behavior has also been clearly documented in the literature (Crandall and Graham, 1989; Evans, 2001; Van Auken and Zellner, 2005; Anderson, 2008; Jacobsen, 2013; Anderson and Auffhammer, 2014), as heavier vehicles generally produce both environmental externalities and additional safety risks to other drivers.

When concerned about safety, a consumer may look to multiple sources of information to make a suitable vehicle choice. While they might directly internalize weight as a safety attribute, consumers looking to purchase a safe vehicle may also gather reputable information from published safety ratings. When these ratings contain objective information about the role of weight in vehicle accidents, the hypothesized vehicle arms race can be exacerbated.

Thus, a socially efficient testing approach may be one that excludes these effects.

This paper explores the efficiency improvements of the current safety ratings methodology which holds constant the role of weight in crash tests. We compare demand for vehicles under the current ratings system to a counterfactual rating which incorporates the marginal effects of weight on fatality risk. While the underlying function describing a safety rating may not be directly observed by the consumer, such a ratings system would create an increased, implied preference for weight by ranking heavier vehicles higher, holding other factors constant. Our estimates show that implied preferences for weight would increase by 10-13 percent following the change in methodology, increasing the demand for Trucks and SUVs by 2 and 6 percent, respectively.

The findings in this paper illustrate the manner in which demand for weight may be exogenoulsy influenced by the ratings agency. Although manipulation of preferences is not likely an objective, the controlled setting of the crash tests provide clear benefits by explicitly holding vehicle size constant. While standard wall tests only simulate collisions between equal sized vehicles, they are far less costly than an alternative which, for example, crashes vehicles of different sizes together. An important consideration is whether ratings agencies would have an incentive to adopt weight-incorporated ratings should testing become more affordable. High demand by consumers and insurance companies for an accurate depiction of risk could easily drive testing methodology in that direction; a practice, that this paper illustrates, could produce costly outcomes, potentially exacerbating the effect of a vehicle arms race.

References

Anderson, Michael. 2008. "Safety for Whom? The Effects of Light Trucks on Traffic Fatalities." *Journal of Health Economics* 27 (4):973–989.

Anderson, Michael L. and Maximillian Auffhammer. 2014. "Pounds That Kill: The External Costs of Vehicle Weight." *Review of Economic Studies* 82 (2):535–571.

Bento, Antonio, Kenneth Gillingham, and Kevin Roth. 2017. "The Effect of Fuel Economy Standards on Vehicle Weight Dispersion and Accident Fatalities." *NBER Working Paper No. 23340* .

Berry, Steven, James Levinsohn, and Ariel Pakes. 1995. "Automobile Prices in Market Equilibrium." *Econometrica: Journal of the Econometric Society* :841–890.

Crandall, Robert W. and John D. Graham. 1989. "The Effects of Fuel Economy Standards on Automobile Safety." *Journal of Law and Economics* 32:97–118.

Evans, Leonard. 2001. "Causal Influence of Car Mass and Size on Driver Fatality Risk." *American Journal of Public Health* 91 (7):1076–1081.

Jacobsen, Mark R. 2013. "Fuel Economy and Safety: The Influences of Vehicle Class and Driver Behavior." *American Economic Journal: Applied Economics* 5 (3):1–26.

Li, Shanjun. 2012. "Traffic Safety and Vehicle Choice: Quantifying the Effects of the 'Arms Race' on American Roads." *Journal of Applied Econometrics* 27:34–62.

Peltzman, Sam. 1975. "The Effects of Automobile Safety Regulation." *Journal of Political Economy* 83:677–726.

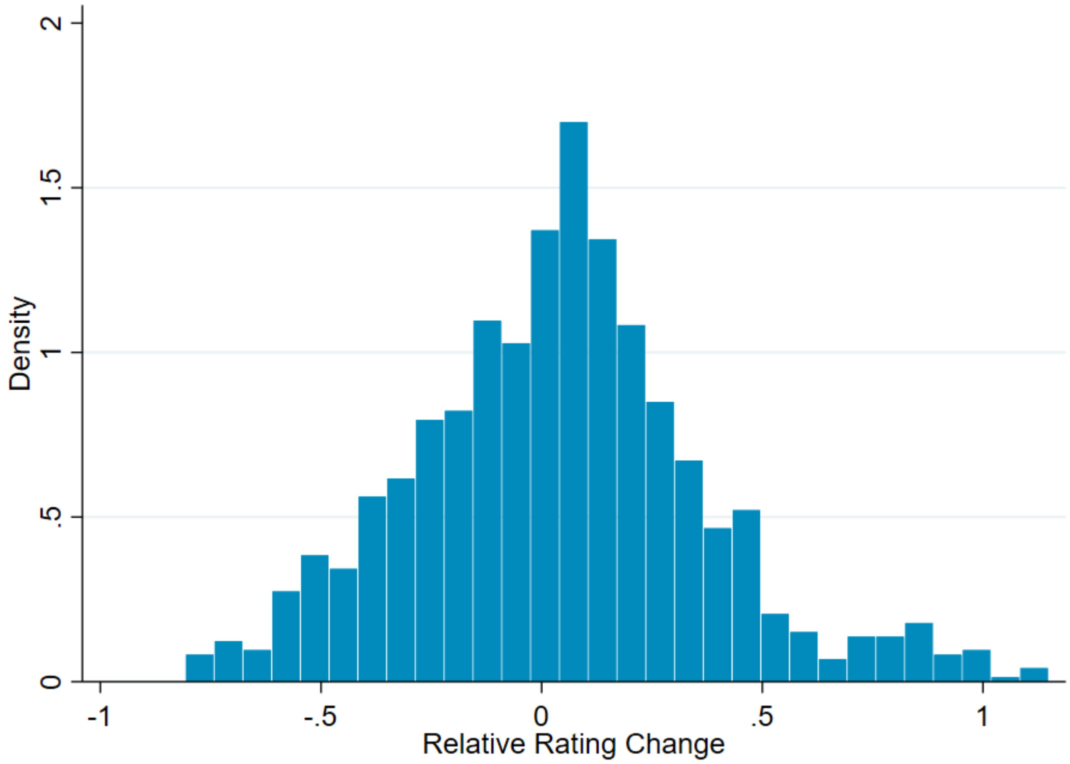
Scott, Jonathan B. 2022. "Pounds That Save: The Role of Preferences for Safety in Demand for Large Vehicles." *Journal of Law and Economics* 65 (3):555–579.

Van Auken, M. and J. W Zellner. 2005. "An Assessment of the Effects of Vehicle Weight and Size on Fatality Risk in 1985 to 1998 Model Year Passenger Cars and 1985 to 1997 Model Year LIght Trucks and Vans." *SAE International* .

White, Michelle J. 2004. "The 'Arms Race' on American Roads: the Effect of SUVs and Pickup Trucks on Traffic Safety." *Journal of Law and Economics* 47:333–356.

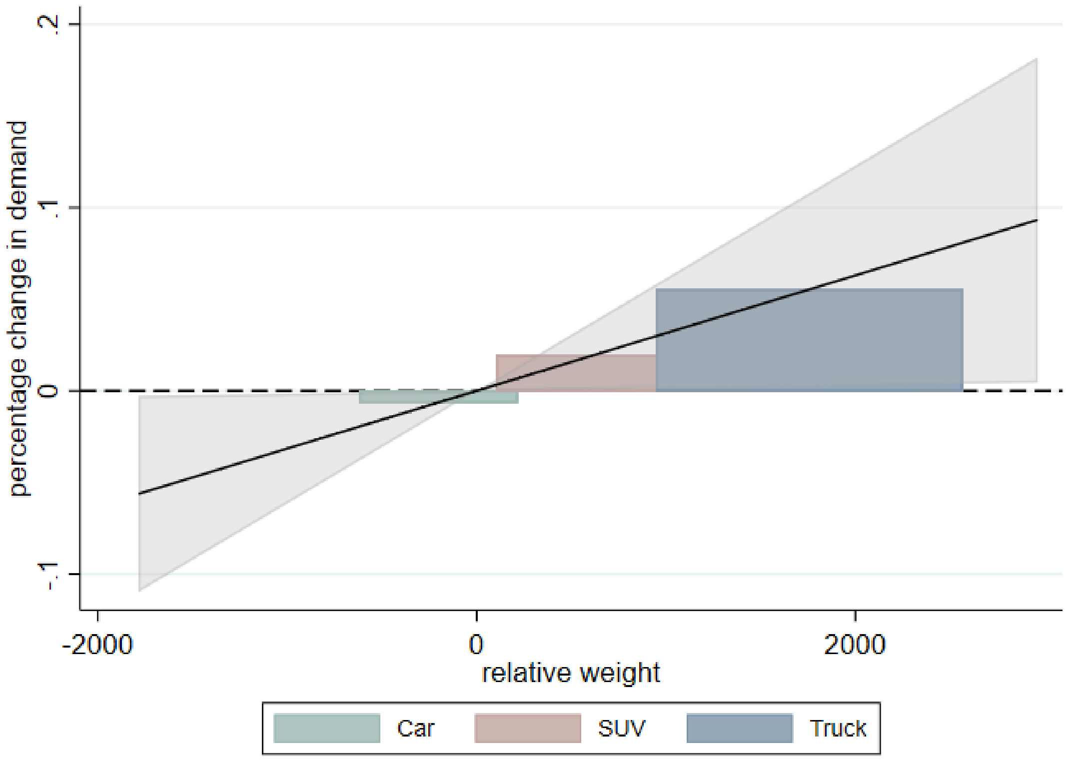
Figures

Figure 1: Relative Change in Safety Ratings



Note: The figure shows the distribution of rating changes across all models. We define the relative change in rating of a vehicle j as $\Delta_j = top_j \cdot (1 - top^*_j) + (1 - top_j) \cdot top^*_j$

Figure 2: Additional Effect of Weight on Demand Through Ratings Change



Note: This figure shows the treatment effect of the weight-incorporated safety ratings across vehicle weights. A bar plot depicts the average effect for vehicles classified as a "Car," "SUV," or a "Truck." The width of each bar represents the interquartile range of relative weight for each vehicle type, and weights are evaluated against the average (sales-weighted) curb weight in our data set

Tables

Table 1: IIHS Top Pick Criteria

Model Year:	2015	2016	2017	2018	2019
Test 1: Driver Side Small Overlap Front	Acceptable	Good	Good	Good	Good
Test 2: Passenger Side Small Overlap Front	N/A	N/A	N/A	N/A	Acceptable
Test 3: Moderate Overlap Front	Good	Good	Good	Good	Good
Test 4: Original Side	Good	Good	Good	Good	Good
Test 5: Roof Strength	Good	Good	Good	Good	Good
Test 6: Head and Seat Restraint Test	Good	Good	Good	Good	Good
Test 7: Headlight Rating	N/A	N/A	N/A	Acceptable	Acceptable
Test 8: Front Crash Prevention Technology	N/A	Basic	Advanced	Advanced	Advanced

Note: Above is the minimum rating requirement in each model year for the 8 primary crash tests to achieve IIHS top pick status. Tests 1-7 are evaluated on a scale: Poor, Marginal, Acceptable, Good. Column 8 is evaluated on a scale: Not Equipped, Basic, Advanced, Superior. "N/A" indicates that there is no minimum requirement for the test.

Table 2: Logit Model of Vehicle Choice

	(1)	(2)	(3)	(4)	(5)
Top Pick	0.0977***	0.0497***	0.0504***	0.0514***	0.0953***
	(0.0150)	(0.0136)	(0.0141)	(0.0139)	(0.0251)
Curb Weight			0.232***	0.245***	0.851**
(1,000 lbs)			(0.0779)	(0.0845)	(0.397)
MPG			0.0250***	0.0252***	0.0371***
			(0.00348)	(0.00351)	(0.00783)
Horsepower			0.00949	0.0243	0.700*
(100s)			(0.0380)	(0.0416)	(0.407)
MSRP				-0.0214	-1.003*
($10,000s)				(0.0273)	(0.591)
Crash Test Dummies	-	Yes	Yes	Yes	Yes
Make × Model FE	Yes	Yes	Yes	Yes	Yes
City × Year FE	Yes	Yes	Yes	Yes	Yes
Prices	-	-	-	OLS	2SLS
Observations	36,466	36,466	36,466	36,466	36,466

Note: * $p < 0.1$, ** $p < .05$, *** $p < .01$. Standard errors clustered by city to account for correlation in unobserved preferences within geographic boundaries. The outcome is log purchases of a vehicle model in a given city. This table shows the marginal effect of top safety picks on city-level vehicle demand across different model specifications.

Table 3: Effect of Top Pick on Fatality Risk

	(1)	(2)	(3)	(4)	(5)	(6)
Top Pick	-0.00008	-0.00013*	-0.00017**	-0.00017**	-0.00017**	-0.00016
	(0.00006)	(0.00008)	(0.00007)	(0.00007)	(0.00008)	(0.00012)
Opposing Vehicle Weight			0.00011***	0.00008***	0.00008***	0.00008***
(1,000 lbs)			(0.00002)	(0.00002)	(0.00002)	(0.00002)
Curb Weight			-0.00027***	-0.00025***	-0.00026**	-0.00025
(1,000 lbs)			(0.00006)	(0.00009)	(0.00012)	(0.00034)
MSRP				-0.00001	0.00002	0.00002
($10,000s)				(0.00005)	(0.00007)	(0.00011)
Speed Limit				0.00004***	0.00004***	0.00004***
				(0.00000)	(0.00000)	(0.00000)
Vehicle-type × Model-year FE	Yes	Yes	Yes	Yes	Yes	Yes
Vehicle-type × Make FE	-	-	-	-	Yes	-
Make-model FE	-	-	-	-	-	Yes
Year × County FE	Yes	Yes	Yes	Yes	Yes	Yes
Top Pick	OLS	2SLS	2SLS	2SLS	2SLS	2SLS
Observations	640,820	640,820	640,820	640,820	640,820	640,816

Note: * $p < 0.1$, ** $p < .05$, *** $p < .01$. Standard errors clustered at the make-model level. The outcome variable, Fatality, is a binary variable that equals 0 for no death count and 1 for a positive number of death counts. This table shows a mapping between safety ratings and fatality risk across different model specifications. We exploit the estimated ratio of the weight coefficient to the top safety pick coefficient ($\frac{\gamma}{\delta}$) to construct counterfactual safety ratings following Equation 3.

Table 4: Vehicles with Largest Rating Change

Make-Model	Curb Weight (lbs)	Top Pick	Counterfactual safety rating
Model Year: 2015 (mean weight = 3,626)			
kia soul	2,800	1	.634
subaru brz	2,800	1	.627
honda fit	2,600	1	.546
chevrolet spark	2,400	1	.438
ford expedition el	5,900	0	1.032
gmc yukon xl	5,700	0	.925
ford expedition	5,600	0	.891
gmc yukon	5,500	0	.817
chevrolet tahoe	5,300	0	.756
Model Year: 2016 (mean weight = 3,655)			
honda civic	2,800	1	.623
kia soul	2,800	1	.605
chevrolet sonic	2,800	1	.603
scion ia	2,400	1	.436
lincoln navigator l	6,100	0	1.095
cadillac escalade esv	5,900	0	1.005
ford expedition el	5,900	0	.998
lincoln navigator	5,900	0	.983
gmc yukon xl	5,700	0	.92
cadillac escalade	5,700	0	.905
Model Year: 2017 (mean weight = 3,678)			
hyundai elantra gt	2,900	1	.634
honda civic	2,800	1	.626
mazda cx3	2,800	1	.623
hyundai elantra	2,800	1	.61
toyota yaris ia	2,400	1	.426
nissan titan xd	6,300	0	1.149
lincoln navigator l	6,200	0	1.099
ford expedition el	5,900	0	.986
lincoln navigator	5,900	0	.976
cadillac escalade esv	5,900	0	.970
gmc yukon xl	5,800	0	.927
Model Year: 2018 (mean weight = 3,715)			
kia forte	2,800	1	.608
hyundai elantra	2,800	1	.599
nissan kicks	2,700	1	.529
kia rio	2,700	1	.527
cadillac escalade esv	6,000	0	.995
gmc yukon xl	5,700	0	.903
cadillac escalade	5,700	0	.880
chevrolet suburban	5,700	0	.860
nissan titan	5,600	0	.834
gmc yukon	5,500	0	.800
Model Year: 2019 (mean weight = 3,764)			
hyundai veloster	2,800	1	.582
kia rio	2,700	1	.534
hyundai accent	2,700	1	.518
nissan kicks	2,700	1	.507
gmc yukon xl	5,800	0	.888
cadillac escalade	5,700	0	.849
chevrolet suburban	5,700	0	.843
nissan titan	5,600	0	.829
gmc yukon	5,500	0	.790
toyota tundra	5,400	0	.743

Note: This table presents vehicle models with the largest relative change in ratings after accounting for weight. We define the relative change in ratings of a vehicle model as: $\Delta_j = top_j \cdot (1 - top_j^*) + (1 - top_j) \cdot top_j^*$.

Table 5: Effect of Weight-Incorporated Safety Ratings

	(1)	(2)	(3)	(4)	(5)	(6)
Combined parameter	0.03149*	0.02166	0.02209	0.03191*	0.02195	0.02239
	(0.01818)	(0.01473)	(0.01680)	(0.01741)	(0.01371)	(0.01591)
Demand Model:						
Crash Test Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Make × Model FE	Yes	Yes	Yes	Yes	Yes	Yes
City × Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls	-	-	-	Yes	Yes	Yes
Crash Model:						
Vehicle-type*Model-year FE	Yes	Yes	Yes	Yes	Yes	Yes
Vehicle-type*Make FE	-	-	Yes		-	Yes
Year*County FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls	-	Yes	Yes	-	Yes	Yes

Note: * $p < 0.1$, ** $p < .05$, *** $p < .01$. Standard errors in parenthesis were derived using the delta method and covariance matrices from the demand and crash models. This table reports the primary parameter of interest, $\alpha\frac{\gamma}{\delta}$, which is the additional demand for weight arising from the new safety rating methodology. α is the marginal effect of top safety pick on vehicle demand obtained from Equation 1. $\frac{\gamma}{\delta}$ is the ratio of the weight coefficient to the top safety pick coefficient following Equation 3. Columns 1-3 present estimates combining specification 2 of Table 2 with Columns 3-5 of Table 3. Columns 4-6 combine Column 3 of Table 2 with Columns 3-5 of Table 3