Valuing Safety: Understanding Household Willingness to Pay for Autonomous Driving Features

Jia Yan

April 3, 2024

1 Research Questions

Autonomous vehicles hold promise as a transformative solution for improving both mobility and road safety. This project focuses on quantifying households' heterogeneous preferences for AI-driven safety features, which are integral steps toward achieving fully autonomous vehicles. We utilize advanced estimation approaches that combine the strengths of modern machine learning methods with traditional econometric models. By integrating causal inference into structural models and analyzing a comprehensive dataset encompassing vehicle ownership and purchasing behavior from over 1.5 million households in Texas from 2009 to 2020, complete with detailed socio-demographic data and more than 80 vehicle attributes, we aim to address the following key questions:

1. What are households willing to pay for AI-driven safety features to reduce their risk of fatal and non-fatal accidents?

2. Is their willingness-to-pay (WTP) consistent with the technological performance of AI-driven safety features in improving safety?

3. What are the total benefits to households from AI-driven safety features and can public policy play a constructive role to facilitate those benefits?

2 Data Summary

We constructed a comprehensive dataset of Texas households spanning from 2010 to 2020, sourced from AIB, Incorporated. The sample encompasses residents of the seventeen counties across the Houston-Galveston-Brazoria, Dallas-Fort Worth, Austin, and El Paso regions. It includes the households' addresses, as well as the model year, make, model, and trim of their automobiles. Additionally, socioeconomic data such as income, family composition, and other demographic details were acquired from Acxiom.

We identified new-vehicle purchases by tracking the yearly changes in vehicle holdings and the model years of the vehicles owned by these households. The resulting dataset

contains records for each year t from M_t households, of which N_t households completed new-vehicle purchases (vehicles ranging from the previous to the following model year). A vehicle product is defined by a specific combination of model year, make, model, and trim. Table 1 summarizes the number of the sampled households, new-vehicle sales, number of vehicle products with positive sales in each year. The last column of Table 1 presents the number of vehicle products, categorized by model year, as observed in the households' vehicle holdings. Under our definition, the number of new-vehicle products in a calendar year is the sum of the number of products from the previous to the following model year. For example, the total number of new-vehicle products in 2009 is 679 + 631 + 677 = 1,987, compared with 1,077 that had positive sales.

Year	No. of sampled households	No. of new-vehicle sales	No. of new vehicles with sales	No. of held-vehicles by model year
2001	N.A.	N.A.	N.A.	494
2002	N.A.	N.A.	N.A.	538
2003	N.A.	N.A.	N.A.	545
2004	N.A.	N.A.	N.A.	579
2005	N.A.	N.A.	N.A.	560
2006	N.A.	N.A.	N.A.	574
2007	N.A.	N.A.	N.A.	654
2008	N.A.	N.A.	N.A.	679
2009	49,901	8,565	1,077	631
2010	72,970	10,696	1,148	677
2011	83,385	11,895	1,237	756
2012	98,859	16,288	1,413	751
2013	110,545	19,894	1,462	776
2014	129,180	23,721	1,515	764
2015	144,588	24,825	1,548	820
2016	166,879	28,377	1,659	802
2017	196,537	27,689	1,700	800
2018	187,604	20,223	1,522	731
2019	155,129	9,928	1,190	646
2020	166,584	3,437	740	360

Table 1: **Overview of Data**. New vehicles in a year are vehicles from the previous model year to the following model year.

The attributes of each vehicle product in our sample were constructed using a commercially available Vehicle Identification Number (VIN) decoder. This tool enables us to identify vehicles down to the trim level, which represents different versions of the same make and model, each with distinct features. Variations in safety features at the trim level play a key role in identifying households' willingness to pay (WTP) for these features.

Figure 1 shows the trend in the share of vehicles equipped with various safety features

related to driver assistance across model years, spanning from 2001 to 2020. The Society of Automotive Engineers has classified these features as Level 1 (L1) autonomous driving features. The graph illustrates a clear trend: the share of vehicles equipped with L1 features has increased significantly over the years, especially from around 2013-2016 onwards. Among these features, the automatic emergency brake is approaching universal adoption, whereas driver alert warnings and the self-driving package remain less common.

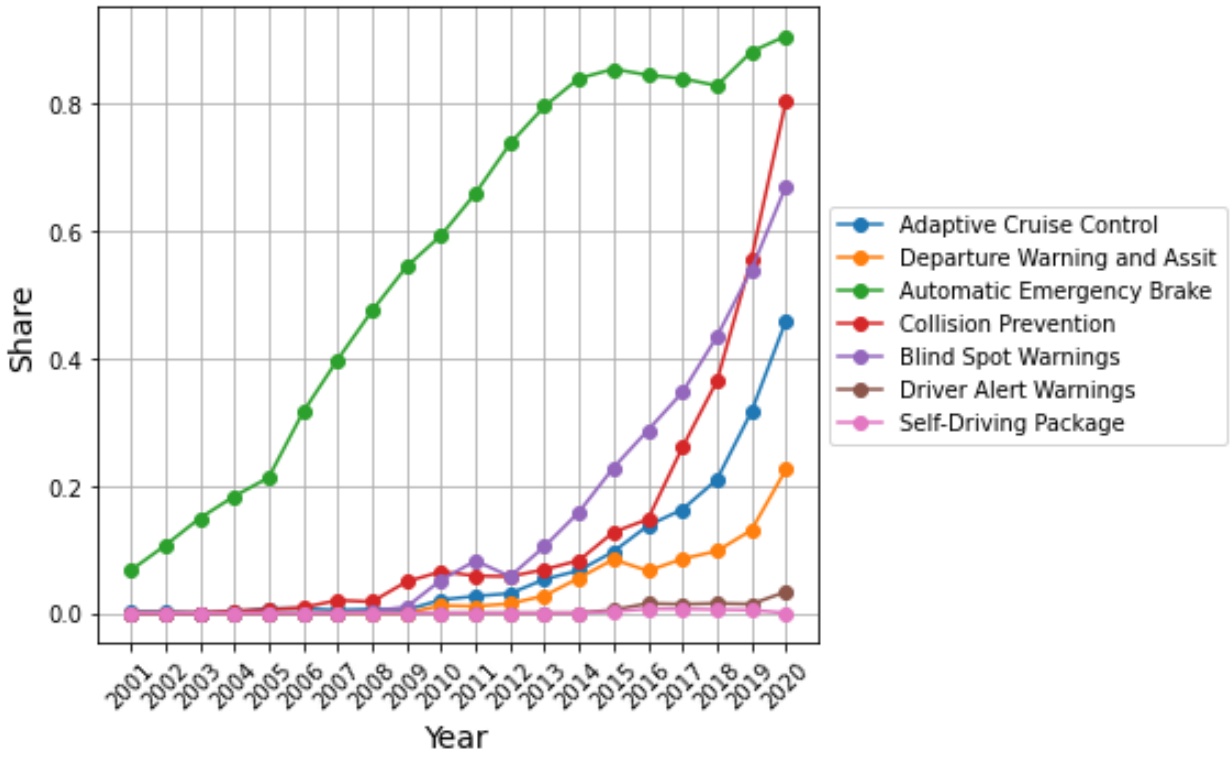


Figure 1: Trends in Vehicle Safety Features Across Model Years

Figure 2 illustrates the proportion of vehicles equipped with Level 1 safety features in new-vehicle sales, segmented by calendar year. The shares of new vehicles sold in a calendar year that are equipped with Level 1 safety features are slightly lower than the shares of vehicles offered with these features in the same model year. This suggests that consumer adoption of new technology lags behind technological advancements.

3 Methodology

The core of our project is a discrete choice model adapted to the data-rich environment and focusing on new-vehicle purchases, through which we estimate the varied household preferences for safety features and conduct counterfactual policy simulations grounded in the structural model. To address the significant challenges encountered in estimating and validating choice models with our compiled data, we embrace recent advancements in empirical

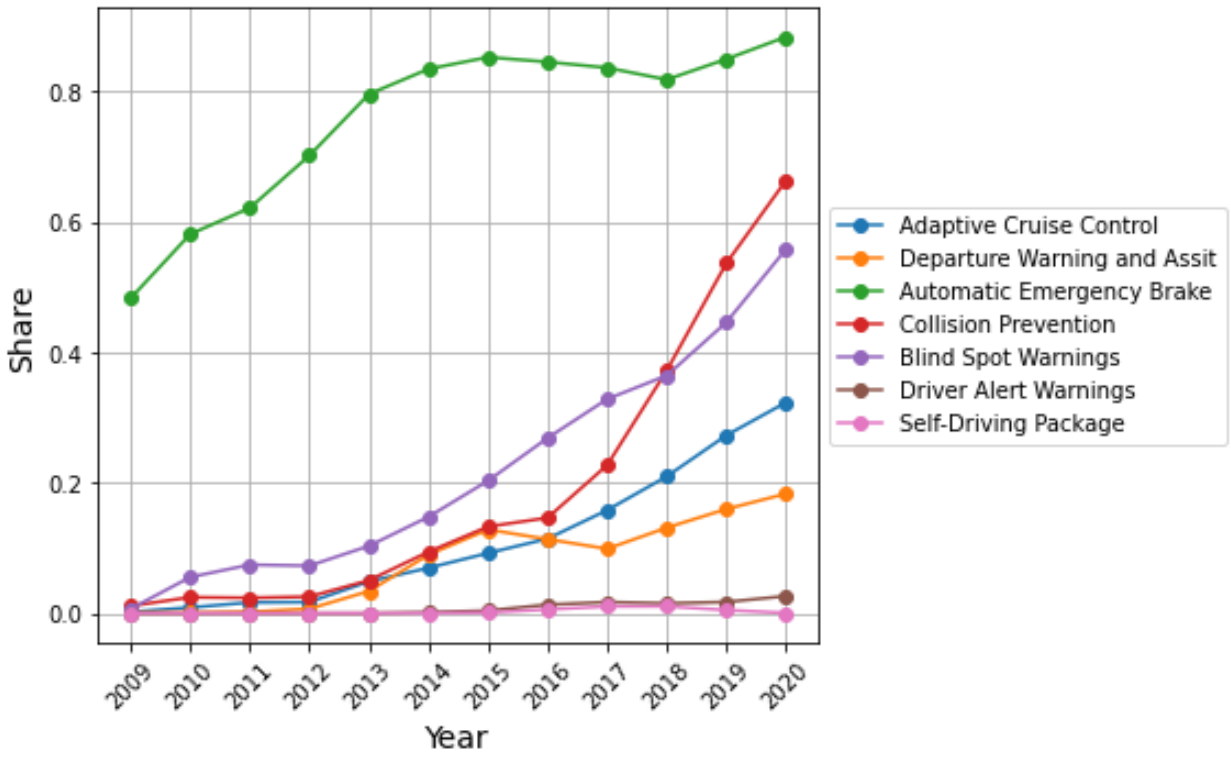


Figure 2: Evolution of Safety Feature Adoption in New Vehicle Sales

methods. These include enhanced causal inference techniques for causal parameters, and the integration of machine learning with traditional econometric analysis. In the sections that follow, we delineate the principal challenges encountered in our estimation process and detail the methodologies employed to address each of these challenges.

3.1 Accounting for Unobserved Product Attributes

A typical challenge in estimating discrete choice models is the presence of unobserved product attributes that may correlate with both price and observed product features. In our modeling, we partition the vehicle attributes, excluding price, into two groups: model attributes, such as make and model information, which are unlikely to change from year to year; and option packages, including the safety features that are the focus of this project. These are add-on features in addition to the make, model, and trim. We allow both the price and the optional packages to be correlated with unobserved attributes.

Given the detailed information on vehicle attributes down to the trim level, we will construct group fixed effects at the make, model, and trim levels to control for biases from unobserved product attributes. Furthermore, we will construct instruments for the endogenous price and option packages by following the extensive literature on applying the BLP (Berry, Levinsohn, and Pakes, 1995) [1] approach to estimate choice models. We will employ a control function type estimator (Petrin and Train, 2009) [6] using these instruments to ad-

dress any remaining concerns about endogeneity. To mitigate concerns regarding the strong assumptions inherent in the control function estimator, we will employ modern machine learning approaches to estimate the control function non-parametrically

3.2 Accounting for Large Choice Set

The large choice set in our data entails several consequences. First, estimating a discrete choice model with thousands of products is computationally demanding. Random sampling from the choice set, as first proposed by McFadden (1976) [5], has been extensively utilized in machine learning methods for classification problems, effectively addressing this issue. Second, the large choice set suggests heterogeneity in households' choice sets. Including irrelevant alternatives into a household's choice set can introduce bias in parameter estimation (Crawford, Griffith and Iaria 2021) [4].

We will utilize a factor specification to model the substitution patterns among the large number of choice alternatives in a flexible manner. This approach will also enable choice set formulation and substitution pattern to vary according to observed socio-demographic information of households and unobserved sources.

3.3 Model Validation

We will enhance the credibility of the estimated structural choice model by validating its out-of-sample prediction performance through experiments constructed based on causal inferences. The rapid advancements in technology and consumers' adoption of new technologies, as depicted in Figures 1 and 2, provide an opportunity for us to identify the causal effects of these technological advancements on market competition outcomes. Given the large size of our dataset and detailed information on vehicle attributes, we will employ the robust double machine learning approach (Chernozhukov et al., 2017, 2018) [2] [3] to identify the effects of technological advancements on the changes in market shares of the same vehicle products across different markets (vehicles of a model-year are included in different markets of calendar years). The identified causal effects will be used to validate the estimated structural choice model.

4 Importance of the Findings for the US Department of Transportation

Autonomous vehicles are coming, and it is vital for DOT to understand their potential benefits from improving highway safety and households' value of those safety improvements. The gradual availability of Level 1 autonomous vehicle safety attributes in the last decade enables us to determine households' value of those attributes in a market setting where they are free to choose new vehicles that offer those attributes. However, estimating households' value of those attributes using a vehicle choice model encounters important methodological issues that must be addressed or the estimates of households' willingness-to-pay (WTP) for Level 1 safety attributes will be biased. This study will use appropriate data and methodologies

to obtain unbiased estimates of households' WTP. The findings will be important to DOT's mission to improve highway safety by enabling DOT officials to assess whether the market is generating effective adoption of new Level 1 safety technologies—that is, households' value of those technologies is consistent with the technologies' benefits and enables automakers to set prices for them that cover their costs. In addition, our estimates would inform DOT about whether they should consider intervening in the market adoption of safety technologies if, for example, households appear to undervalue the benefits of those technologies.

5 Research Team

This research will be conducted by a three person team as follows: **PI**: Jia Yan, Professor, School of Economic Sciences, Washington State University. **Co-Investigator**: Clifford Winston, Senior Fellow, Economic Studies, The Brookings Institution. **Research Assistance**: Economics PhD Student, Washington State University

6 Conflict of Interest Statement

Neither the PI nor the co-investigator have any relevant conflicts of interest to disclose.

References

[1] Steven Berry, James Levinsohn, and Ariel Pakes. Automobile prices in market equilibrium. *Econometrica*, 63(4):841–890, 1995.

[2] Victor Chernozhukov, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, and Whitney Newey. Double/debiased/neyman machine learning of treatment effects. *American Economic Review*, 107(5):261–265, 2018.

[3] Victor Chernozhukov, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins. Double/debiased machine learning for treatment and structural parameters. *Econometrics Journal*, 21(1):1–68, 2018.

[4] Gregory Crawford, Rachel Griffith, and Alessandro Iaria. A survey of preference estimation with unobserved choice set heterogeneity. *Journal of Econometrics*, 222(1):261–265, 2021.

[5] Daniel McFadden. Quantal choice analysis: A survey. *Annals of economic and social measurement*, 5(4):363–390, 1976.

[6] Amil Petrin and Kenneth Train. A control function approach to endogeneity in consumer choice models. *Journal of Marketing Research*, 47(1):3–13, 2010.