Dispersion, Discrimination, and the Price of Your Pickup*

Stephan Sagl†

this version: November 14, 2023

Please click here for the latest draft.

Abstract

Price discrimination is a pervasive, yet controversial feature of the automobile market. Using repeat purchase data on pickup trucks, fully controlling for product heterogeneity, I establish that the same consumers pay persistently high or persistently low prices across vehicle purchases. This result suggests that dealers can learn persistent consumer preferences beyond coarse demographics through direct interactions with consumers and use them for pricing. Using a novel discrete choice model of supply and demand, I study the role of consumer information available to firms in the welfare effects of price discrimination. To do so, I overcome a common problem in settings with transaction data: consumer-specific prices of non-chosen alternatives are unobservable. I solve this problem by recovering unobserved consumer-specific prices and consumer-specific price sensitivity from the observed transaction price via firms' first-order conditions. I simulate two counterfactuals: uniform pricing and price discrimination based on coarse demographic groups – gender, race, and income. Compared to uniform pricing, price discrimination with consumer-specific prices increases profits and total welfare but, on average, harms consumers. On the other hand, compared to uniform pricing, price discrimination based only on demographics is not profitable. This suggests that information beyond demographic groups drives the profitability of price discrimination.

JEL Codes: D43, L13, L62

*I thank Paul Grieco, Charles Murry, Joris Pinkse, Mark Roberts, Connor Ryan, Bradley Setzler, and Karl Schurter for invaluable guidance. I thank all Penn State IO brown bag participants for helpful comments. This paper relies on data on sticker prices and manufacturer-installed options. I thank BlackBook and a leading online used car retailer for providing me with these data.

†Department of Economics, The Pennsylvania State University, stephan.sagl@psu.edu.

1 Introduction

Price discrimination is a pervasive, yet controversial feature of the US automobile market. However, whether price discrimination benefits firms or consumers is theoretically ambiguous. In particular, price discrimination can benefit firms, some consumers, or all consumers. The direction of the welfare effects depends on two factors: the precise form of consumer heterogeneity and the consumer information firms use for pricing. Academic literature has focused on studying price discrimination by assuming firms leverage coarse consumer information such as demographic groups for pricing and found generally small or insignificant effects (e.g. Langer, 2016; D'Haultfoeuille, Durrmeyer and Février, 2019). In this paper, I provide novel evidence that firms use much more granular, individual-level consumer information in their pricing decisions, and develop a novel framework to study the role of consumer information in the welfare effects of price discrimination.

Using data on pickup truck purchases in Texas, I make three main contributions. First, leveraging data on consumers who bought trucks repeatedly and fully controlling for product heterogeneity, I establish that firms set individual consumer-specific prices.[1] After controlling for pickup heterogeneity, around 25% of price dispersion is due to price discrimination based on consumer characteristics. However, only 0.5% of price dispersion is due to demographics observable to the econometrician. Second, I solve a key problem in demand estimation models in which prices are consumer-specific: the consumer-specific prices of non-chosen alternatives can never be observed. I overcome this problem by developing and estimating a novel discrete choice model of supply and demand that recovers unobserved consumer-specific prices and consumer-specific price sensitivity from the observed transaction price exploiting firms' pricing first-order conditions. I do not make parametric assumptions on the joint distribution of price

[1] Consumer-specific prices are also referred to as *personalized prices* (e.g. Buchholz et al., 2022; Dubé and Misra, 2023)

sensitivities and demographics. Price sensitivities can depend arbitrarily on consumer demographics. Third, I use this model to study the role of information in the welfare effects of price discrimination. I find that consumer-specific, personalized prices raise profits by 20% relative to uniform pricing, increase total surplus, and harm consumers on average. However, how much consumer information firms use for pricing is crucial for the direction of the welfare effects. If firms only have information about coarse demographic groups, such as gender, race, and income, profits are 0.3% lower relative to uniform pricing.

I establish these findings by leveraging new data on pickup truck purchases linked to both the actual buyers and the actual pickup truck characteristics including installed optional equipment. Using data on sticker prices inclusive of options and delivery fees, I find that different consumers pay different prices for the same good. After controlling for optional equipment and delivery fees, the average within-model standard deviation is around $4,100. Leveraging data on repeat purchases, I use Abowd, Kramarz and Margolis (1999)'s two-way fixed effects estimator to establish that the same consumers pay persistently high or persistently low prices across vehicle purchases, controlling for product characteristics. Around 25% of price variation after controlling for product heterogeneity is due to price discrimination. This shows that persistent consumer preferences are known to dealers and used when making price offers.

I find that on average, and after controlling for product features, women and Hispanic consumers pay significantly higher prices for the same pickup truck. In particular, women and Hispanics pay around $900 and $500 more for the same pickup truck, respectively. However, price discrimination based on protected classes only explains a small portion, around 0.2%, of observed price dispersion after controlling for product heterogeneity. Furthermore, even a wide array of consumer demographics, including income, only explains 0.5% of price variation after controlling for product heterogeneity. This suggests that dealers have a large amount of consumer information beyond coarse demographics which they may glean from test drives, conversations, credit scores, etc. This finding is reminiscent of the use of personalized pricing - pricing based on a rich set of consumer features in the marketing literature (e.g. Dubé and Misra, 2023).

To understand the importance of consumer information available to firms in the welfare effects of price discrimination, I estimate a structural model of supply and demand where dealers observe key consumer preferences and price discriminate accordingly. The model flexibly incorporates consumer heterogeneity via observed demographics and unobserved tastes, including price sensitivity. The key challenge complicating estimation is that the consumer-specific prices of non-chosen alternatives can never be observed. To overcome this problem, I use firms' pricing first-order conditions to recover unobserved prices of non-chosen alternatives and price sensitivities directly from the observed transaction prices. This allows me to learn the joint distribution of price sensitivities and demographics for all consumers in the market while avoiding making parametric assumptions on the joint distribution.

I use this model to evaluate the role of consumer information in the welfare effects of price discrimination. First, I consider the elimination of price discrimination. It is well known that when firms compete, price discrimination may increase or decrease consumer surplus, depending on the precise form of consumer heterogeneity. In addition, price discrimination may also decrease firms' profits (Corts, 1998). I find that all producers benefit from price discrimination in the pickup truck market, and consumers are harmed on average.

Secondly, I illustrate the importance of consumer information firms use for pricing for the direction of the welfare effects of price discrimination. To do this, I consider a counterfactual where firms price only based on observable consumer demographics but do not use additional information about consumers. In this case, all firms lose, and industry profits are lower by around 0.3% relative to uniform pricing. This demonstrates that the welfare effects of price discrimination are highly sensitive to the amount of consumer information available to firms for pricing. Furthermore, these results show that dealers use much richer information than gender and race to price discriminate. In particular, most of the price dispersion and profits stem from dealers leveraging the information they acquire through direct interactions with consumers rather than coarse demographics.

Why does price discrimination not lead to lower prices in the pickup truck market? Corts (1998) showed that if firms differ sufficiently in their ranking of consumers in terms of demand elasticities, best response asymmetries can induce discriminatory prices below uniform prices. This is a feature of horizontal price discrimination. My estimates imply that the drivers of price dispersion in the pickup truck market are differences in price sensitivities across consumers rather than demographics. Because all dealers rank price-insensitive consumers higher and this channel dominates, there is no sufficient asymmetry in best responses. Thus, price discrimination did not lead to lower prices in the pickup truck market.

Previous research has studied the average effects of demographics on price (e.g. Goldberg, 1996; Harless and Hoffer, 2002; Chandra, Gulati and Sallee, 2017). My estimates leveraging repeat purchases of individual consumers complement and expand on these findings. While I find mean differences in price paid across demographics, my estimates demonstrate that price dispersion is almost entirely driven by unobservable heterogeneity.

This paper also offers a contrasting perspective on recent structural work on price discrimination in the automobile market (Langer, 2016; D'Haultfoeuille, Durrmeyer and Février, 2019). Existing literature has found small effects of price discrimination on firms' profits. My results offer a stark contrast to these findings. Dealers engage in price discrimination, and it is highly profitable to them. However, in line with earlier structural research, I find that price discrimination based on protected classes or observed demographics only leads to small changes in profits relative to uniform pricing.

My work also contributes to the literature on the estimation of discrete choice demand models when either prices or quantities are unobserved (e.g. Thomadsen, 2005; Coşar, Grieco and Tintelnot, 2015). Methodologically closest to my approach is D'Haultfoeuille, Durrmeyer and Février (2019), who develop a model of supply and demand with unobserved price discrimination based on coarse demographic groups. I contribute to this literature by building and estimating an equilibrium model of supply and demand where firms set *individual* rather than group-specific prices when only the transaction price is observable.

I also contribute to the empirical literature on the welfare effects of personalized pricing. A growing subset of this literature has recently focused on personalized pricing in online markets (Dubé and Misra, 2023; Shiller, 2020, 2022). However, there has been less work on offline markets (Waldfogel, 2015; Buchholz et al., 2022). I contribute to this literature by studying an offline, multi-firm, competitive market under a baseline scenario of personalized prices. I show that personalized prices benefit firms, harm consumers on average, and increase total welfare. However, with different information structures welfare effects flip.

Lastly, I contribute to the literature of audit studies documenting price differences based on gender and race across industries, such as in the automobile market (Ayres and Siegelman, 1995), housing market (e.g. Bayer et al., 2017; Goldsmith-Pinkham and Shue, 2023), in car repairs (e.g. Busse, Israeli and Zettelmeyer, 2017), or in the mortgage market (e.g. Ambrose, Conklin and Lopez, 2020). As in the housing market literature, I find that price discrimination based on protected classes is prevalent but only explains a small fraction of price dispersion.

I organize the remainder of this paper as follows: Section 2 describes the data used in this paper and its sources. Section 3 presents evidence for price discrimination in the market for pickup trucks. Section 4 introduces the equilibrium model of supply and demand. Section 5 presents the empirical strategy and results, section 6 studies the role of consumer information in the welfare effects of price discrimination, and section 7 concludes.

2 Pickup Trucks & Data in Texas

This section describes the market for pickup trucks in Texas and the data I use for this paper.

2.1 Industry Background

The automobile market is one of the US's most important consumer goods markets. A defining feature distinguishing the US automobile market from other countries' markets is the popularity of personal-use pickup trucks. Of the around 17 million new light-duty vehicles sold across the US in 2018 (National Automobile Dealers Association, 2019*b*), nearly 2.5 million, or about 14%, were full-size pickup trucks (Drury and Caldwell, 2020). Remarkably, pickup trucks made up three of the top three most popular vehicles in the US in 2020, with the Ford F-Series claiming the top spot as America's most popular vehicle for nearly four decades.[2]

While pickup trucks are popular nationwide, pickup trucks are particularly popular in Texas. Based on my data, between 2011 and 2019, around 22% of all new vehicle sales in Texas were pickup trucks. Focusing on Texas is important because the pickup truck market in Texas is nationally significant. According to data from Experian, Texas was the biggest market for pickup trucks in the US in the first quarter of

[2]Wayland, Michael. 2021. "Pickup trucks dominate America's 10 best-selling vehicles of 2020", *cnbc.com*, January 11. https://www.cnbc.com/2021/01/08/pickup-trucks-dominate-americas-10-best-selling-vehicles-of-2020.html, accessed 2023-08-04.

2019 by volume, accounting for around 13% of all pickup truck sales in the US.[3] The vast majority of new pickup trucks in Texas, with limited exceptions, sell through franchised dealerships.[4]

In 2018, there were around 1,200 new vehicle dealerships in Texas (National Automobile Dealers Association, 2019*a*). New automobile dealerships in Texas are most commonly franchised with a single manufacturer, but sometimes larger dealerships are franchised with more than one manufacturer.[5] Dealerships are restricted to selling new pickup trucks from manufacturers they are franchised with. While franchise regulations are complex and out of the scope of this overview, Murry and Schneider (2016) argue that regulations are de facto prohibiting manufacturers from terminating dealerships unilaterally, requiring manufacturers to charge the same wholesale prices for vehicles across dealerships, and allow dealerships to charge consumers any price for automobiles they sell.

The sale of a pickup truck commonly proceeds in a standardized way. Traditionally, interested potential customers would enter the dealership's lot looking for the pickup truck they are interested in.[6] In most cases, a salesperson then approaches the customer, informing the customer about specific vehicle features, asking questions about the customer's background, and eventually offering a test drive.[7] If the customer expresses interest in a test drive, she is usually allowed to test drive the pickup truck for a limited time. After the test drive, the salesperson asks the customer to discuss the

[3] Miller, Marty. 2019. "King of the Road: Breaking Down the Popularity of Pickup Trucks", *experian.com*, August 30. `https://www.experian.com/blogs/insights/2019/08/king-road-breaking-popularity-pickup-trucks/`, accessed 2023-08-04.

[4] Recently, electric vehicle manufacturers like, for example, Rivian, have started to sell pickup trucks directly to consumers.

[5] For example, Vandergriff owns an Acura, a Chevrolet, a Honda, a Hyundai, and a Toyota dealership in Arlington, TX, located just a few feet apart next to Interstate 20.

[6] Although most new vehicle dealerships in the US offer appointments with salespersons on request, walk-ins remain popular: according to Cox Automotive (2019), around 50% of customers are walk-in. If a consumer schedules such an appointment, she usually enters the dealership to ask for the salesperson instead and proceeds from there.

[7] These interactions allow the salesperson to learn more about their consumers. Common questions are the number of children in the household, the vehicle's intended use (commute/pleasure), etc.

potential purchase terms. Discussing the terms is commonly the most time-consuming part of buying a new pickup truck, with US dealership visits averaging close to 3 hours in 2017 (Cox Automotive, 2019). During the so-called negotiation phase, the salesperson will make multiple offers and sometimes leave to "consult" with their manager behind closed doors.[8] At the end of the negotiations phase, the salesperson will typically state that the current price is the best price they can offer. If the consumer decides to buy, the salesperson usually earns a significant commission on each pickup truck sale, with literature reporting commissions as high as "20 to 30 percent of the profit margin of the dealership" (Murry and Schneider, 2016, p. 345).

The distinguishing feature of the pickup truck market is the lengthy interactions between consumers and dealers. Literature has found that quoted prices during negotiations correlate with demographics even when auditors follow the same negotiation protocol (Ayres and Siegelman, 1995). Zettelmeyer, Scott Morton and Silva-Risso (2006) suggests that the negotiations phase serves as a way for dealers to learn consumers' willingness to pay and to price discriminate. This suggests that sales personnel combine the information from test drives, credit scores, and conversations with observable consumer demographics to infer customers' willingness to pay.

2.2 Data Sources

I rely on a combination of different data sets: *transaction-level* registrations data from the Texas Department of Motor Vehicles (TX DMV), detailed data on buyer characteristics provided by *Infutor*, as well as data on pickup truck characteristics up to the manufacturer installed options obtained from leading industry sources.[9]

TX DMV Registrations data. My primary data set consists of individual

[8] I use "consult" because often there is no consultation, and the salesperson makes customers wait.

[9] I combine data from multiple industry sources to construct my pickup truck characteristics data. Only BlackBook has agreed to have its identity disclosed. Other data sources include a leading online used car retailer.

transaction-level data on all vehicle purchases for the state of Texas obtained from the Texas Department of Motor Vehicles from January 1, 2011, to October 31, 2019. Each of the observations corresponds to a unique vehicle sale. I observe the make/model of the vehicle, the month of purchase, the vehicle identification number (VIN), the transaction price, the name of the dealership selling the vehicle, its five-digit zip code, and the five-digit zip code of the buyer. I limit the sample to pickup trucks and exclude models mostly used commercially.[10] Additionally, I geo-code buyer and dealer locations to obtain purchase distances. These data consist of 1,433,657 transactions spanning the whole state and nine years.

infutor Consumer data. For consumer demographics, I rely on data collected by *infutor Data Solutions*, henceforth infutor. The infutor data has been proven popular for research requiring tracking individuals' US addresses over their life (e.g. Bernstein et al., 2022; Diamond, McQuade and Qian, 2019) because it provides persistent identifiers for and detailed demographic data on individuals. The infutor data is "highly representative of the overall US adult population" and "covers 78% of the overall adult US population" (Bernstein et al., 2022, p. 7). I impute race using the names of individuals contained in the infutor data using the *NamePrism* API (Ye et al., 2017; Ye and Skiena, 2019).[11] The main novelty of my approach is that I take advantage of infutor's auto data. According to Infutor Data Solutions (2023), infutor's auto data contains nearly 200 million vehicle owner records built from information on sales records, service records, auto repairs, and others. Besides information on the owner, the auto data contains basic information about the vehicle plus, importantly, the complete vehicle identification number. I use the vehicle identification number to merge the demographic data from infutor with my transaction-level data. Combining

[10] I exclude trucks such as, e.g., the Ford F-350 and F-450 because these are rarely used as personal vehicles.

[11] NamePrism is a tool to classify race and nationality based on names widely employed in academic research and can be accessed via https://www.name-prism.com

these data allows me to (i) learn the identity and, thus, the demographics of purchasers; and (ii) track automobile purchases of individuals over time.

Sticker Price Options data. In the US, by law, manufacturers must affix a label stating, among other information about the vehicle, the manufacturer's suggested retail price of the automobile and all installed optional equipment to every new vehicle before delivery (15 U.S.C. §1232). These labels are called *window stickers*.[12] Information on the window sticker allows consumers to learn the exact specifications of the pickup truck they are about to purchase. I obtain data on these window stickers from two industry sources. My primary data source is *Black Book*. Black Book is an industry-leading data provider for automobile dealerships in the US. Secondly, I supplement these data with data from a leading online used car retailer. I obtain data on approximately 1.65 million window stickers.[13] Again, I merge these data with my transaction data based on the vehicle identification number.

Merging all three data sets, I end up with 76,675 matched observations. Table 1 documents differences in the purchase price, base price, price of the optional equipment, and income across demographic groups. On average, female and minority consumers choose less expensive truck models and trims and spend less on optional equipment. These demographic groups also have lower incomes.

My data offers several advantages. First, it contains data on the actual pickup truck buyers, their demographics, and complete information on pickup truck characteristics. Second, because infutor tracks consumers over time, I can construct a panel of repeat purchases for a subset of consumers. Across all years, I identify 2,640 repeat pickup truck purchases by the same consumers. Because the repeat purchase data set is relatively small, however, I do not use it for structural estimation. In particular, I construct two different data sets from the 76,675 matched observations. First, the

[12]Window stickers are sometimes called Monroney labels.

[13]These are window stickers for all types of vehicles in the transaction level data, not just pickup trucks. Each window sticker matches one vehicle in the transaction data.

	(1) Female	(2) Male	(3) Non-Hispanic	(4) Hispanic	(5) income $\leq$ \$50k	(6) income $\in$ (\$50k,\$100k]	(7) income >\$100k
Purchase Price	36,140.54	37,650.31	37,534.96	36,907.46	36,683.47	37,051.95	38,465.21
	(7,951.64)	(8,439.60)	(8,344.76)	(8,375.34)	(8,282.32)	(8,227.94)	(8,637.09)
Base Price	37,322.01	39,286.77	39,280.95	38,031.22	37,716.22	38,648.64	40,204.43
	(7,943.15)	(8,123.13)	(8,119.83)	(8,070.98)	(8,069.30)	(8,033.74)	(8,220.08)
Options Price	5,116.72	5,660.08	5,582.44	5,465.05	5,302.06	5,521.23	5,765.59
	(3,607.76)	(3,730.93)	(3,748.43)	(3,635.27)	(3,577.39)	(3,651.53)	(3,934.78)
Income	74,467.07	82,542.82	87,809.69	66,791.61			
	(28,848.58)	(32,013.39)	(31,527.39)	(26,473.93)			
N	16,475	60,200	51,131	25,544	12,463	46,099	18,113

Table 1: Summary Statistics across Demographic Groups

repeat purchase data set which covers 2011 to 2019. Second, the *microdata for demand estimation* I use to estimate my structural model covering 2016 – 2019. The second data set does not cover all years because the merge rate before 2016 is significantly worse than for 2016 – 2019. The drop in the merge rate is due to one of my data sources only becoming an increasingly important player in the used car retailing business after 2015, significantly reducing the amount of sticker data I can match with the transaction level data.

Lastly, a commonly encountered problem in structural work on the US automobile market is that estimating a model at the trim level is impossible because the product space would be prohibitively large. For this reason, I normalize pickup trucks to the base trim in the microdata for demand estimation data set with no installed options. I remove the price of the installed options and the price effect of the trim by running a regression of transaction price on these controls for each pickup truck model separately. The normalized price is then the average price of the base trim plus the individual specific residual from this regression. This preserves the price dispersion and is similar to the homogenization of bids commonly used in the auction literature. Note that I do not need this normalization in the repeat purchase data set. The documentation of the price dispersion in the next section is unaffected.

3 Documenting Price Discrimination

In this section, I present evidence that motivates a model in which firms set individual consumer-specific prices, sometimes called *personalized prices*, based on rich consumer information. In particular, this section documents the following facts about the pickup truck market. First, different consumers pay vastly different prices for the same pickup truck model. Second, the same consumers pay persistently low or persistently high prices across vehicle purchases. Third, coarse demographics such as gender, race, or income only explain a tiny fraction of price dispersion.

Fact 1: Different consumers pay different prices for the same pickup truck

I start with documenting the first fact: different consumers pay vastly different prices for the same pickup truck model. Prices of a pickup truck model can vary because of differences in product offerings or price discrimination. I first establish that differences in product offerings are important for observed price dispersion but that they cannot fully rationalize it. While all pickup truck models in my data exhibit roughly the same patterns, focusing on a single model is easier because it eliminates between-model heterogeneity. In the following, I present statistics for all 2018 Chevrolet Silverado 1500s sold in Texas in 2018.[14]

Pickup trucks are heterogeneous products, even conditional on model and trim. The problem of product heterogeneity is well known but widely ignored in the literature because of data limitations, e.g., Berry, Levinsohn and Pakes (1995). Consumers can customize pickup truck models by selecting from various options. For the 2018 Chevrolet Silverado 1500, the average consumer chooses 16 out of the 200 available

[14] I use the Chevrolet Silverado 1500 as an example because it was the second most popular truck in the US in 2018 (Wardlaw, 2019).

options in 2018.[15]

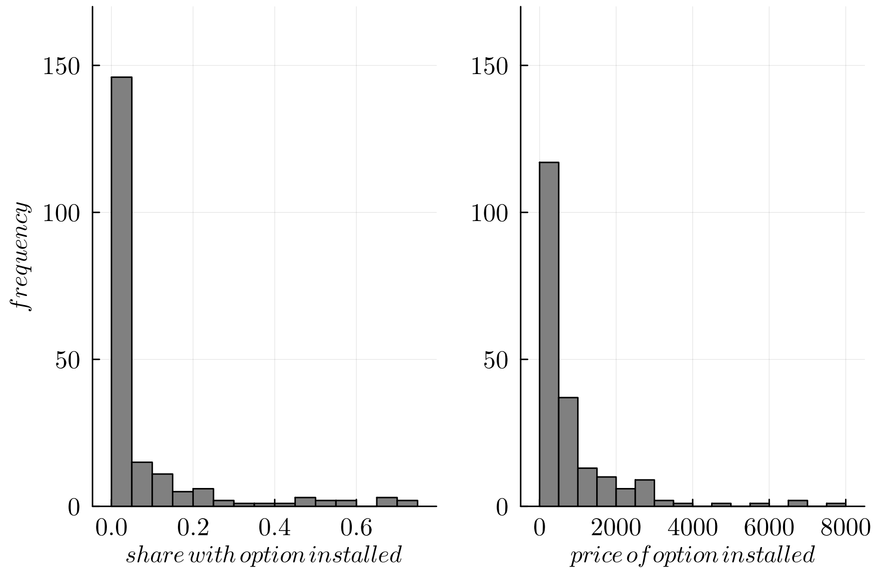

Figure 1: Distributions of optional equipment choices. Left panel plots frequency of specific options installed on share of all 2018 Chevrolet Silverado 1500s. Right panel plots distribution of the price of options.

The left panel in figure 1 shows that most options are only installed on a handful of pickup trucks. Relatively few options are installed on the majority of pickup trucks. These popular options, like 10-way adjustable driver seats, remote start ability, or electric rear windows, usually come in packages. Rarer options like spot lamps, upgraded spare wheels, or suspension options do not occur more than a few times in the data. However, less common options are, on average, more expensive.

The right panel in figure 1 documents the distribution of options prices. The mean total price for all installed options on a 2018 Chevrolet Silverado 1500 is around \$6,500. While most options cost below \$1,000, with an average of around \$825, many installed options can get very expensive, with the maximum exceeding \$7,000.

While options are important factors that drive the dispersion of transaction prices, options alone are not sufficient to explain the variation in prices. Figure 2 plots

[15] Of the 200 options, some seem to be the same but have different option codes. I therefore report them as different options.

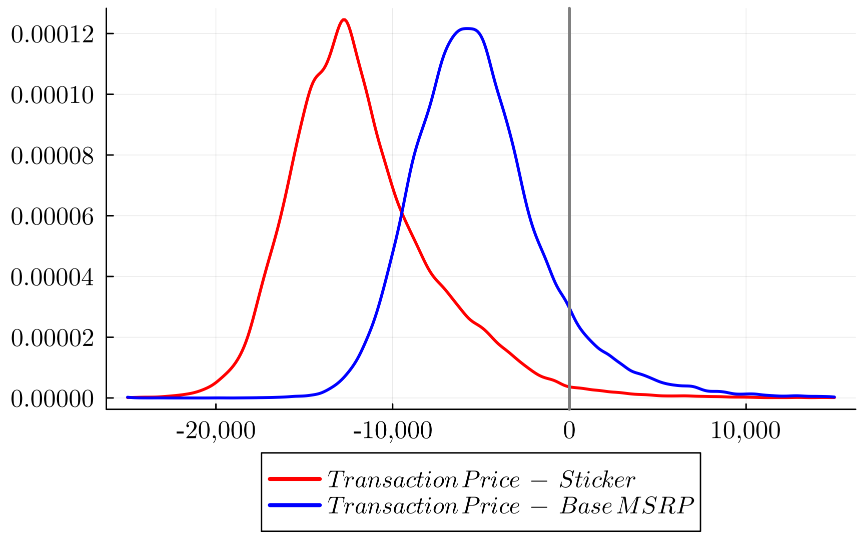

Figure 2: Density of Margins when controlling for Options (red) and when not controlling for Options (blue)

the distribution of two different margins for the 2018 Chevrolet Silverado 1500: the transaction price minus the sticker price (red),[16] and the transaction price minus the base MSRP at the trim level (blue). This figure establishes two facts. First, while often proposed as one reason for price dispersion, accounting for manufacturer-installed options slightly increases the standard deviation by around \$300 to \$4,200. Differences in transaction prices do not only stem from people opting for different levels of packages or manufacturer-installed options. Second, because sticker data fully controls for differences in product offerings, the remaining price differences are solely due to price discrimination.[17]

Fact 2: Prices are individual consumer-specific

Next, I document that pickup truck prices are individual consumer-specific, and that consumers pay persistently high or persistently low prices across vehicle purchases. To do so, I decompose the variance of the difference between the transaction prices and

[16]The sticker price is the manufacturer's suggested retail price including all installed options. It is not binding for dealers.

[17]Note that price discrimination alone does not imply that prices are individual consumer-specific.

sticker prices for all pickup truck models and all years into a consumer and product component. This allows me to explore how much of the variation in transaction prices is due to price discrimination on consumer characteristics. The repeat purchase data enable me to follow a subset of consumers across purchases. Because I observe consumers and pickup trucks repeatedly, I can decompose the variance following Abowd, Kramarz and Margolis (1999).[18] This is a novel approach in the automobile market because researchers usually only have access to cross-sectional data. Following AKM, I estimate the following equation using repeat purchases of the same customers

$$\left(\text{Transaction Price - Sticker}\right)_{ijt} = z_{it}'\beta + \eta_i + \phi_{j(i,t)} + \epsilon_{it} \tag{1}$$

where $\left(\text{Transaction Price - Sticker}\right)_{ijt}$ is the difference between transaction price and sticker price for consumer i purchasing pickup truck j at time t, z_{it} are time-varying characteristics, η_i is a consumer-specific fixed effect, and $\phi_{j(i,t)}$ are pickup truck year-model-trim fixed effects.

Before estimating this equation, I show that the repeat purchase data are consistent with existing literature studying price discrimination in the US automobile market. Literature has repeatedly found mostly statistically insignificant or small effects when regressing transaction prices on a broad set of consumer demographics and vehicle fixed effects even on the model-trim level. Table 2 shows that my repeat purchase data are consistent with these findings. Even when controlling for trim, all demographic variables fail to achieve significance at the 5% level. The lack of statistical significance arises because demographics correlate with trim levels and optional equipment. Column (4) shows that appropriately controlling for trim and optional equipment uncovers that females and Hispanics pay higher prices than males and non-Hispanics. This suggests that, on average, price discrimination based on gender and race is prevalent. However,

[18] This amounts to estimating a two-way fixed effects model with consumer and product-specific fixed effects and time-varying consumer characteristics as controls.

Table 2: Sticker prices uncover mean differences in prices paid across demographics

	Price			Price - Sticker
	(1)	(2)	(3)	(4)
Female	295.81		262.61	729.67***
	(244.79)		(246.34)	(198.09)
Hispanic		417.49*	399.18*	603.40***
		(235.23)	(236.91)	(193.54)
Income (10,000s)	46.18	59.62*	61.60*	-20.01
	(33.33)	(34.05)	(34.09)	(28.35)
Controls:				
Year-Model-Trim FE	✓	✓	✓	✓
MSA FE	✓	✓	✓	✓
N	2,280	2,280	2,280	2,280
R^2	0.86	0.86	0.86	0.72

Note: Number of observations smaller than 2,640 because some observations miss income.
Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

table 2 alone is insufficient to conclude that price discrimination is important for price dispersion or that prices are consumer-specific. In particular, it does not isolate the proportion of the variance of transaction prices driven by dealers price discriminating on demographics. Using AKM, I now go one step further demonstrating that prices are indeed individual consumer-specific, not demographic group-specific.

AKM's variance decomposition allows me to distinguish between variation in transaction prices due to differences in product offerings and price discrimination. I estimate varying specifications of equation (1). In each specification, $\phi_{j(i,t)}$ is a year-model-trim fixed effect. I present the results in table 3. My main result is that time-varying covariates do not contribute to the price dispersion but that time-invariant, consumer-specific fixed effects η_i account for around 25% of the variation in transaction price minus sticker price (column 4).[19] This suggests that dealers have access to large amounts of information about persistent consumer characteristics they use for pricing. It also implies that consumers pay persistently high or persistently low prices across vehicle purchases. This is a novel result not shown before in literature on the US automobile market.

The remaining columns investigate whether other factors also explain the variation in the transaction price. Column 2 explores whether search costs could be the driving factor of price variance by controlling for purchase distance. While search costs might be otherwise significant, variation in purchase distance does not explain the variation in the transaction price. The coefficient on purchase distance has the expected sign, but it is not significant even at the 10% level. The same is true for the financing decision of the consumer, and jointly controlling for all both variables.[20] However, even including all these variables jointly does not explain the variation in transaction

[19] Note that the sum of the fractions of variance explained do not have to sum to one because of the unreported covariances. This is a well known fact in the labor literature.

[20] I also run a separate version where I control for the month of purchase. All indicators except for two are insignificant. Results are qualitatively unchanged and magnitudes are similar. In particular, they do not effect the proportion of variance explained by consumer fixed effects and the results in table 4 are roughly the same. Hence, I do not report it.

Table 3: Decomposition of the variance following AKM

dependent variable: (price - sticker)$_{ijt}$	(1)	(2)	(3)	(4)
Fraction of variance explained:				
$z'_{it}\beta$		.00	.00	.00
individual FE η_i	.25	.25	.25	.25
pickup truck FE $\phi_{j(i,t)}$	.72	.72	.72	.72
match heterogeneity ϵ_{it}	.10	.10	.10	.10
Time varying characteristics:				
Purchase Distance (mi)		-15.35		-15.19
		(16.99)		(16.98)
Financed			789.90	786.41
			(567.61)	(567.67)
Year-Model-Trim FE	✓	✓	✓	✓
N	2,640	2,640	2,640	2,640

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4: Consumer-specific fixed effects correlate with gender and race

dependent variable: individual FE $\hat{\eta}_i$	(1)	(2)	(3)
Female	902.44***		871.47***
	(217.95)		(219.82)
Hispanic		534.05**	485.16**
		(221.57)	(221.12)
Mean Income$_i$ (\$10,000s)	0.24	12.30	20.02
	(34.62)	(36.01)	(35.88)
Mean Purchase Distance$_i$	-15.71	-17.45*	-15.15
	(9.78)	(9.85)	(9.76)
MSA FE	✓	✓	✓
N	1,150	1,150	1,150
Within R^2	0.015	0.009	0.020
R^2	0.082	0.076	0.087

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

price since the contribution to price variation is estimated to be approximately 0%.

Fact 3: Coarse demographics only explain a tiny fraction of price dispersion

Lastly, I establish that coarse demographics do only explain a tiny fraction of price dispersion. To understand the importance of price discrimination based on demographic groups, I project the estimated consumer-specific effects on demographic controls: gender, race, the mean income across purchases for customer i, and the mean purchase distance across purchases for customer i. I add fixed effects for Metropolitan Statistical Areas (MSAs) that control for differences in competition and population density. The results in table 4 show that the consumer-specific fixed effects systematically correlate with consumers' demographics. Women and Hispanic customers pay significantly more for the same pickup truck, even when fully controlling for pickup truck heterogeneity. On average, women have to pay around \$900 more for the same pickup truck than males with similar demographics. Similarly, Hispanics pay around \$500 more than non-Hispanics. These differences are approximately 2% of the average transaction price. While the mean income across purchases is relevant for whether a consumer buys any pickup truck model, conditional on purchase, it does not affect the variance of the transaction price. Lastly the parameter estimate for the mean purchase distance traveled across vehicle purchases is insignificant, showing that different search costs across consumers do not drive the observed price differences.

While documenting systematic differences based on demographics and protected classes, these results also show that price discrimination based on the demographics I observe in the microdata cannot rationalize the observed price dispersion alone. Coarse demographics only account for a vanishingly small fraction of variation in the individual-specific fixed effects. More than 90% of the variation attributable to consumer fixed effects stems from unobservable heterogeneity.[21] These are factors unobservable to the econometrician but observable to the dealer. Dealers engage in lengthy and costly consumer interactions to extract information about their willingness to pay. The information acquisition of dealers can take the form of offering test drives, offering dealer financing to learn income and credit scores, casual chats about what the truck will be used for, or conclusions based on the general appearance or behavior of the consumer during interactions. This soft information dealers learn about consumers explains the overwhelming fraction of price dispersion, and dealers leverage it to price discriminate across consumers.

The facts I presented in this section highlight the importance of the use of consumer information in firms' pricing decisions for observed price dispersion. In particular,

[21] This includes the contribution of MSA fixed effects. Demographics without MSA fixed effects only explain around 2% of the variation in the individual fixed effects.

firms use very detailed consumer information to make personalized price offers to consumers. However, whether the use of detailed consumer information benefits firms or consumers is theoretically ambiguous. To understand the role of consumer information firms use for pricing in the welfare effects of price discrimination, I need a structural model. The patterns I documented in the data motivate a structural model in which firms price discriminate on observable and unobservable consumer characteristics. I present this model in the next section.

4 Model

I develop an equilibrium model of supply and demand with individualized prices to study the pickup truck market. The model broadly falls into the class of differentiated product demand models commonly used in industrial organization (Berry, Levinsohn and Pakes, 1995). The innovation of the model is that it encompasses firms setting *individual-specific prices* based on *observed* and *unobserved* consumer heterogeneity. Thus, the model matches the main features of the pickup truck industry and is consistent with the empirical facts I presented in the preceding section. I assume that dealers and manufacturers act as integrated firms with aligned incentives, abstracting away from the vertical structure of the market. In the following, I will use dealer and firm interchangeably.

4.1 Demand

A market is defined as a year, indexed by t, and is populated by N_t consumers, indexed by i. Each consumer in the market chooses whether to buy a single pickup truck model j among the J_t available alternatives or to choose the outside option of not buying a pickup truck this year. I denote the outside option with $j = 0$ in each market t. Each consumer has full information on all products available and the corresponding prices she faces at the time of her choice.

I assume that each consumer maximizes her indirect utility by choosing either one of the J_t inside goods or opting for the outside good. If consumer i opts for pickup truck model j in market t, her indirect utility is given by

$$u_{ijt} = \delta_{jt} + \mu_{ijt} + \varepsilon_{ijt}, \tag{2}$$

where δ_{jt} is a product fixed effect for pickup truck model j, and μ_{ijt} the individual specific deviations from the product mean utility in the spirit of Berry, Levinsohn and Pakes (1995). I assume that ε_{ijt} is an idiosyncratic i.i.d. distributed type I extreme value taste shock. I further normalize a consumer's indirect utility from choosing the outside option to $u_{i0t} = \varepsilon_{i0t}$.

While I do not specify the exact functional form of the mean utility of pickup truck j in market t, I implicitly assume that it is a function of observable pickup truck characteristics $\boldsymbol{x}_{jt}$ and unobservable pickup truck characteristics ξ_{jt}.

The innovation of my approach lies in the specification of the individual-specific deviations from product mean utilities μ_{ijt}. I model the consumer heterogeneity as follows:

$$\mu_{ijt} = -\alpha_{it}\rho_{ijt} + \sum_k \theta^k \boldsymbol{x}^k_{jt} \boldsymbol{z}^k_{it}. \tag{3}$$

Consumer heterogeneity in preferences for a pickup truck model is captured by the individual specific price sensitivity α_{it}, the individual specific price for a pickup truck ρ_{ijt}, as well as interactions between observed pickup truck characteristics $\boldsymbol{x}_{jt}$ with observed consumer demographics $\boldsymbol{z}_{it}$. Allowing for interactions between consumer demographics, observed and unobserved, and product characteristics allows for more flexible substitution patterns. I further assume that the support of α is the positive real line excluding zero and that the distribution of α has a finite first moment.

Next, let y_{ijt} denote the choice of consumer i. It equals 1 if the consumer chooses pickup truck model j in market t, and 0 otherwise. Integrating over the i.i.d. type I extreme value shocks yields the familiar logit choice probabilities for each consumer i

$$\begin{aligned} s_{ijt} &= \Pr(y_{ijt} = 1 | \boldsymbol{\rho}_{it}, \boldsymbol{x}_t, \boldsymbol{z}_{it}; \alpha_{it}, \boldsymbol{\delta}, \boldsymbol{\theta}) = \\ &= \frac{\exp(\delta_{jt} - \alpha_{it}\rho_{ijt} + \sum_k \theta^k x^k_{jt} z^k_{it})}{\sum_{l \in J_t} \exp(\delta_{lt} - \alpha_{it}\rho_{ilt} + \sum_k \theta^k x^k_{lt} z^k_{it})}, \end{aligned} \tag{4}$$

where $\boldsymbol{\rho}_{it}$ is the vector of all individual specific prices.

4.2 Supply

In section 2, I found that firms price discriminate on observable and unobservable consumer characteristics. Sales personnel at pickup truck dealerships are trained to infer the willingness to pay of consumers. The difference between most consumer goods markets and the US automobile market that makes this strategy viable is the interaction length between consumers and sales personnel. Before a salesperson offers a price to consumers, she can observe various consumer demographics and rely on information learned during test drives and conversations with the consumer. Due to the nature of these lengthy interactions between consumers and dealership personnel, sales personnel have a good understanding of the consumer's willingness to pay when making an offer.[22] However, sales personnel are unlikely to learn the willingness to pay perfectly. This is reminiscent of personalized pricing in the marketing literature, in which firms set prices based on large sets of consumer characteristics. Following this literature, I assume that firms can learn consumers' preferences up to a prediction or classification error (e.g. Dubé and Misra, 2023). While my approach to modeling is

[22] This appears to be well-known in literature. For example, Zettelmeyer, Scott Morton and Silva-Risso (2006) suggest that the negotiation process serves as a way for firms to learn consumers' willingness to pay.

closely related to the personalized pricing literature, it is novel because this literature did not simultaneously allow for heterogeneity in price sensitivity and competition. Following the personalized pricing literature, I model dealers' prediction errors as i.i.d. type I extreme value errors. I further assume that firms are aware that they cannot learn the willingness to pay of consumers perfectly but know the distribution of the prediction errors. Firms engage in a static, full-information Nash-Bertrand pricing game, simultaneously setting individual-specific prices for each consumer in the market.

I denote the set of all products within a market with $\mathcal{J}_t$ with cardinality J_t. Let $\mathcal{J}_t^d$ be the set of all products sold by firm d in market t. Firm d's expected profit from a single consumer i is given by

$$E[\pi_{i,t}^d] = \sum_{j \in \mathcal{J}_t^d} (\rho_{ijt} - c_{jt}) \cdot \mathrm{P}(u_{ijt} = \max\{u_{i0t}, ..., u_{iJ_t t}\}), \tag{5}$$

where c_{jt} denotes the firm's constant marginal cost for producing pickup truck model j in year t. Integrating over the distribution of the i.i.d. Gumbel errors yields

$$E[\pi_{i,t}^d] = \sum_{j \in \mathcal{J}_t^d} (\rho_{ijt} - c_{jt}) \cdot s_{ijt}(\boldsymbol{\rho}_{it}, \cdot). \tag{6}$$

Maximizing (6) with respect to the prices of the pickup truck models a firm offers yields the following first-order conditions

$$s_{ijt} + \sum_{j \in \mathcal{J}_t^d} (\rho_{ijt} - c_{jt}) \frac{\partial s_{ijt}}{\partial \rho_{ijt}} = 0. \tag{7}$$

Then, stacking all first-order conditions of all firms, one gets the usual, albeit consumer-specific, result for markups

$$\boldsymbol{\rho}_{it} - \mathbf{c}_t = (-\Omega_t \times \frac{\partial \boldsymbol{s}_{it}}{\partial \boldsymbol{\rho}_{it}})^{-1} \boldsymbol{s}_{it}, \tag{8}$$

where Ω_t represents the ownership matrix, $\frac{\partial \boldsymbol{s}_{it}}{\partial \boldsymbol{\rho}_{it}}$ is the matrix of all first partial derivatives of choice probabilities with respect to individual prices, and $\times$ represents the element-by-element multiplication.

Note that each consumer represents an independent market in this model. Each consumer has multinomial logit demand, $\boldsymbol{s}_{it}$, since the random coefficient α_{it} is consumer-specific. By theorem 7 of Konovalov and Sándor (2010), there exists a unique Nash-equilibrium satisfying the first-order conditions (8) in $[\boldsymbol{c}_t, +\infty)$.

5 Estimation

Estimation is complicated for two reasons. First, I need to jointly estimate demand and supply because prices are individual consumer-specific and only the transaction price can ever be observed. In particular, for consumers who bought a pickup truck, the prices for the $J_t - 1$ other pickup trucks available in the market are unobservable. Second, price offers for consumers who did not buy a pickup truck can never be observed. This additionally creates a selection problem.

5.1 Estimation Strategy

I jointly estimate supply and demand using maximum likelihood. The model's loglikelihood combines information on individual purchases ("microdata"), information on aggregate shares ("macrodata"), as well as supply-side optimality conditions. Before stating the loglikelihood of the model, I introduce notation.

To clearly distinguish between observed and unobserved prices, I define the observed, scalar-valued transaction price p_{it} as

$$p_{it} = \boldsymbol{\rho}_{it}' \cdot \boldsymbol{y}_{it}, \tag{9}$$

where $\boldsymbol{\rho}_{it}$ is the $J_t \times 1$ vector of prices the individual faces in the market, and $\boldsymbol{y}_{it} = [y_{i1t}, \ldots, y_{iJ_t t}]$ is a $J_t \times 1$ vector indicating the chosen pickup truck model of individual i: its elements equal 1 at the index of the chosen pickup truck and 0 for all other elements.

For each of the S_t consumers in market t who purchased a pickup truck model, I observe the triple $(p_{it}, \boldsymbol{y}_{it}, \boldsymbol{z}_{it})$: the transaction price, the vector indicating the consumer's choice, as well as the consumer's demographics. For each pickup truck model, I observe the tuple $(\boldsymbol{x}_{jt}, s_{jt})$: the vector of the pickup truck models' characteristics and the market share of the pickup truck model. Lastly, I denote the vector of model parameters as $\boldsymbol{\Theta} = (\boldsymbol{\theta}, \boldsymbol{\delta}, \boldsymbol{c})$. Then, I can write the model's loglikelihood as

$$\sum_{t=1}^{T} \sum_{i=1}^{S_t} \sum_{j=0}^{J_t} \mathbb{1}\{y_{ijt} = 1\} \log \mathscr{s}_{ijt}^*(p_{it}, \boldsymbol{z}_{it}, \boldsymbol{x}_t; \boldsymbol{\Theta}) + \sum_{t=1}^{T} \sum_{j=0}^{J_t} \left(N_t s_{jt} - \sum_{i=1}^{S_t} \mathbb{1}\{y_{ijt} = 1\} \right) \log \mathscr{s}_{jt}^*(\boldsymbol{x}_t; \boldsymbol{\Theta}) \tag{10}$$

where $\mathscr{s}_{ijt}^*$ are the individual specific equilibrium choice probabilities of consumers in the microdata, and $\mathscr{s}_{jt}^*$ the model's predicted equilibrium market shares. The main challenge for estimation is to recover both types of equilibrium choice probabilities to be able to form the loglikelihood. For expositional clarity, I present the strategy to form the loglikelihood for a single market, but no such restriction is necessary, and forming the loglikelihood proceeds analogously for all markets. Forming the loglikelihood given a guess of $\boldsymbol{\Theta}$ proceeds roughly in 4 steps, which I describe in the following.

5.1.1 Step 1: estimating price sensitivities from microdata

To recover the unobserved prices and unobserved consumer heterogeneity α_{it} for every consumer in the microdata, I rely on the supply side optimality conditions. I impose that the observed transaction price p_{it} satisfies the equilibrium conditions. As shown in the previous section, there exists a unique price vector satisfying the optimality conditions

$$\boldsymbol{\rho}_{it} - \mathbf{c}_t = (-\Omega_t \times \frac{\partial \boldsymbol{s}_{it}}{\partial \boldsymbol{\rho}_{it}})^{-1} \boldsymbol{s}_{it}. \tag{11}$$

These J_t optimality conditions define the price vector $\boldsymbol{\rho}_{it}$ and unobserved consumer heterogeneity α_{it} as implicit functions of the model's parameters $\boldsymbol{\Theta}$ and the data. Note that since we observe one element of the price vector, this is a system of J_t equations with J_t unknowns for each consumer. Therefore, we can write α_{it} and $\boldsymbol{\rho}_{it}$ as implicit functions of observables and the models parameters

$$\begin{aligned} \alpha_{it} &= \alpha(p_{it}, \boldsymbol{z}_{it}, \boldsymbol{x}_t; \boldsymbol{\Theta}), \\ \boldsymbol{\rho}_{it} &= \boldsymbol{\rho}^p(p_{it}, \boldsymbol{z}_{it}, \boldsymbol{x}_t; \boldsymbol{\Theta}). \end{aligned} \tag{12}$$

Next, because these functions are defined implicitly only, we have to numerically solve the first-order conditions of firms in (8) for the corresponding vector of prices $\boldsymbol{\rho}_{it}$ and price sensitivity α_{it}. This requires solving J_t equations for J_t unknowns.

However, instead of directly using (8) to solve for J_t prices and price sensitivity, we can rely on a well-known property of multinomial logit demand models: firms set the same profit margin, defined as price minus marginal cost, for all products. Therefore, we only need to solve for D_t instead of J_t unknowns, where D_t is the number of firms active in market t. This considerably speeds up estimation. After solving for α_{it} for all consumers, we have the vector of consumers' price sensitivities conditional on purchase.

Substituting the functions in (12) for prices and price sensitivity also allows us to rewrite the choice probabilities s_{ijt} for consumers in the microdata entirely in terms of data and parameters imposing the optimality conditions of the firms

$$s^*_{ijt}(p_{it}, \boldsymbol{z}_{it}, \boldsymbol{x}_t; \Theta) = \frac{\exp(\delta_{jt} - \alpha(p_{it}, \boldsymbol{z}_{it}, \boldsymbol{x}_t; \boldsymbol{\Theta}) \cdot \rho^p_j(p_{it}, \boldsymbol{z}_{it}, \boldsymbol{x}_t; \boldsymbol{\Theta}) + \sum_k \theta^k x^k_{jt} z^k_{it})}{1 + \sum_{l \in J_t} \exp(\delta_{lt} - \alpha(p_{it}, \boldsymbol{z}_{it}, \boldsymbol{x}_t; \boldsymbol{\Theta}) \cdot \rho^p_k(p_{it}, \boldsymbol{z}_{it}, \boldsymbol{x}_t; \boldsymbol{\Theta}) + \sum_k \theta^k x^k_{lt} z^k_{it})}. \tag{13}$$

Thus, after this step, we have the distribution of price sensitivities conditional on purchase and demographics, as well as the vector of choice probabilities for each consumer in the microdata. In the next step, we need to solve the selection into purchase problem.

5.1.2 Step 2: estimating the distribution of price sensitivities conditional on demographics

Let $\mathcal{Y}$ the random variable indicating whether a consumer bought any pickup truck model or opted for the outside good. Its realization y equals 1 if the consumer chooses a pickup truck model, and 0 if she chooses the outside option.

From step 1, we have the vector of consumers' price sensitivities conditional on purchase and demographics. I estimate the distribution of price sensitivities conditional on purchase and demographics, $dF_{\alpha|\mathcal{Y}=y,Z=\boldsymbol{z}}(a|y \neq 0, Z = \boldsymbol{z})$, using a kernel density estimator.[23] To form predicted market shares, however, we need the distribution of price sensitivities conditional on demographics only.

To solve the selection into purchase problem, we can again exploit the optimality conditions in (8): for every point $(a, \boldsymbol{z})$ in the support of the joint distribution of price sensitivities and demographics, $F_{\alpha,Z}(a, \boldsymbol{z})$, the supply side optimality conditions uniquely pin down the corresponding price vector

[23] Note that here I assume that demographics are following a discrete distribution. While none of this is necessary and can be relaxed, my demographic data comes in bins.

$$\boldsymbol{\rho} = \boldsymbol{\rho}^{\alpha}(a, \boldsymbol{z}, \boldsymbol{x}; \boldsymbol{\Theta}). \tag{14}$$

As we will see, this relation is key to recovering the distribution of price sensitivities conditional on demographics only. From the functional form of the logit choice probabilities, it follows that

$$P(y = 0|\alpha = a, Z = \boldsymbol{z}) = \frac{1}{1 + \sum_{l=1}^{J_t} \exp(\delta_{lt} - a \cdot \rho_j^{\alpha}(a, \boldsymbol{z}, \boldsymbol{x}_t; \boldsymbol{\Theta}) + \sum_k \theta^k x_{lt}^k z^k)}, \tag{15}$$

where ρ_l^{α} is the l-th element of the implicit pricing function from the supply side first-order conditions defined in (14).

Next, $Pr(y \neq 0|Z = \boldsymbol{z})$ is the sum of the conditional inside shares for a given demographic group. We can express these as a function of the data by an application of Bayes' rule

$$P(y \neq 0|Z = \boldsymbol{z}) = \frac{P(\boldsymbol{z}|y \neq 0)P(y \neq 0)}{P(Z = \boldsymbol{z})}, \tag{16}$$

where $P(\boldsymbol{z}|y \neq 0)$ can be estimated directly from the microdata, $P(y \neq 0)$ are the unconditional inside shares from the transaction level data, and $P(Z = \boldsymbol{z})$ is the joint distribution of demographics in Texas from the Current Population Survey.

By applying Bayes' rule, we can now obtain the distribution of price sensitivities conditional on demographics only, solving the selection into purchase problem

$$dF_{\alpha|Z=\boldsymbol{z}}(a|Z = \boldsymbol{z}) = \frac{dF_{\alpha|y=y,Z=\boldsymbol{z}}(a|y \neq 0, Z = \boldsymbol{z})P(y \neq 0|Z = \boldsymbol{z})}{1 - P(y = 0|\alpha = a, Z = \boldsymbol{z})}. \tag{17}$$

Now, we can use this distribution to form the model's predicted shares in the next step.

5.1.3 Step 3: forming model's predicted shares

To obtain predicted market shares, we need to integrate over the individual-specific choice probabilities of consumers in the market. Using (14), the choice probabilities at the point $(a, \boldsymbol{z})$ are given by

$$\mathscr{s}_{jt}^{a,z}(a, \boldsymbol{z}, \boldsymbol{x}_t; \boldsymbol{\Theta}) = \frac{\exp(\delta_{jt} - a \cdot \rho_j^{\alpha}(\boldsymbol{z}, \boldsymbol{x}_t, a; \boldsymbol{\Theta}) + \sum_k \theta^k x_{jt}^k z^k)}{1 + \sum_{l \in J_t} \exp(\delta_{lt} - a \cdot \rho_l^{\alpha}(\boldsymbol{z}, \boldsymbol{x}_t, a; \boldsymbol{\Theta}) + \sum_k \theta x_{lt}^k z^k)}. \tag{18}$$

Then, integrating over the joint distribution of price sensitivities and consumer demographics yields the predicted market shares $\mathscr{s}_{jt}^*(\cdot)$:

$$\begin{aligned} \mathscr{s}_{jt}^*(\boldsymbol{x}_t; \Theta) &= \int \mathscr{s}_{jt}^{a,z}(a, \boldsymbol{z}, \boldsymbol{x}_t; \boldsymbol{\Theta}) dF_{\alpha,Z}(a, \boldsymbol{z}) \\ &= \int \mathscr{s}_{jt}^{a,z}(a, \boldsymbol{z}, \boldsymbol{x}_t; \boldsymbol{\Theta}) dF_{\alpha|Z=\boldsymbol{z}}(a|Z = \boldsymbol{z}) dF_Z(\boldsymbol{z}), \end{aligned} \tag{19}$$

which are the model analogs to observed market shares in the data. Here, we used the distribution of price sensitivities conditional on demographics from equation (17).

5.1.4 Step 4: forming and maximizing the Loglikelihood

From step 1, we have the choice probabilities $\mathscr{s}_{ijt}^*$ of the consumers in the microdata; from step 3 we have the model's predicted market shares $\mathscr{s}_{jt}^*$ for all products. Thus, we can form the model's loglikelihood for market t by

$$\begin{aligned} &\log \mathcal{L}_t(\boldsymbol{p}_t, \boldsymbol{z}_t, \boldsymbol{x}_t, \boldsymbol{s}_t; \boldsymbol{\Theta}) = \\ &\sum_{i=1}^{S_t} \sum_{j=0}^{J_t} \mathbb{1}\{y_{ijt} = 1\} \log \mathscr{s}_{ijt}^*(p_{it}, \boldsymbol{z}_{it}, \boldsymbol{x}_t; \boldsymbol{\Theta}) + \sum_{j=0}^{J_t} \left(N_t s_{jt} - \sum_{i=1}^{S_t} \mathbb{1}\{y_{ijt} = 1\} \right) \log \mathscr{s}_{jt}^*(\boldsymbol{x}_t; \boldsymbol{\Theta}). \end{aligned} \tag{20}$$

Note that I did not impose that the distribution of price sensitivities is the same across markets.[24] The model's loglikelihood is then the sum of all the markets' loglikelihoods,

$$\log \mathcal{L}(\cdot; \boldsymbol{\Theta}) = \sum_{t=1}^{T} \log \mathcal{L}_t(\cdot; \boldsymbol{\Theta}). \tag{21}$$

In practice, not every parameter guess will be consistent with all first-order conditions of all consumers in my data. Therefore, if a parameter guess creates such a problem, I add a smooth, quadratic penalty term to the loglikelihood.

5.1.5 Measuring the correlation of price sensitivity and consumer demographics

After estimation, we have estimates for $\boldsymbol{\Theta}$ and $F_{\alpha,Z}(a,z)$. Since I do not assume that α_{it} follows a specific parametric distribution, I can use the estimate $\hat{F}_{\alpha,Z}(a,z)$ of the distribution of price sensitivities to explore the correlation between observed consumer demographics and price sensitivities. Given the estimates for $\boldsymbol{\Theta}$, I draw 10,000 times from the estimated joint distribution. I then regress these draws on the demographics to understand how price sensitivities and consumer demographics correlate.

5.2 Results

I estimate the model by maximum likelihood using the *microdata for demand estimation* covering 2016 – 2019 following the procedure outlined in section 5.1. In the first step, I obtain estimates for the parameters in $\boldsymbol{\Theta}$. In the second step, I show how price sensitivity and consumer demographics correlate.

Table 5: Parameter Estimates of Structural Model

	pickup truck characteristics			
	constant	Size	Miles/Gal.	US-Brand
income	0.498 (0.025)	0.145 (0.017)	0.090 (0.018)	-0.214 (0.030)
female	-1.962 (0.020)	—	—	—
Hispanic	-0.719 (0.020)	—	—	—

Note: Standard errors in parentheses.
All continuous pickup truck characteristics and income standardized.

5.2.1 Parameter Estimates

I specify consumer i's indirect utility from choosing pickup truck model j in market t as

$$\begin{aligned} u_{ijt} = \delta_{jt} - \alpha_{it}\rho_{ijt} + \theta_1 \text{income}_{it} + \theta_2 \text{size}_{jt} \times \text{income}_{it} + \theta_3 \text{mpg}_{jt} \times \text{income}_{it} \\ + \theta_4 \text{US-Brand}_{jt} \times \text{income}_{it} + \theta_5 \text{female}_{it} + \theta_6 \text{Hispanic}_{it} + \varepsilon_{ijt} \end{aligned} \tag{22}$$

where income is standardized to have zero mean and unit variance, size is the length times the height of a pickup truck, mpg is the miles per gallon rating, and US-Brand is an indicator variable for US makes. I standardize all continuous variables except income. This specification allows consumers of different sociodemographic groups to differ in their taste for buying a pickup truck model. I do not parametrize the

[24]While this allows me to consistently estimate the distribution of α if the distribution of α is the same across markets, a more efficient estimator that does not separate estimation by market in the inner loop would exist. Hence, my estimator is only efficient if the distribution of price sensitivities differs across years.

distribution of α_{it} but nonparametrically estimate it from the microdata exploiting the supply side equilibrium conditions in (8). Price sensitivities may be arbitrarily correlated with consumer demographics. The flexible specification allows the model to capture asymmetric rankings in the best responses across firms. If price discrimination indeed lead to more competitive pressure for firms, as considered by Corts (1998), this specification can capture it.

Table 5 presents the results for the estimates of the structural parameters θ. These parameters govern the taste for product characteristics by consumer demographics. The results intuitively match the correlations in the microdata uncovered in section 3. Higher-income individuals are likelier to purchase a pickup truck in any given year than low-income individuals. Female buyers are less likely to buy a pickup truck than males. Similarly, Hispanic consumers are less likely to purchase a pickup truck in any given year compared to non-Hispanic consumers.

The estimates for α_{it} show consumers' strong distaste for price. I estimate the mean price sensitivity to be 2.95, with a standard deviation of 2.11, highlighting the substantial unobserved heterogeneity in price sensitivity. Figure 3 plots the estimated unconditional distribution of price sensitivities.

5.2.2 Correlations of Price Sensitivities with Protected Classes

Table 6 presents the results for the correlations of α with protected classes (column α), as well as the standard deviations of α conditional on protected classes (column $\sigma(\alpha|Z)$). The correlations between price sensitivities and protected classes are intuitively matching the explorative analysis. On average, female and Hispanic buyers pay more for the same pickup truck than non-Hispanics and males. This maps directly into the estimates for price sensitivities: I estimate females and Hispanics to be less price-sensitive.

While table 6 summarizes the differences in means *across* demographic groups,

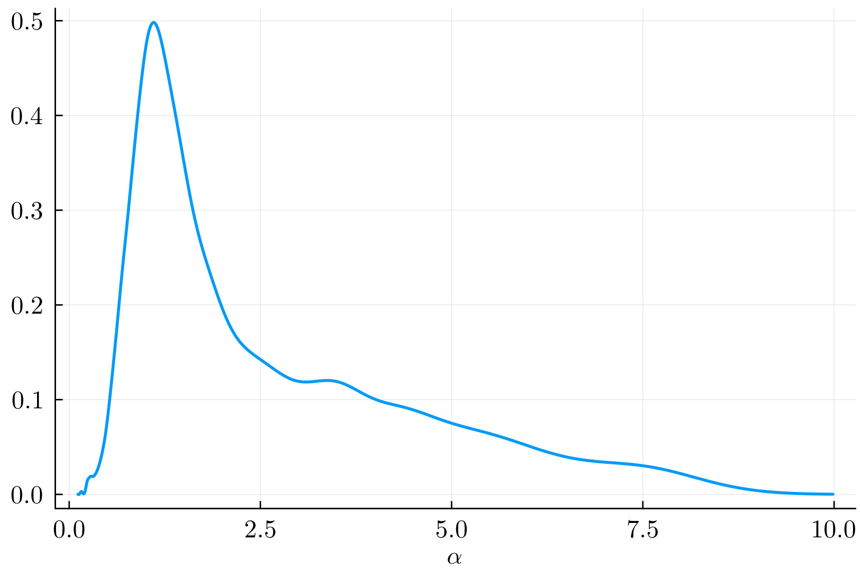

Figure 3: Unconditional distribution of price sensitivities

figure 4 plots the distributions of price sensitivities for females and males, Hispanics, and non-Hispanics. These plots highlight the substantial heterogeneity in price sensitivity even conditional on protected classes. The distributions of price sensitivities for women and men, and Hispanics and non-Hispanics considerably overlap. This implies that while mean differences across protected classes exist, heterogeneity within protected classes is much more significant.

5.2.3 Determinants of Marginal Costs

The average estimated marginal cost for a pickup truck model is $18,911. This aligns well with industry reports. The estimates exhibit intuitive rankings regarding manufacturers: Toyota produces the most expensive pickup trucks, while Nissan produces the cheapest. These estimates match the price levels observed in the transaction level data well.

As an additional check, I project estimated marginal costs of producing a pickup truck on pickup truck characteristics. Results are reported in table 7. Pickup trucks of American manufacturers are estimated to be cheaper to produce than foreign pickup

	α	$\sigma(\alpha\|Z)$
female	-1.434***	1.316
	(0.027)	
Hispanic	-0.700***	1.626
	(0.028)	
N	10,000	

Note: Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 6: Correlations of price sensitivity with protected classes

	marginal cost
Size	0.495***
	(0.167)
MPG	0.181***
	(0.064)
Horsepower	0.001
	(0.001)
US-Brand	-1.185***
	(0.269)
R^2	0.501
Year FE	✓

Note: Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Size in thousands of square inches.

Table 7: Regression of marginal costs on vehicle characteristics

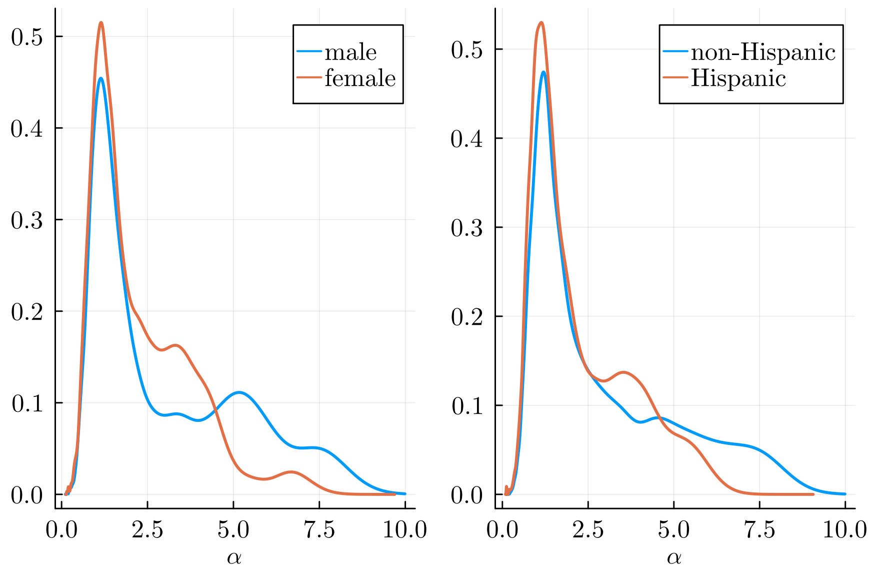

Figure 4: Conditional distributions of price sensitivities

trucks. Additionally, larger pickup trucks are more costly to produce. Trucks with higher per gallon ratings are more expensive to produce but horsepower is estimated to have no significant effect on marginal cost.

5.2.4 Price Elasticities and Markups

A complication for calculating price elasticities and markups given the transaction-by-transaction pricing decisions of firms is that separate own-price elasticities exist for each consumer, which is non-standard. To obtain a comparable statistic to the standard notion of an own-price elasticity, I define the market-level own-price elasticity for pickup truck model j in market t as follows

$$\epsilon_{jt} = -\int a \cdot \rho_j^{\alpha}(\cdot)(1 - s_{jt}^{a,z}(\cdot))dF_{\alpha,Z}(a, z). \tag{23}$$

The interpretation of this statistic changes relative to the standard notion of price elasticity in the following way: It measures the change in demand for product j when the average price level in the market for product j increases by 1%. As defined above, the estimated parameters imply that the average market-level own-price elasticity across all products and markets is -7.31. Therefore, demand is relatively more elastic than in previous studies on the automobile market. However, this model's more price-elastic demand curve is not surprising since preceding studies have relied on MSRPs or average transaction prices as proxies for all quoted prices in a market. As it turns out, the consumers who ultimately purchase a pickup truck are the least price-sensitive consumers in the market, paying the highest markups. Using the average transaction price thus significantly overstates the quoted prices in the market. Consequently, these studies' estimated mean price sensitivity parameters can be severely downward biased. Similarly, as shown in figure 2, most purchasers pay significantly less than MSRP for their pickup trucks. Therefore, using MSRP exacerbates the problems that the average transaction price induces if the MSRP is larger than the average transaction price.

Similar to the definition of the market-level own-price elasticity for a pickup truck model, I define the market-level markup for pickup truck model j in market t as

$$\text{markup}_{jt} = \int \frac{\rho_j^{\alpha}(\cdot) - c_{jt}}{\rho_j^{\alpha}(\cdot)} dF_{\alpha,Z}(a,z). \tag{24}$$

This definition of the average market-level markup for a pickup truck model is analogous to the familiar product-level markups measured by the Lerner index. I estimate that the average markup is around 0.21, while the average share-weighted markup is 0.35. The difference in both again highlights the importance of distinguishing between the models' estimated price quotes for the whole market and the realized prices in the transaction data. Since most consumers in the market are very price-sensitive, average markups are substantially lower than average share-weighted markups.

While average markups are informative about the competitiveness of a market,

	markup (share-weighted)	markup (share-weighted)
female		0.007***
		(0.002)
Hispanic		0.011***
		(0.002)
R^2	0.758	0.785
Year-Model FE	✓	✓

Note: Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 8: Regression of share-weighted markups on consumer demographics

they are uninformative about the variation of markups across protected classes. To study the correlation between markups set by firms, I project the estimated margins on consumers' demographics. The results of this regression are reported in table 8. The correlations of markups again broadly match the results in the descriptive evidence. As price sensitivity directly maps into markups, Hispanic and female buyers are charged significantly higher margins for the same pickup truck model. On average, females are charged 0.7 percentage points higher markups than males, and Hispanics pay around 1 percentage point higher markups than non-Hispanics. This amounts to around \$200 more for females and \$400 more for Hispanics for the same truck. However, these demographics do not explain a large fraction of the variation in markups. Including controls for protected classes in addition to year-model fixed effects only increases the R^2 by around 2%. This highlights again that unobserved heterogeneity plays a much larger role in explaining the variation in markups than whether a consumer is a member of a protected class.

To compare the estimates of the model with industry sources on price-cost margins, I define the average price-cost margin for pickup truck model j in market t as

$$\text{margin}_{jt} = \int (\rho_j^{\alpha}(\cdot) - c_{jt}) dF_{\alpha,Z}(a, z). \tag{25}$$

Again, price-cost margins are consumer-specific, but we can define a market-level analog to the familiar statistic. Using the definition of the market-level price-cost margin for pickup truck model j, I estimate that the average market-level margin across all products is around \$5,592. The average share-weighted margin is \$10,683. These margins might seem high but are in line with industry reports. For example, industry reports suggest that Ford makes around \$13,000 in profit per pickup truck.[25]

To sum up, the model produces reasonable estimates for demand, costs, and markups. While I have to resort to new definitions for elasticities and margins, these statistics align well with what has been reported in industry sources. In the next section, I will use these estimates to study the welfare effects of price discrimination.

6 The Value of Consumer Information in Discriminatory Pricing

A crucial component in whether price discrimination using consumer-specific prices benefits or harms consumers is the amount of consumer information firms use to price discriminate. To isolate the role of information in the welfare effects of price discrimination, I conduct multiple counterfactual simulations. Each of the counterfactuals uses a specific form of pricing in which firms can use *less* information than the actual amount of information firms use to set consumer-specific or personalized prices under the baseline. In particular, I consider the following counterfactual simulations:

[25] Automotive News 2015. "The F-150, Ford's heavy-duty profit hauler", *Automotive News*, April 28. https://www.autonews.com/article/20150428/BLOG06/150429797/the-f-150-ford-s-heavy-duty-profit-hauler, accessed 2023-07-01.

1. *Uniform pricing*: First, I conduct a counterfactual in which firms cannot use any information beyond the expected values of price sensitivities and demographics in the population. Firms then set a single price for all consumers for each pickup truck model they produce as in the canonical Berry, Levinsohn and Pakes (1995) model. It is well known that the stacked first-order conditions then take the form of

$$\boldsymbol{\rho}_t^u - \mathbf{c}_t = (-\boldsymbol{\Omega}_t \times \frac{\partial \boldsymbol{s}_t}{\partial \boldsymbol{\rho}_t^u})^{-1} \boldsymbol{s}_t. \tag{26}$$

where $\boldsymbol{\rho}_t^u$ is the vector of uniform prices. The first-order conditions define the vector of uniform prices $\boldsymbol{\rho}_t^u$ as an implicit function of the model's parameters and the data.[26]

2. *Price discrimination on gender, race, and income only*: Second, I conduct a counterfactual in which firms have information on gender, race, and income only. Firms do not observe the consumer-specific price sensitivity but only the distribution of price sensitivities conditional on consumer demographics. Firms then set a single price for each demographic group. For example, firms set a single price for all Hispanic females with incomes between \$50,000 and \$75,000 per year for each pickup truck model they offer. This corresponds to the dominant approach studying price discrimination in the automobile market (e.g. Langer, 2016; D'Haultfoeuille, Durrmeyer and Février, 2019). The corresponding stacked first-order conditions are equivalent to (26) but are now *group-specific*

$$\boldsymbol{\rho}_t^d - \mathbf{c}_t = (-\boldsymbol{\Omega}_t \times \frac{\partial \boldsymbol{s}_t^d}{\partial \boldsymbol{\rho}_t^d})^{-1} \boldsymbol{s}_t^d. \tag{27}$$

[26] It is well known that a Nash equilibrium does not need to exist in random coefficients discrete choice models. Therefore, I assume the existence of a pure strategy Nash equilibrium satisfying the first-order conditions as is standard in literature (e.g. Berry, Levinsohn and Pakes, 1995).

where $\boldsymbol{\rho}_t^d$ is the price vector and $\boldsymbol{\mathscr{s}}_t^d$ the model predicted shares for demographic group d.

I use the compensating variation to quantify consumer welfare changes under the different informational settings. The compensating variation in market t comparing consumer surplus under uniform prices to discriminatory pricing is

$$\begin{aligned}\Delta CS_t = &\int \frac{1}{a} \log(1 + \sum_{j \in \mathcal{J}_t} \exp(\delta_{jt} - a\rho_{jt}^u + \sum_k \theta^k x_{jt}^k z^k)) dF_{\alpha,Z}(a,z) \\ &- \int \frac{1}{a} \log(1 + \sum_{j \in \mathcal{J}_t} \exp(\delta_{jt} - a\rho_j^{a,z}(\cdot) + \sum_k \theta^k x_{jt}^k z^k)) dF_{\alpha,Z}(a,z),\end{aligned} \tag{28}$$

As usual, I measure welfare changes for firms in profits.

6.1 Counterfactual results

The counterfactuals let me evaluate the welfare effects of using consumer information beyond the distribution of demographics and price sensitivities in population to firms. First, the uniform pricing counterfactual isolates the value of consumer information relative to only knowing the distribution of consumer characteristics in the population. Second, the counterfactual in which firms only use demographic consumer information observable to the econometrician isolates the importance of observable vs unobservable consumer heterogeneity.

The value of consumer information for price discrimination: The *baseline* columns of table 9 show that firms benefit significantly from using information beyond simple population averages. Acquiring detailed, fine-grained information through interactions with individual consumers and using this information to charge consumer-specific, personalized prices substantially raises profits by around 20% over the profits under uniform prices. In particular, not only industry profits but profits across all manufacturers are unequivocally higher than ignoring the information about consumers

	price discrimination *baseline*		price discrimination *gender, race, income only*	
	\$ millions	%	\$ millions	%
Toyota	+75.67	+22.57%	-0.93	-0.37%
Honda	+4.06	+23.37%	-0.01	-0.05%
Nissan	+10.86	+15.73%	-0.17	-0.26%
Ford	+72.75	+14.76%	-1.10	-0.26%
GM	+90.65	+14.59%	-2.11	-0.40%
RAM	+36.04	+16.69%	-0.27	-0.14%
Industry profits	+290.05	+19.75%	-7.30	-0.31%

Table 9: Gains/losses in profit relative to uniform pricing due to price discrimination on all consumer characteristics (baseline) and gender, race, and income only.

and charging uniform prices. While the profitability of price discrimination based on detailed consumer features is unambiguous across firms, there is heterogeneity in how much firms profit from using this information. Figure 6 in the appendix documents this heterogeneity.

Next, I show that price discrimination need *not* raise profits in this setting if firms information about consumers is coarse. The *gender, race, income only* columns of table 9 show that different information structures can completely reverse the profitability of using consumer information for pricing. In particular, industry profits from price discrimination only on demographics observable to the econometrician, gender, race, and income, would be around 0.3% lower than under uniform pricing. Table 9 furthermore shows that when firms are restricted to using information on demographic groups only, *all* firms lose from price discrimination. Notably, these results are in line with findings of the preceding literature studying price discrimination in the US automobile market based on demographic groups.[27]

These results highlight that dealers' primary source of profits is consumer information beyond coarse demographic groups. The model estimates imply significant heterogeneity in price sensitivity across and within demographic groups. Large heterogeneity makes learning about price sensitivity dispersion and personalize price offerings highly profitable for dealers. Sales personnel invest considerable effort to learn the willingness to pay of consumers via long conversations during test drives, the negotiations phase, and even financial information when running consumers' credit. While dealers use information on protected classes and income for pricing, these demographics only offer limited information about consumers' willingness to pay because the dispersion of price sensitivity is significant, even within demographic groups.

This result also shows why price discrimination does not lead to lower prices in the pickup truck market. Corts (1998) showed that price discrimination can improve consumer welfare. These results rest on a property called best response asymmetry. The asymmetries that can induce discriminatory prices below uniform prices are a feature of horizontal price discrimination. The model is flexible enough to capture these asymmetries. For example, my estimates imply that higher-income consumers prefer larger, foreign-made pickup trucks while low-income consumers prefer smaller, domestic ones. US manufacturers have an incentive to offer lower prices to high-income consumers, while foreign firms have the same incentives for low-income consumers. Corts (1998) showed that this can lead to more pricing pressure and lower prices for all consumers. However, the estimates also imply that the drivers of price dispersion in the pickup truck market are differences in price sensitivities across consumers. The institutional details of the market allow dealers to learn these price sensitivities and base their prices on them. However, all dealers use price sensitivities in the same way:

[27]Using a model with price discrimination based on demographic groups, Langer (2016) finds that price discrimination based on gender and marital status would reduce industry profits by around 3% per year.

less price-sensitive consumers are offered higher prices. Because Corts (1998) insight applies to horizontal but not vertical price discrimination, it did not lead to lower prices for all consumers in the pickup truck market. Subsequently, using very fine consumer information raises profits for firms.

Relative to uniform pricing, the increase in industry profits from using consumer information beyond coarse demographic groups stems from two sources. First, dealers leverage consumer information to offer price-sensitive consumers better deals than under uniform pricing. Thus, under price discrimination, price-sensitive consumers are likelier to buy a pickup truck than under uniform prices. This realizes profits dealers could not have earned under uniform pricing. Second, dealers can use the same consumer information to charge price-insensitive consumers higher prices than under uniform prices while making them less likely to buy pickup trucks. Figure 9 in the appendix illustrates how choice probabilities across different levels of price sensitivities change when moving from uniform pricing to discriminatory pricing to illustrate this tradeoff. These sources have opposing forces on the average prices offered in the market as well as aggregate demand. While price discrimination, compared to uniform pricing, using consumer-specific prices reduces prices on *average* by 15%, share-weighted prices increase by 2%. Share-weighted prices increase because price-insensitive consumers are more likely to buy pickup trucks. Finally, price discrimination also leads to higher demand. On average, across years, dealers' access to detailed consumer information increases pickup truck sales by around 10% compared to under uniform pricing.

Effect of Price Discrimination on Consumer Surplus: Price discrimination on very detailed levels of consumer information reduces aggregate consumer surplus. In particular, as figure 5 illustrates, price discrimination by dealers decreases *average* compensating variation by around 7% relative to a situation with posted uniform prices. This corresponds to losses of around $197 million in consumer surplus per year. Turning to the issue of protected classes, figure 5 also shows that, on average,

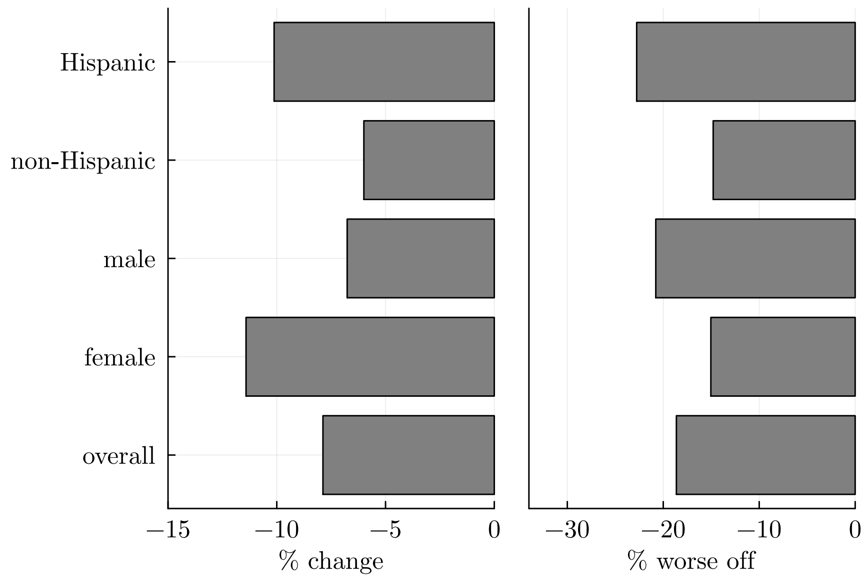

Figure 5: Losses in consumer surplus due to price discrimination by demographic group. Left panel shows average percentage change in consumer surplus, right panel percentage of people in demographic group worse off than under uniform prices.

Hispanics and women are disproportionally losing from consumer-specific pricing under price discrimination.

While the average consumer in the pickup truck market loses from dealers using granular consumer information to price discriminate, most consumers are better off. Some consumers lose a lot, while most gain little compared to uniform prices. The left panel in figure 5 shows that the losses in consumer surplus stem from a minority of around 19% of consumers. Moreover, even most female and Hispanic consumers benefit from price discrimination. This result arises because most consumers are very price-sensitive, even across and within demographic groups. The burden of price discrimination is born by a minority of price-insensitive consumers who pay significantly higher prices than under uniform pricing. The losses of this subgroup of consumers are, however, those that cause consumer surplus to decrease. A more detailed discussion of why consumer surplus decreases can be found in appendix A.2.

Lastly, it is important to notice that while consumer-specific prices harm consumers *on average*, reducing consumer surplus, the losses in consumer surplus are less than the gains in profit for firms. Since dealers can leverage very granular consumer information for their price quotes, prices are closer to the actual willingness to pay of consumers.[28] Therefore, total surplus increases compared to uniform pricing, and dealers leveraging detailed consumer information for pricing is more efficient than uniform pricing. In particular, across years, the average total surplus increases by around $92 million per year.

7 Conclusions

In this paper, I showed that dealers leverage large amounts of consumer information when making pricing decisions. Information about protected classes plays a small but significant role in the pricing decisions of dealers. Controlling for the heterogeneity of pickup trucks and leveraging panel data on consumers, I showed that protected classes pay significant premiums for the same trucks. Using a novel equilibrium model of supply and demand with individualized pricing, I show that while the use of detailed consumer information for pricing is controversial, it increases profits and total surplus, but *on average* harms consumers. However, whether or not consumer information increases profits crucially depends on how detailed these data are. In particular, I show that if firms can only price discriminate on demographic information, including protected classes, firms lose from price discrimination. These results on how the granularity of consumer information affects welfare results complement the recent theoretical literature studying the role of information segmentation in price discrimination for welfare with real-world evidence.

This paper also offers important insights for the contemporaneous personalized pricing literature. I show that the concerns about restricting firms from employing personalized prices raised in theoretical literature do not apply to one of the most

[28] Note that there is still uncertainty because firms cannot observe the Gumbel error term.

important consumer goods markets in the United States. Indeed, allowing firms to engage in price discrimination leads to a reduction in average consumer surplus in the Texas market for pickup trucks and a redistribution of surplus from price-insensitive consumers to price-sensitive consumers as well as firms. However, personalized prices in the pickup truck market increase total welfare. While the theoretical literature has echoed concerns about the possible consumer welfare losses from restricting price discrimination, my results call for more empirical studies validating these concerns in real-world markets.

Lastly, there are fruitful directions for future research. My approach focuses on the main features of the pickup truck market. Although I have excellent and complete data on the consumer and product level for the US automobile market, I cannot consider markets closely tied to the market for pickup trucks. An exciting direction for future research might be combining the individualized pricing of this model with the trade-in or financing negotiations process. These prices are usually highly correlated with the actual transaction price in the new pickup truck market and might offer additional insights about consumers' price sensitivities and firms' information about consumers' willingness to pay.

References

Abowd, John M., Francis Kramarz, and David N. Margolis. 1999. “High Wage Workers and High Wage Firms.” *Econometrica*, 67(2): 251–333.

Ambrose, Brent W, James N Conklin, and Luis A Lopez. 2020. “Does Borrower and Broker Race Affect the Cost of Mortgage Credit?” *The Review of Financial Studies*, 34(2): 790–826.

Ayres, Ian, and Peter Siegelman. 1995. “Race and Gender Discrimination in Bargaining for a New Car.” *American Economic Review*, 3(85): 304–321.

Bayer, Patrick, Marcus Casey, Fernando Ferreira, and Robert McMillan. 2017. “Racial and ethnic price differentials in the housing market.” *Journal of Urban Economics*, 102: 91–105.

Bernstein, Shai, Rebecca Diamond, Abhisit Jiranaphawiboon, Timothy McQuade, and Beatriz Pousada. 2022. “The Contribution of High-Skilled Immigrants to Innovation in the United States.” National Bureau of Economic Research Working Paper 30797.

Berry, Steven, James Levinsohn, and Ariel Pakes. 1995. “Automobile prices in market equilibrium.” *Econometrica: Journal of the Econometric Society*, 841–890.

Buchholz, Nicholas, Laura Doval, Jakub Kastl, Filip Matejka, and Tobias Salz. 2022. “The Value of Time: Evidence from Auctioned Cab Rides.” Princeton University. Economics Department. Working Papers 2022-22.

Busse, Meghan R., Ayelet Israeli, and Florian Zettelmeyer. 2017. “Repairing the Damage: The Effect of Price Knowledge and Gender on Auto Repair Price Quotes.” *Journal of Marketing Research*, 54: 75 – 95.

Chandra, Ambarish, Sumeet Gulati, and James M. Sallee. 2017. "Who Loses When Prices Are Negotiated? An Analysis of the New Car Market." *The Journal of Industrial Economics*, 65(2): 235–274.

Corts, Kenneth. 1998. "Third-Degree Price Discrimination in Oligopoly: All-Out Competition and Strategic Commitment." *RAND Journal of Economics*, 29(2): 306–323.

Coşar, A. Kerem, Paul L. E. Grieco, and Felix Tintelnot. 2015. "Borders, Geography, and Oligopoly: Evidence from the Wind Turbine Industry." *The Review of Economics and Statistics*, 97(3): 623–637.

Cox Automotive. 2019. "2019 Car Buyer Journey Study." https://www.coxautoinc.com/learning-center/2019-car-buyer-journey-study/, Last accessed on 2023-06-10.

D'Haultfoeuille, Xavier, Isis Durrmeyer, and Philippe Février. 2019. "Automobile Prices in Market Equilibrium with Unobserved Price Discrimination." *The Review of Economic Studies*, 86(5): 1973–1998.

Diamond, Rebecca, Tim McQuade, and Franklin Qian. 2019. "The Effects of Rent Control Expansion on Tenants, Landlords, and Inequality: Evidence from San Francisco." *American Economic Review*, 109(9): 3365–94.

Drury, Ivan, and Jessica Caldwell. 2020. "edmunds Truck Report 2020." Available at https://static.ed.edmunds-media.com/unversioned/img/industry-center/insights/2020-truck-report.pdf, Last accessed on 2023-08-04.

Dubé, Jean-Pierre, and Sanjog Misra. 2023. "Personalized Pricing and Consumer Welfare." *Journal of Political Economy*, 131(1): 131 – 189.

Elliott, Matthew, Andrea Galeotti, Andrew Koh, and Wenhao Li. 2022. "Market Segmentation Through Information." *Working Paper*.

Goldberg, Pinelopi Koujianou. 1996. "Dealer Price Discrimination in New Car Purchases: Evidence from the Consumer Expenditure Survey." *Journal of Political Economy*, 3(104): 622–654.

Goldsmith-Pinkham, Paul, and Kelly Shue. 2023. "The Gender Gap in Housing Returns." *The Journal of Finance*, 78(2): 1097–1145.

Grieco, Paul, Charles Murry, and Ali Yurukoglu. 2021. "The Evolution of Market Power in the US Automobile Industry." *Working Paper*.

Harless, David W., and George E. Hoffer. 2002. "Do Women Pay More for New Vehicles? Evidence from Transaction Price Data." *American Economic Review*, 1(92): 270–279.

Infutor Data Solutions. 2023. "Auto Data Appends." https://infutor.com/profile-appends/automotive-appends/, Last accessed on 2023-06-08.

Konovalov, Alexander, and Zsolt Sándor. 2010. "On price equilibrium with multi-product firms." *Economic Theory*, 44(2): 271–292.

Langer, Ashley. 2016. "(Dis)Incentives for Demographic Price Discrimination in the New Vehicle Market." *Working Paper*.

Murry, Charles, and Henry S. Schneider. 2016. "The economics of retail markets for new and used cars." In *Handbook on the Economics of Retail and Distribution.* , ed. Emek Basker, 343–367. Edward Elgar.

National Automobile Dealers Association. 2019*a*. "Driving Texas' Economy: Annual Contribution of Texas' New-Car Dealers." https://web.archive.org/we b/20190417192940/https://www.nada.org/WorkArea/DownloadAsset.aspx?id=21474837361, 04-17-2019, Last accessed on 2023-06-11.

National Automobile Dealers Association. 2019*b*. "NADA Data 2018: Annual Financial Profile of America's Franchised New-Car Dealerships." https://web.archive.org/web/20190417205114/https://www.nada.org/WorkArea/DownloadAsset.aspx?id=21474857318, 04-17-2019, Last accessed on 2023-06-11.

Scott Morton, Fion, Florian Zettelmeyer, and Jorge Silva-Risso. 2003. "Consumer Information and Discrimination: Does the Internet Affect the Pricing of New Cars to Women and Minorities?" *Quantitative Marketing and Economics*, 1(1): 65–92.

Shiller, Benjamin Reed. 2020. "Approximating Purchase Propensities and Reservation Prices from Broad Consumer Tracking." *International Economic Review*, 61(2): 847–870.

Shiller, Benjamin Reed. 2022. "Discreet Personalized Pricing." Brandeis University, Department of Economics and International Business School Working Papers 133.

Thomadsen, Raphael. 2005. "The Effect of Ownership Structure on Prices in Geographically Differentiated Industries." *The RAND Journal of Economics*, 36(4): 908–929.

Waldfogel, Joel. 2015. "First Degree Price Discrimination goes to School." *The Journal of Industrial Economics*, 63(4): 569–597.

Wardlaw, Christian. 2019. "10 Most Popular Trucks." https://www.jdpower.com/cars/trucks/10-most-popular-trucks, Last accessed on 2023-06-09.

Ye, Junting, and Steven Skiena. 2019. "The Secret Lives of Names? Name Embeddings from Social Media."

Ye, Junting, Shuchu Han, Yifan Hu, Baris Coskun, Meizhu Liu, Hong Qin, and Steven Skiena. 2017. "Nationality Classification Using Name Embeddings." *CIKM '17*, 1897–1906. New York, NY, USA:Association for Computing Machinery.

Zettelmeyer, Florian, Fiona Scott Morton, and Jorge Silva-Risso. 2006. "Scarcity Rents in Car Retailing: Evidence from Inventory Fluctuations at Dealerships." National Bureau of Economic Research Working Paper 12177.

A APPENDIX

A.1 Value of consumer information for firms' profits

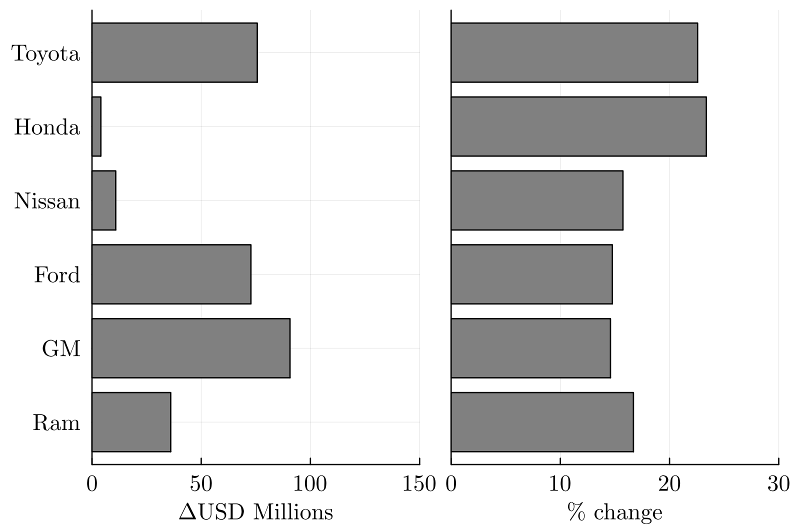

Figure 6: Changes in yearly profit per manufacturer by using discriminatory pricing in $ Millions (left panel) and percentages (right panel) when compared to uniform pricing.

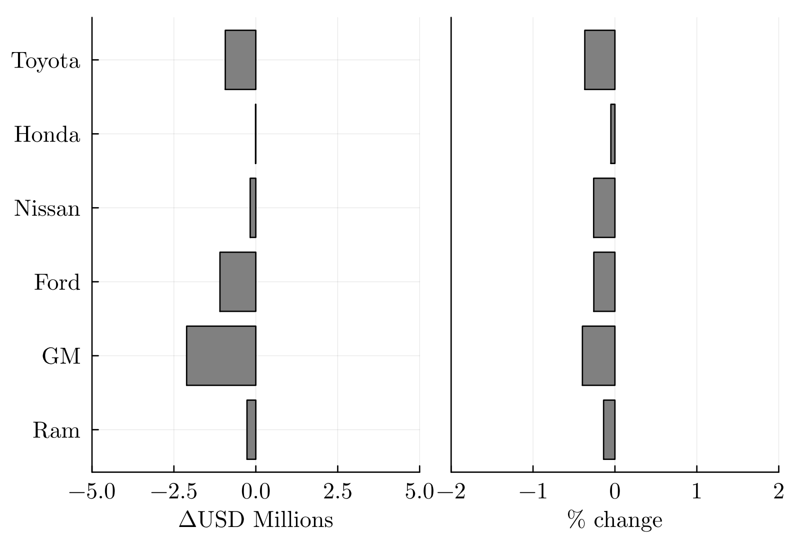

Figure 7: Changes in yearly profit per manufacturer by using discriminatory pricing based only on demographics in $ Millions (left panel) and percentages (right panel) when compared to uniform pricing.

A.2 Why does consumer surplus decrease under uniform pricing if most people are better off?

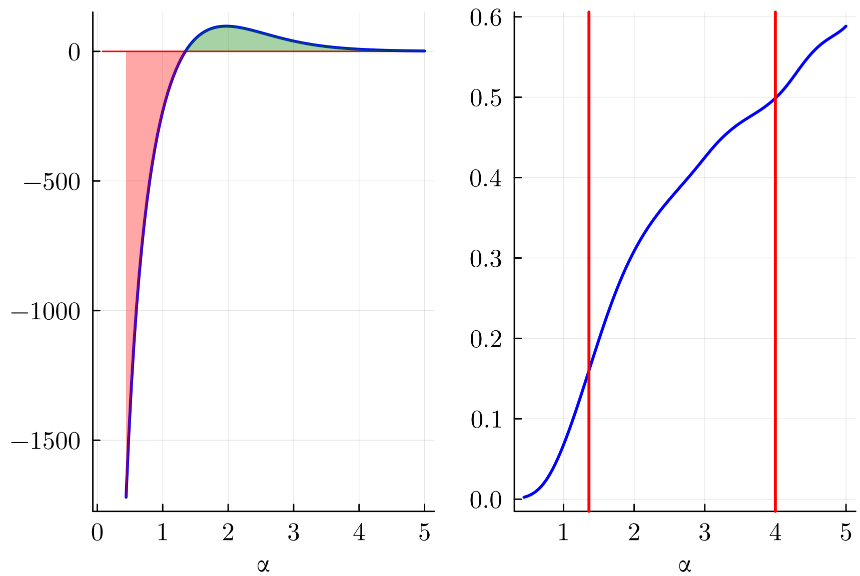

Figure 8: Changes in Consumer Surplus due to Price Discrimination as function of Price Sensitivity

How could total consumer surplus decrease if only a small fraction of consumers are worse off and most are quoted lower prices by dealers? Figure 8 explores this question more thoroughly. The left panel of figure 8 shows the change in consumer welfare as a function of individual price sensitivity α_{it}, the right panel shows the distribution of price sensitivities. The red area between the curve and the horizontal line marking no change in consumer surplus marks the losses in consumer surplus by price-insensitive consumers. Due to dealers' ability to infer consumers' willingness to pay, price-insensitive consumers face much higher prices for pickup trucks under price discrimination than under uniform prices. On the other hand, the area shaded in green represents the gains for more price-sensitive consumers. However, the gains for very price-sensitive consumers from discriminatory pricing quickly become negligible with α_{it}. This is because these consumers rarely buy a pickup truck model, even when

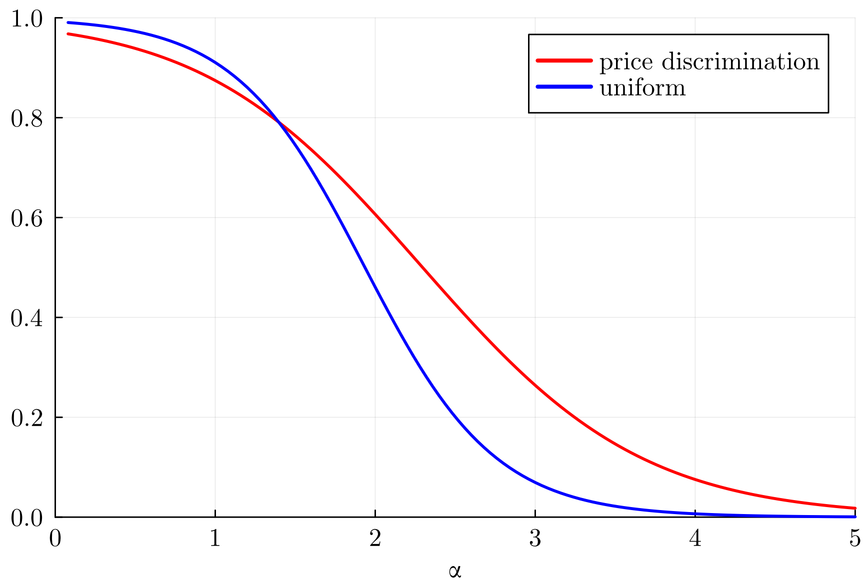

Figure 9: Changes in probability of purchasing any pickup truck model under price discrimination (red) and uniform pricing (blue) in 2017 for a non-Hispanic male as a function of price sensitivity.

facing low prices.

The two vertical red lines in the right panel of figure 8 mark the cutoffs: the left vertical red line marks the cutoff for consumers losing from dealers engaging in price discrimination, and the right vertical red line marks when gains from price discrimination become negligible.[29] As the right panel shows, there are around 19% of consumers who lose a lot from discriminatory pricing. About 30% of consumers gain little from discriminatory pricing, while the rest of consumers see only a very small but positive change in their consumer surplus. From this figure, it is clear that the losses of the price-insensitive consumers outweigh the gains from most consumers. Thus, consumer surplus decreases.

While figure 8 illustrates that consumer surplus decreases as a result of dealers engaging in price discrimination, figure 9 shows why some consumers are better off, and some are worse off. Figure 9 plots the probability of buying any pickup truck

[29]The right cutoff is based on my definition and arbitrary since gains never become exactly 0.

model in a given year as a function of price sensitivity for a non-Hispanic male in 2017. The figure shows that the probability of purchasing any pickup truck model under price discrimination is rotated around a point for which consumers are indifferent between dealers engaging in price discrimination and uniform pricing. When dealers price discriminate, price-insensitive consumers' probability of purchasing any pickup truck model decreases. On the other hand, the probability of purchase increases for more price-sensitive consumers. Since compensating variation is a function of purchase probabilities, less price-sensitive consumers lose while more price-sensitive consumers benefit.