Demographic Preferences and Price Discrimination in New Vehicle Sales*

Ashley Langer
University of Michigan

October 2011

Abstract

Understanding why different demographic groups pay different prices for the same goods is central to questions of consumer welfare and equity in many markets. In a monopolistically competitive setting like the new car market, sellers have an incentive to charge higher prices to consumers with more inelastic demand. Unlike seller animus or differences in bargaining skills, such "third-degree" price discrimination implies a distinctive pattern of *product-specific* price differentials across groups. This paper investigates the importance of third-degree price discrimination in the new vehicle market in the U.S. Specifically, I use micro data for a large sample of recent buyers to estimate separate random-coefficient discrete choice models by marital status and gender, and calculate optimal markups for each group. I find that controlling for optimal markups substantially changes our understanding of the correlation between prices paid and demographics, suggesting that both price discrimination and differences in negotiating ability across demographic groups are important in understanding the prices diffferent groups pay in the new car market.

*I would like to thank David Card, Patrick Kline, Enrico Moretti, Kenneth Train, Clifford Winston, and Catherine Wolfram for their guidance throughout this project. Dan Ackerberg, Megan Busse, Fred Finan, Jeremy Fox, Brad Howells, Ryan Kellogg, Ulrike Malmendier, Nathan Miller, Asaf Plan, Kevin Stange, and seminar participants at UC Berkeley, the San Francisco Federal Reserve, Collegio Carlo Alberto, the University of British Columbia, Arizona State University, Tufts University, the University of Toronto, Columbia University, the Ohio State Univsersity, the University of Wisconsin, Stanford University, the University of Maryland, and the University of Michigan provided valuable comments and discussion. Andrew Lumsden and Zachary Anderson were extraordinarily helpful with procuring the data.

1 Introduction

Why do different demographic groups pay different prices for the same goods? Three primary explanations have been proposed in the economics literature, with different implications for consumer welfare and equity. Sellers may dislike interacting with certain groups and in equilibrium charge them higher prices (Becker, 1957). Alternatively, some groups may have different negotiating abilities or search costs and therefore pay more for all vehicles (Babcock and Laschever, 2003). Finally, sellers may use consumers' demographics to infer their preferences and practice third-degree price discrimination. In a market with a homogeneous product, all three explanations predict a difference between groups in average prices and are therefore hard to distinguish empirically. In a differentiated product market, however, third-degree price discrimination implies a distinctive pattern of product-specific price differences across groups that are related to their relative demand elasticities. Thus, even if average prices paid by women (for example) reflect some degree of seller animus, or limited negotiating skills, the impact of third degree price discrimination can be identified from the relative variation of prices and elasticities across different products.

This paper investigates the role of third-degree price discrimination in the market for new automobiles by estimating group-specific product demand functions and comparing the correlation between price and demographic characteristics that is estimated in the literature to this correllation once I controll for predicted markups. The new vehicle market is well suited to this investigation because prices are set by individual negotiation and vary across consumers, there are many different (but closely substitutable) models, and different buyers appear to have distinct valuations for each product attribute. Moreover, previous research has shown that a random-coefficient model of consumer choice among vehicle models can capture many of the most important features of the structure of demand in this industry.[1] When combined with a simple model of price-setting, these models yield good descriptions of the determinants of price and quantity in the market.

This approach extends the literature on discrimination in new vehicle sales. Earlier papers have focused on measuring the difference in the average price paid by demographic groups controlling for vehicle attributes. This work has included both paired audit studies (Ayers and Siegelman (1995)) and cross-sectional investigations (Goldberg (1996)). Researchers have looked at whether dealer profit varies over consumers of different demographic groups (Harless and Hoffer (2002)) and whether negotiating online rather than in person changes prices paid differentially across demographic groups (Scott Morton, Zettelmeyer, and Silva-

[1]See, for example, Berry, Levinsohn, and Pakes (1995, henceforth BLP), Berry, Levinsohn, and Pakes (2004, henceforth MicroBLP), and Train and Winston (2007).

Risso (2003)). All of these papers have acknowledged that while they are measuring differences in the treatment of different demographic groups in the market, there are multiple explanations for why those differences may arise. Yet these papers look at the average price paid (or profit made) by demographic group controlling for vehicle attributes, rather than looking at the product-specific price differences across demographic groups.[2] My approach of estimating demand functions for different demographic groups investigates price differences at a finer level than has been done in the past.[3] This approach could also be useful for understanding the determinants of price differences in markets like housing (Yinger (1998)) and loans (Charles, Hurst, and Stephens (2008)).

This estimation relies on the assumption that the demographics of the person purchasing the new vehicle are the same as the demographics of the person whose preferences lead to the vehicle choice. If consumers rely on friends or family members to purchase a new vehicle for them, this assumption could be problematic. I circumvent this concern by using a unique survey dataset of new vehicle purchasers from the second quarter of 2005. I use only those 12,014 new vehicle purchasers who confirmed that they were "both the principle buyer and driver" of a new vehicle, provided full demographic information, and reported the price they paid for their new vehicles. I augment this data with information from the Current Population Survey on the total number of Americans in each demographic group in order to include information on consumers who chose not to purchase a new vehicle in the second quarter of 2005.

Using this dataset, I estimate separate random coefficient discrete choice models for gender and marital status groups.[4] Consumers are assumed to have utility specifications similar to Berry, Levinsohn, and Pakes (1995, henceforth BLP) and Berry, Levinsohn, and Pakes (2004, henceforth MicroBLP), but estimation is via maximum likelihood in the style of Train and Winston (2007) paired with the Berry (1994) market share inversion. This approach allows each demographic group to value all vehicle attributes, including those unobserved to the econometrician, differently from other demographic groups. It also allows for consumer heterogeneity within each demographic group. Using these demand estimates, I calculate the optimal markup for each firm to charge each demographic group for each vehicle, and I use this estimate to calculate the extent to which the price paid exceeeds the

[2]Goldberg (1995) looks at the variance in prices paid by demographic groups and concludes that the results suggest that differences in negotiating behavior may play a role in generating different price distributions across demograhpic groups.

[3]A similar approach has been taken in Verboven (1996) and Goldberg and Verboven (2001) to investigate price discrimination across country borders, but not to distinguish between consumers in the same area.

[4]Data limitations require that the demographic groups be large in order to facilitate estimation, which excludes some interesting demographic variables, such as race, from analysis. In the appendix I also consider gender and household income groups and get very similar results.

optimal manufacturer markup, a term which will include the marginal cost of the vehicle, any taste-based or animus costs faced by the dealer, as well as any profits the dealer can extract through negotiation. I then estimate whether these prices in excess of optimal markups vary with consumer demographics which would indicate that there are differences in negotiating ability or dealer animus across different demographic groups.[5]

I find that vehicle preferences do vary substantially over different demographic groups. Single women are the most price sensitive, while married women are substantially less price sensitive. Medium SUVs have broad appeal across demographic groups, while expensive luxury vehicles and sport utility trucks (crosses between SUVs and pickups) do not appeal to consumers on average. However, preferences for vehicles vary substantially within demographic groups as well, especially with respect to differences in household income. The estimated preferences yield reasonable estimates of optimal dealer markups over marginal cost that are generally in line with estimates of markups used by the United States Environmental Protection Agency to estimate the price impact of increased emissions technologies (RTI International, 2009).

When I control for the optimal markup that manufacturers should charge consumers of different demographic groups, I find that single men pay slightly less than single women while married people pay substantially less than single people. Interestingly, in contrast to the pattern with single consumers, married women pay less than married men once I control for optimal markups. This pattern seems to suggest that both the ability to obscure one's demographic group by sending a friend or family member to negotiate for a new vehicle and differences in negotiating ability across groups are important for understanding the variation in prices controlling for preferences.[6] These results also suggest that earlier papers which found small differences in prices paid across demographic groups were not finding an absence of differential treatment of consumers by demographics but rather were finding offsetting effects of optimal third-degree price discrimination and differences in negotiating ability.

The remainder of the paper is organized as follows: section 2 lays out some simple correlations between vehicle prices and consumer demographics in the spirit of the earlier literature to motivate a more preference-oriented approach to the data. In section 3 I describe my empirical specification. The data used is explained in detail in section 4. I then present

[5]This will assume that each vehicle's marginal cost is the same regardless of the demographic group that purchases the vehicle.

[6]I limit my analysis to consumers who report being both the "principle buyer and driver" of the new vehicle. While this allows me to match consumer demographics to the purchase price, it also induces some selection based on preference for negotiation that should be particularly strong for married couples. This selection is what leads to the differential result of married women paying less than married men while single women pay more than single men.

results of the demand estimation and the correlations between price in excess of markup and consumer demographics in section 5. Section 6 concludes.

2 Price Correlations

In the literature on price discrimination in new vehicle sales, the standard approach has been to examine the correlation between vehicle prices (negotiated or transaction) and consumer demographics while controlling for vehicle attributes.[7] These studies generally find insignificant or small differences between the prices paid by consumers of different genders and have mixed results on the differences in prices paid by consumers of different races. Table 1 shows these correlations with my price data (which will be described in more detail in Section 4), controlling for vehicle model fixed effects. As in earlier studies, I do not find a significant difference in the prices paid by men and women regardless of the other demographics that are included. Income significantly increases the transaction price, and, when they are included together, being single or high income increases the price, and being more educated decreases the price paid.[8]

The first response to the lack of significant differences between men and women in price paid for new vehicle might be to suspect that men and women are being treated similarly in this market. Yet the fact that marital status, income, and education do seem to affect transactions prices when included together does suggest that prices are not set uniformly across consumers.[9] One way to ask whether prices differ between men and women is to conduct an F-test of whether there is a difference between the prices men and women pay for each vehicle model. This test is overwhelmingly rejected (F(210,10607) statistic = 1.80, Pr>F = 0.0000), suggesting that the average prices men and women pay for each model are quite different, even if these differences are not statistically different from zero when averaged across all models.[10]

Yet just because men and women appear to be paying different prices for the same vehicles does not mean that there is some sort of systematic discrimination (taste-based or statistical) leading to these differences. Goldberg (1996) includes vehicle segment indicator variables as well as those segment indicators interacted with a minority indicator variable,

[7] See, for example, Ayers and Siegelman (1995), Goldberg (1996), Harless and Hoffer (2002), or Scott Morton, Zettelmeyer, and Silva-Risso (2003).

[8] The results presented in Table 1 are given in levels, but the results are fundamentally unchanged when run with the natural log of price paid.

[9] These correlations could also potentially result from differential pricing across markets.

[10] Similarly, when I compare high and low income groups (those with household incomes above or below \$100k) I not surprisingly reject that they pay the same amount for all vehicles. F(210,10607) statistic = 1.91, Pr>F=0.0000

and finds insignificant results on the interaction for most segments. In Table 2 I present the coefficients on these interactions between segment indicator variables and indicators for male, marital status, and high income (households with annual income over $100k) for my data. The results by gender are particularly interesting. Men pay significantly less than women for compact cars, while men pay significantly more than women for luxury cars, sport utility vehicles, and sporty cars. This pattern of women paying more for smaller, less expensive vehicles with men paying more for luxury and performance vehicles might start to suggest that differences in preferences between vehicles allow dealers to statistically discriminate based on gender. Clearly these differences are merely suggestive, but they tend to contradict the idea that a lack of correlation between price paid and gender on average is evidence of a lack of price discrimination based on gender. In order to better understand whether demographic group preferences are playing a role in dealers' pricing decisions, I will have to directly estimate the preferences for different demographic groups.

3 Empirical Specification

The empirical model aims to compare the firms' optimal price for each demographic group under third-degree price discrimination to the observed price for that group in order to identify the extent of price discrimination in the market. This requires estimating the vehicle demand functions of each demographic group for each vehicle and pairing these estimates with a model of manufacturer and dealer pricing behavior. In this sense, the model is similar to work done on international price discrimination in the European market by Verboven (1996) and Goldberg and Verboven (2001).

The data I use will be described in detail in section 4, but it consists of individual-level data on vehicle purchasers' demographics and the price paid for the vehicle. I also observe other vehicles considered by most consumers, allowing for a ranking of both the first and second choice vehicle for those consumers. This will allow me to directly tie the demographics of each consumer to the vehicle choice, thus allowing preferences to vary based on both observable (to the econometrician) and unobservable factors. Here I will describe the demand and supply approaches before explaining how I will incorporate these preference estimates into a model of third-degree price discrimination.

3.1 Demand Functions

The demand function follows directly from MicroBLP but is estimated using maximum likelihood as in Train and Winston (2007). I use the Berry (1994) market share inversion to

reduce the dimensionality of the coefficient space.

Consumers are each assumed to belong to a single demographic group, $d = 1, ..., D$. Within these demographic groups, consumers are heterogeneous along both observable and unobservable individual characteristics. Consumer i's utility for vehicle $j = 0, 1, ..., J$ is assumed to be:

$$U_{idj} = p_{jd}\tilde{\alpha}_{id} + \sum_k x_{jk}\tilde{\beta}_{idk} + \xi_{dj} + \epsilon_{idj} \tag{1}$$

where p_{jd} is the price charged to i's demographic group d; $x_{j1}, ..., x_{jK}$ are the non-price attributes of vehicle j; ξ_{dj} is the preference of demographic group d for the unobservable attributes of vehicle j; and ϵ_{idj} is an extreme value type 1 residual preference parameter. The $\tilde{\alpha}_{id}$ and $\tilde{\beta}_{idk}$ are the individual's preference for vehicle attributes p_{jd} and x_k respectively, which are assumed to have the form:

$$\begin{aligned} \tilde{\alpha}_{id} &= \bar{\alpha}_d + \sum_r z_{idr}\alpha^o_{dr} + \nu_{idp}\alpha^u_d \\ \tilde{\beta}_{idk} &= \bar{\beta}_{dk} + \sum_r z_{idr}\beta^o_{dkr} + \nu_{idk}\beta^u_{dk} \end{aligned} \tag{2}$$

Thus the individual's preference for vehicle attribute x_k is decomposed into a component ($\bar{\beta}_{dk}$) that is constant within that individual's demographic group, a component (β^o_{dkr}) that varies with consumer characteristics z_{idr} that are observed by the econometrician, and a component (β^u_{dk}) that varies with consumer characteristics ν_{idk} that are unobserved by the econometrician, but are assumed to have a known distribution.[11] These unobserved consumer characteristics capture the fact that there is heterogeneity in preferences for different vehicle attributes in every demographic group, although we may not have variables that allow us to identify those consumers who get particular utility from horsepower, for instance, rather than fuel economy.

As noted, consumer demographics enter the model through both the d term that defines demographic groups and the z_{idr} term. The difference between these two components of preference is whether the demographic characteristic is observed by the dealer and therefore a potential basis for price discrimination. Since the objective of the model is to understand price discrimination based on demographics that are observable to the dealer, estimated preferences are allowed to be substantially more flexible in d than in z_{idr} (all coefficients may

[11] I will generally assume that unobserved consumer characteristics have normal distributions.

vary in d while only the preferences for price and certain vehicle characteristics may vary in z_{idr}). In particular, the unobservable quality of the vehicle, ξ_{dj}, varies in d but not in z_{idr}. As discussed later, gender and marital status will be considered observable to the dealer, and thus enter as a part of d, while income will be considered unobservable to the dealer and enter as a part of z_{idr}.[12] Variations of these assumptions are tested in the appendix.

Combining equations (1) and (2) leads to the consumer's choice model:

$$U_{idj} = \delta_{dj} + \sum_r p_{jd} z_{idr} \alpha^o_{dr} + \sum_{k,r} x_{jk} z_{idr} \beta^o_{dkr} + p_{jd} \nu_{idp} \alpha^u_d + \sum_k x_{jk} \nu_{idk} \beta^u_{dk} + \epsilon_{idj} \tag{3}$$

$$\text{where} \qquad \delta_{dj} = p_{jd} \bar{\alpha}_d + \sum_k x_{jk} \bar{\beta}_{dk} + \xi_{dj} \quad \text{for each} \quad j = 1, ..., J \tag{4}$$

The consumer chooses the vehicle $j = 1, ..., J$ or the outside option ($j = 0$, not purchasing a new vehicle) that maximizes this utility function.[13] As this notation makes clear, there is a component (δ_{dj}) to each individual's utility for each vehicle that is common across all members of his or her demographic group d. Additionally, the term $\sum_r p_{jd} z_{idr} \alpha^o_{dr} + \sum_{kr} x_{jk} z_{idr} \beta^o_{dkr}$ allows consumers with different observable characteristics to have different tastes for certain vehicle attributes, and thus specifies the extent to which vehicle substitution varies with observable consumer demographics. Finally, there is a component of consumer preference ($p_{jd} \nu_{idp} \alpha^u_{dk} + \sum_k x_{jk} \nu_{idk} \beta^u_{dk}$) that is unobserved by the econometrician but helps explain why certain consumers have stronger preferences for some vehicle attributes than other consumers, and helps to explain why individuals may substitute more strongly between certain vehicles. The β^u_{dk} and α^u_d coefficients can be thought of as representing the standard deviation in the unobserved preference within demographic group d for the vehicle attribute conditional on the consumer's observed attributes. For notational ease, I define the vector of distributional coefficients $\theta_d \equiv [\alpha^o_{dr},\ \alpha^u_{dr},\ \beta^o_{dkr},\ \beta^u_{dkr}]'$.

I estimate the θ_d and δ_d coefficients via maximum-likelihood. The extreme-value error term guarantees that the probability of vehicle j maximizing consumer i's utility conditional on the observable attributes of the vehicle (p_{jd}, x_{jk}) and the consumer's observable (z_{idr}) and unobservable ($\nu_{id} = [\nu_{idp}, \nu'_{idk}]'$) characteristics has a closed form such that the expected probability of a consumer i in demographic group d choosing vehicle j can be expressed as

[12] Demographics that are not directly observable to the dealer, z_{idr}, enter into price discrimination in expectation. Therefore, if married men are generally for higher income households than single women, the expected household income will enter into price discrimination but the variation in household income across married men will not.

[13] The outside option of not purchasing a new vehicle is assumed to have utility equal to $U_{id0} = \beta^o_d(\frac{1}{\text{Income}_i}) + \beta^u_{d0}\nu_{id0} + \epsilon_{id0}$, where ν_i is a draw from a standard normal distribution and ϵ_{id0} is a draw from an EV1 distribution.

an integral over the distribution of ν_{id}. Because the θ_d coefficients determine how consumers substitute between vehicles as attributes change, information on consumers' first and second choice vehicles aids identification of θ_d. Thus, the joint probability that consumer i chooses vehicle $j = 1$ out of the full choice set, and $j = 2$ out of the choice set with $j = 1$ and the outside good removed is:[14]

$$Pr_{i1}(\theta_d, \delta_d) Pr_{i2}(\theta_d, \delta_d|1) = \int \frac{\exp(V_{id1}(\nu_{id}; \theta_d, \delta_d))}{\sum_{l=0}^{J} \exp(V_{idl}(\nu_{id}; \theta_d, \delta_d))} \left(\frac{\exp(V_{id2}(\nu_{id}; \theta_d, \delta_d))}{\sum_{l=2}^{J} \exp(V_{idl}(\nu_{id}; \theta_d, \delta_d))} \right) f(\nu) d\nu$$

where $V_{idj}(\nu_{id}; \theta_d, \delta_d)$ is the non-stochastic component of consumer i's utility for vehicle j from equation (3). To condense notation, I write $V_{idj}(\nu_{id}; \theta_d, \delta_d)$ and $Pr_{idj}(\nu_{id}; \theta_d, \delta_d)$ with the understanding that the non-stochastic utility and probability are also a function of the observable data. Since the probability of observing a particular first and second choice for an individual is conditional upon the individual's ν_{id} vector, the integration over the distribution of ν must be over the joint probability of the first and second vehicle choices. I approximate this integral using simulation, and then sum the log of this probability over consumers i in demographic group d to calculate the log-likelihood function.[15]

The log-likelihood function is maximized over θ_d. For each value of θ_d, I choose δ_d to set the predicted market shares for each demographic group equal to the observed market shares for that group as in Berry (1994):

$$\begin{aligned} S_{dj} &= \int_{z_{idr}} \int_{\nu} Pr_{idj}(\theta_d, \delta(\theta_d)) f(\nu) f(z_{idr}) d\nu dz_{idr} \\ &= Pr_{dj}(\theta_d, \delta(\theta_d)) \end{aligned} \tag{5}$$

where $f(z_{idr})$ is the pdf of the consumer characteristics z_{idr} in the demographic group d. Therefore it should be understood that the δ_d vector is estimated conditional on θ_d and is thus formally $\delta(\theta_d)$. The maximum-likelihood procedure solves for the value of θ_d that maximizes the likelihood function subject to a market-share constraint that is a function of both θ_d and $\delta(\theta_d)$.

This model differs from previous random-coefficient demand models in an important way: the model interacts demographic group membership, d, with the full distribution of

[14] I remove the outside good from the second-choice choice set because the second choice information is based on the vehicle the consumer said she considered but did not purchase. It is not clear whether she would have purchased the second choice vehicle if the first choice were not available (she may not have purchased any vehicle), but it is her preferred alternative out of the set of vehicles once her first choice is removed.

[15] Simulation generally uses 100 scrambled Halton draws to approximate the integral for each consumer.

preferences rather than limiting the influence of demographic group membership to only one or two vehicle characteristics (the demographics z_{idr} are still only interacted with a few vehicle characteristics). This dramatically increases the flexibility of the model to capture preference differences across the demographic groups that are observable to the dealer.[16] While this means that demographic groups may value the observable (to the econometrician) attributes of the vehicles differently, it is particularly important that the unobservable (to the econometrician) characteristics of a vehicle (ξ_{dj}) are also allowed to be valued differently by members of different demographic groups. A prime example of such varying preference for vehicle unobservables would be the vehicles that are commonly referred to as "chick cars" or "guy cars"[17] such that the opposite gender might be interested in the vehicle for its physical attributes, but dissuaded from buying the car because of its social connotation. Additionally, options packages that appeal to one group rather than another (for instance spoilers or wheel rims) would potentially change the unobservable quality of the car for different groups differently.

Once I have estimated θ_d and calculated $\delta(\theta_d)$, I can use the δ_d vector to extract information about the $\bar{\alpha}_d$ and $\bar{\beta}_{dk}$ coefficients rather than just the θ_d coefficients. Recall from equation 4 that unobservable vehicle quality, ξ_{dj} may include vehicle attributes that allow firms to charge more for the vehicle. Therefore, an OLS regression of the δ_{dj} vector on vehicle price and attributes will estimate that consumers are less price sensitive than they actually are. In order to correct for this bias, I run a weighted IV regression of δ_{dj} on the vehicle price and attributes.[18] I use the standard Bresnahan (1987)/BLP instruments plus a distance instrument used in Train and Winston (2007):

$$\sum_{l \in f_j, l \neq j} x_{lk}, \qquad \sum_{l \in f_j, l \neq j} (x_{jk} - x_{lk})^2, \qquad \sum_{l \notin f_j} x_{lk}, \qquad and \qquad \sum_{l \notin f_j} (x_{jk} - x_{lk})^2 \tag{6}$$

which are the sum of each vehicle attribute for competing vehicles produced by the same firm as vehicle j, f_j, the sum of each vehicle attribute for competing vehicles produced by other firms, and the sum of the distance in attribute space between the vehicle and all others sold by the same firm and all others sold by other firms. These instruments are intended to capture the extent of price competition faced by vehicle j in the market. For instance, if a vehicle is competing with a set of vehicles that have particularly high horsepower, then competitive pressure will keep the vehicle's price fairly low conditional on its attributes. If

[16] This also means that the preferences in the population (as estimated in the previous discrete choice literature) are a combination of the preferences in each demographic group.

[17] See, for instance: http://www.cartalk.com/content/features/Guy-Chick-Cars/index.html

[18] The weights are equal to the number of consumers of demographic group d who chose vehicle j.

the observed price is actually high conditional on attributes, it must be that the vehicle has a high level of of unobservable quality that is increasing its demand.

Because demographic groups face different prices and value vehicle attributes differently, the competitive pressure on price created by competing vehicles' attributes should vary over demographic groups. Therefore the instrumental variables regression is run separately for each demographic group.

The estimated demand coefficients allow me to calculate demand elasticities. Because the predicted demand of demographic group d for vehicle j is just the number of people in group d times the predicted market share of group d for vehicle j, the own-price elasticity of demand is just:

$$\frac{\partial Pr_{dj}(\hat{\theta}_d)}{\partial p_{dj}} \left(\frac{p_{dj}}{Pr_{dj}(\hat{\theta}_d)} \right) \tag{7}$$

which is straightforward to calculate given $\hat{\theta}_d$. This is the key formulation of demand for firms that are choosing prices to maximize profits.

3.2 Supply

I pair this demand specification with a model of vehicle supply that closely follows BLP, MicroBLP and Bresnahan (1981). Firms maximize profits over the set of vehicles they offer, which are assumed to have constant marginal cost. The equilibrium is Nash in prices. The complication from the standard supply model is that firms choose an optimal price for each vehicle to offer to *each* demographic group. I explicitly assume that "firms," which include both the manufacturer and its dealer network, are able to charge optimal prices to each demographic group. This means that there is perfect contracting between the manufacturer and its dealers; dealers are perfectly able to identify the demographic group of each consumer, and consumers cannot engage in across-demographic-group arbitrage. This final assumption is stronger than the typical no-arbitrage assumption where consumers are assumed to not participate in a secondary resale market.[19] In this case consumers are also assumed not to obscure their demographic characteristics by sending someone of a different demographic

[19]Note that the used car market would not function as a means of arbitrage in this case because the price difference between a new and a barely used vehicle is substantial, likely stemming from consumers fear that a brand new vehicle that is on the used car market has shown itself to be a lemon (Akerlof, 1970). In particular, the gain from reselling a car to a different demographic group is small relative to the loss of selling a "used" car rather than a new one.

group to purchase the new vehicle for them.[20]

Thus firms $f = 1, ..., F$ set prices to maximize profits over the vehicles they sell:

$$\pi_f = \sum_{d=1}^{D} \sum_{j \in f} Q_{dj}(p_d)(p_{dj} - c_j - D_d) \tag{8}$$

where the demand function of demographic group d for vehicle j, $Q_{dj}(p_d)$, is a function of the vector of the demographic group's prices for all vehicles, p_d. I allow for the possibility of animus or differences in average bargaining abilities by including a fixed cost of selling to demographic group d, D_d. The maximization of this set of profit functions for all firms leads to the vector of optimal prices given the vector of marginal costs, c:

$$P_d^* = c + D_d - \Omega_d^{-1} Q_{dj} \tag{9}$$

$$\equiv c + D_d + M_d \tag{10}$$

where P_d^* is the optimal price vector for group d, M_d is the vector of optimal markups, and Ω_d is the matrix of own and cross-price derivatives of demand:

$$[\Omega_{djk}] = \begin{cases} \frac{\partial Q_{dk}(\theta_d, p_d)}{\partial p_{dj}} & \text{if } j \text{ and } k \in F \\ 0 & \text{otherwise} \end{cases}$$

From the demand estimation, I have estimates of $\hat{\theta}_d$ and $\bar{\alpha}_d$, and I can therefore construct estimates of the demographic group's demand and price derivative matrix, $Q_{dj}(\hat{\theta}_d)$ and $\Omega_d(\hat{\theta}_d, \bar{\alpha}_d)$.[21] Thus I have enough information to construct estimates of the optimal markups for each vehicle j sold to demographic group d, $\hat{M}_{dj}$. While I do not have information on the costs of vehicle j, I do assume that the marginal vehicle costs are the same for all demographic groups, and therefore that they may be estimated with vehicle fixed effects.

This supply model assumes that the firms are able to observe the demographic groups, d, perfectly, but that they don't observe any other consumer characteristics, such as those

[20] Recall that the in the estimation I will use data on those survey respondents who said that they are both the principle buyer and driver of the new vehicle. I will also consider the selection that this restriction induces in the sample.

[21] In BLP and MicroBLP, the authors use a moment similar to equation 9 to estimate their model, allowing the cost of each vehicle to be a linear combination of the vehicle's observed attributes. I do not exploit this moment, and therefore do not assume that the observed prices, p_{dj}, are optimal. This leaves open the possibility of animus or differences in average bargaining ability across groups in the observed data.

included in the z vector in the demand specification. I will discuss this assumption, along with the potential for consumers to obscure their demographic group, in the context of my specific choice of d and z characteristics.

3.3 Incorporating Preferences into Demographic Price Correlations

With the optimal markup for each demographic group for each vehicle in hand, I can return to the correlations between transaction prices and consumer demographics from the literature to better understand how prices vary across demographic groups once we consider groups' preferences. Equation 10 shows that the transaction price should be the combination of three terms: the marginal cost of vehicle production and sales, the "cost" of selling vehicles to a particular demographic group (which is constant over vehicles), and the optimal markup.[22] By subtracting the optimal markup from the transaction price, I get a simple functional form that can be used to better understand how transactions prices vary across demographic groups, controlling for the optimal markup:

$$P_{idj} - \hat{M}_{dj} = c_j + D_d + \epsilon_{idj} \tag{11}$$

Here the price paid by individual i from demographic group d for vehicle j minus the optimal markup that the manufacturer would charge demographic group d for vehicle j should equal the cost of selling the vehicle, which includes the marginal cost of production and sales, c_j, as well as the cost of selling a vehicle to demographic group d in particular. There is a residual, ϵ_{idj},which includes and measurement error in the individual's transaction price relative to the optimal price that the manufacturer would like to charge, as well as any estimation error in the $\hat{M}_{dj}$.

This specification has the appeal of following directly from the earlier price discrimination in new vehicle sales literature while incorporating the price that the manufacturers would like to charge consumers of different demographic groups for each vehicle. However, there are a few complications with this functional form that focusing on demographic group preferences highlights. First, I am including the optimal markup as viewed by the manufacturer rather than the dealer while the dealer is actually interacting with the consumers and potentially enforcing the discriminatory pricing schedule. This approach is appealing because, if the manufacturer has control over the extent and location of its dealerships, then the objective should be to charge the optimal price as viewed by the manufacturer. In that

[22] "Cost" is in quotations because it not only includes a taste-based cost component but also any excess profit that can be extracted from a group via negotiation.

context, any deviations from this optimal price schedule might be considered random and enter the residual.[23] Additionally, by allowing costs to vary over demographic groups, this specification makes clear that D_d should be thought of as both any differences across groups in "taste-based" discrimination that affects the dealers' cost of interacting with a particular demographic group as well as any differences in the benefit from negotiating with different demographic groups. Since this specification provides no way to distinguish between dealer animus against a demographic group and a high cost of negotiating with that group, I will be unable to make a strong statement about the reason that D_d varies over groups.[24]

4 Data

The primary data for this analysis is a survey of new vehicle buyers conducted by a major market research firm. This data is augmented with data from the Current Population Survey, the Automotive News Market Data book, and Autodata Solutions.

The survey of new vehicle buyers includes 25,875 respondents who purchased new vehicles in the second quarter of 2005.[25] The survey includes information on the model of vehicle purchased and the other models considered,[26] but does not include information on the trim level or the options packages of the vehicle. The survey asks respondents a series of questions about their purchase, including the price they paid for the vehicle, and whether they paid cash for the vehicle, leased it, or secured a loan. Additionally, respondents indicated their

[23]This assumes perfect contracting between the manufacturer and the dealer. Alternatively, one could structure the model as a double-marginalization problem between the manfuacturer and the dealer. While this might affect the magnitude of the estimates of discrimination, it is unlikely to affect the result that certain demographic groups pay more than others, controlling for optimal markups.

[24]While the above specification follows directly from the model, I will also run a more data-driven specification which allows for the possibility that dealers are unable to fully extract the variation in optimal markups across demographic groups. This could be because of consumer arbitrage (in this context, where a consumer of one demographic group sends a friend or relative to purchase a new vehicle for her) or because of localized competition between dealers of the same manufacturer's vehicles. This specification will merely allow for a coefficient on the estimated optimal markup from equation 10:

$$P_{idj} = c_j + D_d + \gamma \hat{M}_{dj} + \epsilon^*_{idj}$$

In this specification, I will interpret γ a measure of firms ability to price to demographic groups according to their preferences without assuming perfect contracting between manufacturers and dealers.

[25]Because the survey is limited to consumers who purchased new vehicles in the second quarter of 2005 I abstract from concerns about the changing demographic sales patterns over the calendar year that are raised in Aizcorbe, Bridgman, and Nalewaik (2009).

[26]I will follow the standard practice of assuming that the other models considered are listed in the order in which they were considered in order to identify the consumer's second choice vehicle. I only use the second choice information rather than the third and fourth because the number of respondents who entered a third or fourth choice is low. The results using only the first choice information are similar to those presented here.

age, gender, marital status, education, household income, and race on the survey.[27] In a particularly relevant question, the survey asks whether the respondent is both the "principle buyer and driver" of the vehicle. 21,085 respondents indicated that he or she was both the principle buyer and driver, and I will limit my analysis to these respondents in order to assure that the demographic information matches the driver of the vehicle and the person who physically purchased the vehicle.[28]

My analysis will focus on four demographic groups: married and unmarried men and women. These groups are large enough to estimate demand functions for each. Gender and marital status are attractive groups to use for this analysis because they are fairly evenly distributed geographically, so it is likely that all dealerships have experience with consumers of all demographic groups. Additionally, gender is a readily observable variable to dealers and is often thought of as a dimension along which vehicle preferences may vary. Marital status may be less observable to dealers, so any differences in the amount of price discrimination based on marital status relative to gender might be related to consumers' ability to obscure their demographic group. Additionally, to the extent that married consumers are more likely than single consumers to be older and have larger households that potentially include children, I would expect married consumers' preferences to differ from single consumers of the same gender.[29] Additionally, as discussed in section 2 and shown in table 1, these groups seem to pay similar amounts for new vehicles on average, yet men and women pay different amounts for different vehicle classes as shown in table 2 while married and single people do not. Thus if preferences are playing a role in pricing, gender and marital status are particularly interesting places to examine it.[30]

In order to understand preference heterogeneity within these demographic groups, I allow

[27]Consumers were asked to indicate the range in which their education and household income fell, rather than the exact amount.

[28]Of course, many people may take a friend or family member with them to purchase a vehicle, in which case the dealer may not be completely sure who the primary driver of the vehicle will be.

[29]These groups have the advantage of being fairly observable to dealers, but gender and marital status are clearly only a subset of the demographics that a dealer may observe or infer. In this analysis, differences in average income, age, education, and race across these four demographic groups will enter into the mean preference coefficients, $\bar{\alpha}$ and $\bar{\beta}$. I use household income as an observed (to the econometrician but not to the dealer) determinant of consumer heterogeneity within demographic groups, but assume that prices are set for the demographic group as a whole rather than for different income classes within the demographic group. Age, education, and race differences within demographic groups will contribute to unobserved consumer heterogeneity while differences across demographic groups will enter into the mean preference coefficients. These assumptions are further explored in the appendix.

[30]Price discrimination based on race would be of particular interest to policy makers. Unfortunately, there are too few African American new vehicle purchasers in my sample to accurately estimate preferences. Additionally, if we are concerned that African Americans face discrimination in the used vehicle market as well as the new vehicle market (e.g. Charles, Hurst, and Stephens (2008)), then the assumption of the outside good as being the same for both African American and white consumers would not be valid.

preferences for price and the outside good to vary with income.[31] As noted earlier, I assume that dealers are not able to price discriminate on income, although I report a complete set of results with price discrimination based on gender and broad income group (household income above or below $100k) in the appendix.

I remove from consideration any consumers who purchased a vehicle with an average sales price of over 75 thousand dollars in order to limit the analysis to commonly purchased vehicles. Finally, I only consider respondents who reported their gender, marital status, education level, household income level, and the price paid for thier vehicle, which brings my dataset down to 12,014 consumers. 60.28% of my sample is male and 68.40% is married.[32] New car buyers tend to be wealthier than the average American, with 38.39% of respondents coming from households making over $100,000 and only 32.23% coming from households making less than $60,000. 31.00% of respondents in my sample have a college degree and 21.18% have more education than a college degree. In order to calculate prices for each demographic group for every vehicle, I only include vehicles purchased by at least one member of each sex and marital status group. If a consumer purchases a vehicle that is not included in the choice set, then he or she is assumed to have chosen the outside good.

By limiting the data to those consumers who are both principle buyers and drivers of their new vehicles, I do introduce some selection into my analysis. This selection is most likely strongest for married consumers, who have another adult in the household who might negotiate for the new vehicle in the consumer's place, while single consumers may not have another adult who could easily replace him or her in purchasing the vehicle. In fact, of the respondents with complete data, 22% of married men report not being the principle buyer and driver and 11% of married women report not being the principle buyer and driver. For single men, only 5% of respondents are not the principle buyer and driver, and for single women 4% are not the principle buyer and driver. If a survey respondent reports that she is not the principle buyer and driver of the new vehicle, it is impossible to tell whether she is the principle buyer but someone else is driving the vehicle or whether someone else purchased the vehicle for her to drive. Research on women's propensity to avoid negotiation (Babcock and Laschever, 2003) would indicate that married men may be purchasing vehicles for their wives and then sometimes filling out the accompanying survey. This might mean that the women who purchase vehicles for themselves are particularly strong negotiators, leading selection to bias my results towards finding that married men pay more for all new vehicles. Regardless

[31] Income categories include <$25k, $25-30k, $30-40k, $40-50k, $50-60k, $60-100k, $100-150k, and >150k. The income of the lowest category is assumed to be $25k and the income of the highest category is assumed to be $150k. Otherwise, the midpoints of the range are used as the consumer's household income.

[32] Of the 12,014 observations in my sample, 2,852 are married women, 5,366 are married men, 1,876 are single women, and 1,920 are single men.

of the predicted sign of the selection bias, the fact that single people generally are both the principle buyer and driver of the new vehicle would indicate that selection will be low for single consumers. Similarly, if individuals with high household income are more likely to be married, then I would expect the high income demographic group to include the more skilled negotiator of the couple, leading them to potentially pay less than other income groups.

The other type of selection that may enter my analysis is the selection incurred by limiting the dataset to only those consumers who provide full price and demographic information on the survey. If a consumer feels that she got a particularly bad deal on a car she might choose to "forget" how much she paid when it comes time to complete the survey. As long as this selective omission is similar for different demographic groups and similar over vehicles for which a group has different preference intensities, I would not expect missing data to impact my conclusions. However, future work might benefit from using a dataset that is matched to actual transaction data both to remove the possibility of missing values and to confirm that consumers' recollections of the price they paid for their new vehicles matches the actual transaction price.[33]

In order to account for customers who decided not to purchase a new vehicle in the second quarter of 2005, I append observations to my sample with consumers from each demographic group who purchased the outside good. This approach is valid if the population distributions of the observed consumer characteristics, d (demographic group) and z (other consumer characteristics), are known, the consumer characteristics, d and z, have discrete distributions, and the fraction of non-buyers is known for each $\{d, z\}$ cell.

Working through the requirements, I begin by using information from GfK Automotive Research which says that approximately 20% of Americans considered buying a new vehicle in the previous year. I therefore assume that 10% of Americans considered buying a new vehicle in the second quarter of 2005.[34] I assume that that consumer characteristics of this 10% of Americans mirror the attributes of the American population as a whole, as measured by the Current Population Survey for May 2005.[35] These assumptions give me the population distributions of d and z. Since d is by definition a discrete demographic group and z is the consumer's response to discretely-valued survey questions, the consumer characteristics have

[33]The Scott Morton, Zettelmeyer, and Silva-Risso (2003) dataset would avoid the issues of self-reported prices and has substantially more observations, but does not have information directly from the consumer on demographics, consumers' second choices, or whether a consumer is both the principle buyer and driver.

[34]Assuming 5% or 15% of Americans considered buying a new vehicle in the second quarter of 2005 does not change the results substantially.

[35]This assumption mirrors assumptions in BLP and MicroBLP that all Americans consider buying a new vehicle each year, but limits the population to a more relevant group. The other extreme assumption would be that consumers who consider buying a new vehicle have the characteristics of the consumers who do buy new vehicles, although I am not aware of this having been done in the literature.

discrete distributions.

In order to know the fraction of non-buyers in each $\{d, z\}$ cell, I start with information from the Automotive News Market Data Book on the total number of vehicles of each model sold in the second quarter of 2005.[36] I assume that the distribution of consumer characteristics for purchasers of each vehicle is the same as the distribution of consumer characteristics for purchasers of that vehicle in my survey data. By summing the characteristics of consumers over the total number of vehicles sold in the quarter, I then know the number of new vehicle purchasers in each $\{d, z\}$ cell. Thus I have both the total number of consumers in each cell who considered purchasing a new vehicle and the total number who did purchase a new vehicle. The difference is the weight I place on the non-purchase observation for that consumer characteristic cell, thus satisfying requirement 3 above. Thus my data satisfies the four requirements above and the data augmentation procedure gives me a sample of vehicle purchasers and non-purchasers.

I pair this data with data from AutoData Solutions on the attributes of model year 2005 vehicles. This data includes extensive information on the vehicle, including the manufacturer's suggested retail price (MSRP), horsepower, curb weight, wheel base, fuel economy, turning radius, and whether the vehicle has stability control, traction control, or side airbags. This data is at the vehicle trim level, which allows it to differ for the same vehicle model based on differences such as engine type (e.g. V6 vs V8) or body style (e.g. hatchback vs sedan). Since my consumer choice data only specifies a consumer's purchase decision at the model level, I use the vehicle attributes of the trim with the lowest MSRP as the model attributes and consider any deviations from this unobserved quality. This reinforces the idea that consumers of different demographic groups might have different valuations of unobserved quality, since not only the vehicle's styling may be valued differently but also the average trim level chosen may vary by demographic group. To the extent that many options such as leather seats, rear spoilers, or sunroofs may be fairly inexpensive to produce but command a large markup, these options packages may be a way for firms to encourage consumers to self-select into options packages that are priced to further price discriminate.[37]

[36] Note that in the second quarter of the year almost every vehicle has the model year equal to the calendar year, which alleviates the issue of the mix of model years of vehicles sold.

[37] Although this is technically second-degree price discrimination, where firms configure product offerings such that consumers will sort by willingness to pay, I will estimate it as a part of what I call third-degree price discrimination. Ireland (1992) provides an interesting discussion of the identification of this type of second-degree price discrimination using the assumption that costs are linear in vehicle attributes.

4.1 Posted Prices

While my data includes the actual transaction price for the vehicle purchased, I clearly cannot observe the price that each consumer would have negotiated for every vehicle that she did not purchase. This is a long-standing problem in the vehicle-choice literature, to the extent that the nearly every discrete-choice vehicle demand paper uses the manufacturers' suggested retail price for all vehicles as the price the consumer actually pays. When studying the variation in prices over consumers, this is clearly not a reasonable approach. Instead, as mentioned in section 3.3, I will use the average price paid by a consumer's demographic group for each vehicle in the choice set as the price the consumer pays for that vehicle. This assumes that the consumer is aware of the price paid for each vehicle by other consumers in her demographic group, and that she choses which vehicle to purchase based on those commonly known prices. Thus any actual variation in the price that a specific consumer would pay relative to other consumers in her demographic group must be random. While this is clearly a strong assumption, it avoids issues of heterogeneity in negotiation across individuals within a demographic group that are beyond the scope of this paper, although potentially interesting for future research.

5 Results

The results are presented in three steps: the demand coefficients are presented first and include the δs (the mean preference of each demographic group for each vehicle), mean preference coefficients (how different vehicle attributes contribute to the δ vector for each demographic group), and the coefficients governing observed and unobserved preference heterogeneity within demographic groups. I then present the elasticities and optimal markups that are calculated from these demand coefficients, and finally I compare the predicted markup differences between pairs of demographic groups to the observed average price differences for those groups.

5.1 Demand Estimation Results

The δ vector of mean preference parameters contains the values of the mean preference for each vehicle for each demographic group that set the predicted market share for each vehicle equal to its observed market share, conditional upon the consumer heterogeneity coefficients. In this respect, the δ vector acts as an adjusted market share, where the adjustment comes from the fact that some consumers with extreme preferences will buy certain vehicles frequently enough to match a substantial portion of the vehicle's market share even when the

mean consumer strongly dislikes the vehicle. A good example of such a scenario occurs with muscle cars like the Chevrolet Corvette or the Pontiac GTO. While the average consumer of any demographic group would find such vehicles unattractive, there may be some consumers in each group with a high demand for vehicle performance and a low preference for fuel economy who find these cars attractive. These mean preference coefficients order vehicles by the preference of the average consumer of each demographic group, making them a useful reality-check before moving on to the coefficient estimates. The highest and lowest five vehicles for each demographic group are listed in Table 3. Across all demographic groups, small and mid-sized SUVs such as the Jeep Liberty and the Honda CR-V frequently appear among the top 5 vehicles in terms of mean preference. Similarly, the Hummer H2 SUT and the Cadillac Escalade EXT[38] frequently appear among the bottom 5 vehicles, likely because of their unusual features and their extremely high price. It is additionally interesting to note some of the differences between different demographic groups that might conform to our priors. Low income women have strong average preferences for inported sedans such as the Nissan Altima and the Toyota Corolla that do not rank highly for other demographic groups. Additionally. the lower income groups, both men and women, have luxury vehicles among their bottom five while the high income groups do not.

Regressing the δ vector on vehicle attributes using weighted instrumental variables generates the mean preference coefficients for each demographic group.[39] In these regressions, I include price (instrumented with the BLP instruments as discussed earlier), whether the vehicle is a car, and if so whether it is a "sporty" car (generally a small, high horsepower car like the Acura RSX, the Mazda Miata, or the VW GTI), and whether the vehicle is a truck. I also include vehicle curb weight, number of passengers, turning radius, and whether the vehicle is manufactured by a domestic manufacturer in the mean preference specification.

Table 4 reports the mean preference coefficients for each demographic group. Married women are the least price sensitive on average while single women are the most price sensitive, although these differences are not substantial. All four groups dislike cars relative to the omitted groups (SUVs and vans), while only married consumers dislike pickup trucks relative to SUVs and Vans. All groups dislike sporty cars even more than they dislike cars (the total preference for sporty cars relative to SUVs and Vans is the sum of the car coefficient plus the sporty car coefficient), but this distaste is particularly intense for married consumers. The preferences for vehicle characteristics are generally what we might expect. All demographic groups like bigger vehicles, as measured by curbweight, which are also gen-

[38]SUT stands for sport utility truck, which is effectively a large SUV with a pickup bed.

[39]Weights are equal to the inverse of the number of consumers purchasing each vehicle in each demographic group in my sample.

erally safer. Conditional on vehicle weight and type, however, consumers dislike vehicles with higher seating capacity, a distaste which is stronger for men than women. All consumers prefer vehicles with smaller turning radii, which likely captures other aspects of the vehicle's performance as well. Finally, all demographic groups prefer imported vehicles in domestic vehicles controlling for attributes, but this preference is particularly strong for women.

The specification of consumer demand heterogeneity includes price, price divided by household income, the four vehicle types (car, truck, SUV, and van) as well as the fuel use (in gallons per mile), horsepower, and curbweight, as well as an indicator variable for the outside good and the outside good divided by household income. I specify all vehicle characteristics that are not interacted with an observable consumer characteristics as having normally distributed unobservable heterogeneity. The coefficients on all of the normally distributed unobservable heterogeneity terms can be interpreted as the standard deviation in the demographic group's preference for the vehicle attribute, while the coefficients on price divided by income and the outside good divided by income are the extent to which the preference for price and the outside good vary with income.

Table 5 presents the coefficient estimates for these consumer heterogeneity terms by income group and gender. The first interesting thing to notice is that there is very little variation in the price coefficient one houshold income is taken into account, but household income dramatically affects consumers' sensitivity to a vehicle's price. The sign of this effect is what would be expected: consumers in lower income households react more negatively to a vehicle's price than consumers in higher income households. There is substantial variation in preferences for vehicle types (SUV, truck, van, car) for all groups except for single women, although single women do show variation in the strength of their preference for cars, and the variation in their preference for vans is large but not statistically significant. Interestingly, the variation in preference for vehicle attributes is less generic across groups. Only married men show large variation in their preference for horsepower, and only men (both single and married) have variation in their preference for fuel use. No group shows substantial variation in their preference for curb weight. Finally, groups generally do have substantial variation in their preference for the outside good, and this variation is also correlated with household income. Women have particularly large variation in their preference for the outside good. Perhaps the most surprising coefficient in the demand estimation results is that consumers from households with lower income have a stronger distaste for the outside good (which, recall includes purchasing a used vehicle or continuing to drive your current vehicle). This result is less surprising in light of the fact that the price of the outside good has already been controlled for in the price interacted with household income coefficient, so this result must be picking up differences in preferences for the attributes of new cars relative to the

outside good or differences in the methods of financing new vehicle purchases. If low income households are using leases to purchase new vehicles while high income households purchase new vehicles with cash, then I would expect a stronger preference for the outside good among high income households as I find here.

5.2 Elasticities and Optimal Markups

When converting these coefficient estimates into estimated markups, a useful statistic with some economic intuition is the aggregate own-price demand elasticity for each demographic group for each vehicle. Table give some descriptive statistics on own-price elasticities and markups across demographic groups. Generally, the elasticities average from approximately 6 (in absolute value) for married men and education groups to over 17 for single consumers. While these elasticities are fairly high, the markups that result from them seem quite reasonable. The average predicted markup across all demographic groups is 30.2% of the transaction price, and the US Environmental Protection Agency uses manufacturer markups of 31.5% to estimate the impact of emissions control regulations (RTI International (2009)). We see relatively high markups relative to transaction prices in the new vehicle markup because the high fixed costs for each model produced prevent firms from bringing vehicles to the market that aren't expected to earn high markups over marginal costs. While the range in the percent markups is fairly large, ranging from 15 to 60%, it is not surprising that these different demographic groups, with substantially different household incomes, are willing to pay substantially different prices for the same vehicles.

5.3 Understanding Variation in Prices

As explained in section 3.3, by subtracting the optimal markup from the price paid by each individual, I can look at how prices paid by consumers of different demographic groups vary, controlling for the optimal markup for the manufactuerer to charge each demographic group. This variation will include the marginal cost of each vehicle (which is assumed to be constant over demographic groups) as well as any taste-based cost of interacting with consumers of different demographic groups. Yet this estimate will also include any differences in dealers' ability to extract higher prices from some groups than others through negotiation, and therefore we must be careful not to interpret differences across demographic groups purely as dealer animus. In fact, I will suggest that many of the results are indicative of differences in bargaining ability being a primary driver of price differences across demographic groups once I controll for optimal markups.

Table 7 shows correlations similar to those presented in Table 1, but this time with price paid minus the optimal markup as the dependent variable. The omitted demographic group is single women, so table 7 indicates that single men pay slightly less than single women, while married men and women pay substantially less than single women, controlling for optimal markups. The fact that married people pay less than single people for the same vehicles indicates that the ability to obscure one's demographic group by sending a spouse or other family member to purchase a new vehicle may be an important way to mitigate price discrimination. It may also indicate that married consumers, who are generally older, more educated, and from higher income households than single consumers are stronger negotiators than single consumers. The idea that marital status might be correlated with income and therefore affecting preferences and negotiating ability via channels other than marriage is explored in greater detail in the appendix.

Interestingly, married men pay more than married women for same vehicles. This pattern seems to indicate the strong role of negotiation in the variation in prices paid by consumers of different demographic groups. If men are generally stronger negotiators than women, then I would expect men to pay less than women for new cars when the sample includes no selection. The closest I can come to a sample that is free of selection based on negotiating ability is with the single male and female purchasers, since they likely do not have a household member of the opposite sex who could purchase a vehicle for them. The result that single men pay less than single women once I control for optimal markups suggests that women may be in a weaker negotiating position than men, as suggested by Babcock and Laschever (2003).[40] Married men and women do have a family member who could purchase the vehicle for them, so we would expect the stronger negotiator to be the likely purchaser of new vehicles for the household. Since my data only contains consumers who are both the primary buyer and driver of the new vehicle, it is likely that the married women in my dataset are particularly strong negotiators relative to married women as a whole. Thus differences in negotiating skill, interacted with the availability of another family member to purchase the vehicle, explains the how men might get better deals on new vehicles than women when they are both single, but worse deals than women when both are married.

Of course, these results do not rule out the possibility that dealers are engaging in taste-based discrimination (animus) against single people, or single women in particular. The fact that women pay less than men when they are married but more than men when they are single seems to argue against gender discrimination in general, but if married women who purchase their own cars are substantially better negotiators than married men who purchase

[40] Alternatively, a type of animus may explain this result if dealers negotiate more aggressively with single women than with single men.

their own cars, then this result may still include some taste-based component. More likely is that there is a systematic increase in the price paid by younger consumers, who are more likely to be single. While there may be some dealers who dislike interacting with younger consumers, the fact that consumers tend to purchase the same brand of vehicle repeatedly (Winston and Train (2007)) would seem to indicate that the long-run profits from selling to a younger consumer are larger than the long-run profits from selling to an older consumer, which might lead to the opposite result. Generally, although I find the evidence to lean toward differences in negotiating ability across demographic groups, I cannot rule out that taste-based discrimination is also playing a role.

Finally, in order to confirm that the prices paid by consumers of different demographic groups are actually varying with the optimal markups, I run the price correlation regression again, but controlling for the optimal markup as an independent variable. The results are shown in column 2 of table 7, and indicate that while prices do rise with optimal markups, the relationship is not one-for-one. This could be because the markups I calculate are optimal for manufacturers, while dealers face slightly different incentives. In particular, dealers are spatially differentiated from each other, but a consumer who is willing to travel can purchase the same vehicle from multiple dealerships. This would decrease the dealers' ability to price discriminate, and could explain why only 39.4% of the variation in the optimal markup is being internalized in the variation in prices. Additionally, this less-than-complete price discrimination could result from consumers ability to obscure their demographic group by sending a friend or family member to negotiate for a vehicle in their place.[41] However, controlling for the optimal markup in this way does not change the basic demographic group results. While single men are still estimated to pay less than single women, the difference is not statistically insignificant. Married men and women are still expected to pay less than single men and women, and that difference, while smaller than before, is still statistically significant.[42] Finally, the pattern of single men paying less than single men but married men paying more than married women remains when I control for the optimal markup as an independent variable rather than subtracting it from the price paid.[43]

[41] Additionally, when the markup is interacted with marital status, markups are transferred into prices at a much higher rate for single consumers than for married consumers, indicating that consumers' ability to obscure their type by sending a family member to purchase a vehicle limits dealers ability to price discriminate.

[42] Of course, all of these results could also be explained by consumers who purchase inexpensive vehicles being stronger negotiators than consumers who purchase more expensive vehicles, but given that I have controlled for the expected price sensitivity of each group, these differences in negotiating ability would have to stem from something other than pure price sensitivity, which seems the reverse of what we would actually expect.

[43] If this regression is run with log price paid as the dependent varible and log markup as an independent variable the relative price results are unchanged and all of the demographic group coefficients are significant

6 Conclusion

This paper explores the role of consumer preferences that vary with demographic group in new vehicle pricing. I find that dealers do appear to price discriminate between consumers based on demographic group preferences, but that this discrimination is mitigated by consumers ability to obscure their preferences by sending a family member to purchase the new vehicle for them. The results also suggest that differences in the prices paid for new vehicles controling for demographic group preferences appear to stem from differences in negotiating ability. While these results do not rule out dealer animus against a particular demographic group, I do not find evidence of such behavior either. Price discrimination based on demographic group preferences offsets some of the price variation across demographic groups that would result from just differences in negotiating ability since low educated, low income individuals tend to be more price sensitive but potentially have lower negotiating ability than those with more education or wealth.

Ideally, future reseach could take this approach to identifying optimal third-degree price discrimination to demographic distinctions where there is substantial concern about taste-based or animus discrimination, such as race. There are very few black new vehicle purchasers, and the possibility of discrimination in the used car market makes preference estimation that assumes a uniform outside good potentially confounding. However, a model that includes both new and used vehicle purchases and transaction costs might be able to understand how the differences paid by consumers of different races are related to differences in vehicle preferences across races.

Finally, this paper develops an approach to identifying the effect of preferences in market outcomes that could be taken to many other settings. Preferences could affect where consumers choose to live, work, and study, and could interact with home sellers, employers, and school admissions committees in ways that obscure whether taste-based discrimination is occurring in a market. Understanding the role of preferences in consumers' decision-making would allow policymakers to better target markets where discrimination is leading to adverse distributional outcomes.

at the 1% level.

References

[1] Aizcorbe, Ana, Benjamin Bridgman, and Jeremy Nalewaik 2009 "Heterogeneous car buyers: a stylized fact" *Finance and Economics Discussion Series* 2009-12, Board of Governors of the Federal Reserve System

[2] Akerlof, George A 1970 "The Market for 'Lemons': Quality Uncertainty and the Market Mechanism" Quarterly Journal of Economics 84(3): 488-500

[3] Akerlof, George A., and Rachel E. Kranton 2000 "Economics and identity" Quarterly Journal of Economics 115(3): 715-753

[4] Altonji, Joseph G. and Charles R. Pierret 2001 "Employer learning and statistical discrimination" *Quarterly Journal of Economics* CXVI: 293-312

[5] Arrow, Kenneth 1973 "The theory of discrimination" in Discrimination in Labor Markets, O. Ashenfelter and A. Rees, eds (Princeton, NJ: Princeton University Press)

[6] Ayres, Ian. 1991 "Fair Driving: Gender and Race Discrimination in Retail Car Negotiations" 104 *Harvard Law Review* 817.

[7] Ayres, Ian and Peter Siegelman 1995 "Race and Gender Discrimination in Bargaining for a New Car" *American Economic Review* 85(3): 304-21

[8] Azar, Omar 2003 "Can price discrimination be bad for firms and good for all consumers? A theoretical analysis of cross-market price constraints with entry and price discrimination" *Topics in Economic Analysis & Policy*, 3(1)

[9] Babcock, Linda, and Sara Laschever 2003 *Women Don't Ask: Negotiation and the Gender Divide* (Princeton, NJ: Princeton University Press)

[10] Berry, Steve, 1994 "Estimating discrete choice models of product differentiation" *RAND Journal of Economics* 25:242-262

[11] Berry, Steve, Jim Levinsohn, and Ariel Pakes. 2004. "Differentiated product demand systems from a combination of micro and macro data: The new car market." *Journal of Political Economy*, 112(1): 68-105.

[12] Berry, Steve, Jim Levinsohn, and Ariel Pakes. 1995 "Automobile prices in market equilibrium" *Econometrica*, 63(4): 841-890

[13] Bresnahan, Timothy F. 1981 "Departures from marginal-cost pricing in the American automobile industry: Estimates for 1977-1978" *Journal of Econometrics*, 17(2): 201-227

[14] Bresnahan, Timothy F. 1987 "Competition and collusion in the American automobile industry: The 1955 price war" *Journal of Industrial Economics* 35(4): 457-482

[15] Bresnahan, Timothy F., and Peter C. Reiss. 1985 "Dealer and manufacturer margins" *The RAND Journal of Economics*, 16(2): 253-268

[16] Borenstein, Severin 1985 "Price discrimination in free-entry markets" *Rand Journal of Economics* 16(3): 380-397

[17] Borenstein, Severin 1991 "Selling costs and switching costs: explaining retail gasoline margins" *RAND Journal of Economics* 22(3): 354-369

[18] Charles, Kerwin Kofi, and Jonathan Guryan 2008 "Prejudice and wages: An empirical assessment of Becker's *The Economics of Discrimination*" *Journal of Political Economy* 116(5): 773-809

[19] Charles, Kerwin Kofi, Jonathan Guryan, and Jessica Pan "Sexism and women's labor market outcomes" Working Paper

[20] Charles, Kerwin Kofi, Erik Hurst, and Melvin Stephens Jr. 2008 "Rates for vehicle loans: race and loan source" *American Economic Review, Papers and Proceedings* 315-320

[21] Corts, Kenneth S. "Third-degree price discrimination in oligopoly: All-out competition and strategic commitment" RAND Journal of Economics 29(2): 306-323

[22] Gneezy, Uri, Kenneth L. Leonard, and John A List 2009 "Gender differences in competition: Evidence from a matrilineal and a patriarchal society" *Econometrica* 77(5): 1637-1664

[23] Goldberg, Pinelopi. 1998. "The effects of the corporate average fuel economy standards in the automobile industry." *Journal of Industrial Economics*, 46: 1-33.

[24] Goldberg, Pinelopi. 1996 "Price discrimination in new car purchases: evidence from the consumer expenditure survey" *Journal of Political Economy*: 622-54

[25] Goldberg, Pinelope and Frank Verboven. 2001 "The evolution of price dispersion in the European Car Market" Review of Economic Studies, 68: 811-848

[26] Graddy, Kathryn 1995 "Testing for imperfect competition at the Fulton fish market" *RAND Journal of Economics* 26(1):75-92

[27] Graddy, Kathryn 1997 "Do fast-food chains price discriminate on the race and income characteristics of an area?" *Journal of Business and Economic Statistics* 15(4): 391-401

[28] Graddy, Kathryn, and George Hall 2009 "A dynamic model of price discrimination and inventory management at the Fulton fish market" *NBER Working Paper* 15019

[29] Harless, David W. and George E Hoffer (2002) "Do women pay more for new vehicles? Evidence from transaction price data" *American Economic Review* 92(1): 270-279

[30] Heckman, James J., and Peter Siegelman 1993 "The Urban Institute audit studies: Their methods and findings" in *Clear and Convincing Evidence: Measurement of Discrimination in America*, M. Fix and R. Struyk, eds. (Washington, DC: The Urban Institute Press)

[31] Holmes, Thomas J. "The effects of third-degree price discrimination in oligopoly" *American Economic Review* 79(1): 244-250

[32] Ireland, Norman J. "On the welfare effects of regulating price discrimination" *Journal of Industrial Economics* 40(3): 237-248

[33] List, John. 2004 "The nature and extent of discrimination in the marketplace: Evidence from the field." *Quarterly Journal of Economics* 119(1): 49-89

[34] Lott, John R. Jr and Russell D. Roberts "A guide to the pitfalls of identifying price discrimination" *Economic Inquiry* 29: 14-23

[35] McFadden, Daniel. 1974 "Conditional logit analysis of qualitative choice behavior" in P. Zarembka (ed.) *Frontiers in Econometrics*: 105-142

[36] Phelps, Edmund S. 1972 "The statistical theory of racism and sexism" *American Economic Review* 62: 659-661

[37] Revelt, D. and K. Train. 1998 "Mixed logit with repeated choices: Households' choices of appliance efficiency level" *Review of Economics and Statistics* LXXX(4): 647-657

[38] RTI International 2009 "Automobile industry retail price equivalent and indirect cost multipliers" prepared for US Environmental Protection Agency. Project No. EPA-420-R-09-003 February 2009

[39] Salop, Steven, and Joseph Stiglitz. 1977 "Bargains and ripoffs: A model of monopolistically competitive price dispersion," *Review of Economic Studies* XLIV: 493-510

[40] Schmalensee, Richard 1981 "Output and welfare implications of monopolistic third-degree price discrimination" *American Economic Review* 71: 242-247

[41] Scott Morton, Fiona M., 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: 65-92

[42] Shepard, Andrea 1991 "Price discrimination and retail configuration" *Journal of Political Economy* 99(1): 30-53

[43] Verboven, Frank. 1996 "International price discrimination in the European car market" *RAND Journal of Economics* 27(2): 240-268

[44] Villas-Boas, Sofia Berto "An empirical investigation of the welfare effects of banning wholesale price discrimination" *RAND Journal of Economics* 40(1): 20-46

[45] Train, Kenneth 2003 *Discrete Choice Methods with Simulation* (Cambrigde, MA: Cambridge University Press)

[46] Train, Kenneth and Clifford Winston. 2007. "Vehicle choice behavior and the declining market share of U.S. automakers." *International Economic Review*, 48: 1469-1496.

[47] Yinger, John 1998 "Evidence on discrimination in consumer markets" *Journal of Economic Perspectives* 12(2): 23-40

Table 1: Correlations Between Price Paid and Demographic Group

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Male	32.01					-35.04	48.32	36.52	23.76	19.27
	(116.95)					(117.41)	(118.87)	(117.70)	(116.72)	(118.03)
Household Income		106.30***				106.67***				141.93***
(tens of thousands of $)		(17.93)				(18.10)				(20.33)
Married			-103.94				-111.56			-462.97***
			(133.73)				(135.92)			(147.00)
Years of Education				-37.71				-37.95		-114.99***
				(33.29)				(33.42)		(34.39)
Urban					219.14				218.19	129.40
					(147.37)				(147.16)	(149.31)
Model fixed effects	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
F						17.64	0.37	0.65	1.11	11.06
Joint Significance						0.0000	0.6900	0.5210	0.3292	0.0000
R^2	0.8047	0.8061	0.8047	0.8047	0.8048	0.8062	0.8047	0.8047	0.8048	0.8070

Regression of price paid on consumer characteristics and vehicle fixed effects. Robust standard errors in parentheses. Coefficients marked with a *, **, or *** are significant at the 10%, 5%, or 1% level respectively. All regressions include 11029 observations and 210 vehicle fixed effects. The number of observations in these regressions is smaller than the number of observations in the discrete choice model because of the exclusion of consumers who purchase the outside good from these regressions.

Table 2: Correlations Between Price Paid, Demographic Group, and Vehicle Type

Dependent Variable: Price Paid Independent Variable Interacted with:	Male	Single	High Income
Compact Car	-338.73**	185.97	660.05**
	(171.93)	(165.68)	(275.99)
Fullsize Car	220.42	-236.07	-1037.61*
	(455.44)	(595.94)	(568.21)
Luxury Car	1424.65***	-17.50	953.94*
	(454.81)	(506.44)	(512.36)
Midsize Car	200.40	-6.07	-186.22
	(276.52)	(270.11)	(365.45)
Pickup	-1272.62	17.70	297.16
	(796.95)	(494.26)	(586.86)
Sport Utility Vehicle	682.69***	-377.01	239.92
	(258.21)	(277.58)	(354.61)
Sporty Car	793.21*	192.54	-492.06
	(423.37)	(398.34)	(510.75)
Van	487.34	-368.23	459.88
	(509.83)	(552.82)	(615.53)
F-test rejection level	0.0058	0.8725	0.0660
R^2	0.8055	0.8041	0.8064

Regression of price paid on consumer characteristics and vehicle fixed effects. Robust standard errors in parentheses. Coefficients marked with *, **, and *** are significant at the 10%, 5%, and 1% level respectively. All regressions include 11,029 observations and 210 vehicle fixed effects. The number of observations in these regressions is smaller than the number of observations in the discrete choice model because of the exclusion of consumers who purchase the outside good from these regressions.

Table 3: Mean Preferences by Gender and Marital Status

	Demographic Group			
	Married		Single	
	Women	Men	Women	Men
Top Five Vehicles:	Toyota Highlander	Jeep Grand Cherokee	Honda CR-V	Dodge Caravan
(in order, highest to lowest)	Dodge Caravan	Chevrolet Silverado LD	Toyota Rav4	Chevrolet Express
	Jeep Liberty	Honda CR-V	Jeep Liberty	Jeep Grand Cherokee
	Honda CR-V	Toyota Highlander	Ford Escape	Jeep Wrangler
	Toyota Sienna	Jeep Liberty	Jeep Grand Cherokee	Jeep Liberty
Bottom Five Vehicles:	Chevrolet SSR	Hummer H2 SUT	Buick Park Avenue	Buick Park Avenue
(in order, lowest to highest)	Buick Park Avenue	Audi TT	Lexus LS 430	Lexus SC 430
	Hummer H2 SUT	Lexus SC 430	Lexus LX 470	Buick Century
	Buick Century	Jaguar S-Type	Chevrolet Corvette	Toyota Echo
	Volkswagen GTI	Toyota Echo	Lexus SC 430	Saab 9-5

Vehicles are ordered by the estimated mean preference of each demographic group, controling for heterogeneity in group preference. The mean preference is estimated using the Berry (1994) inversion in the maximum likelihood estimation of the coefficients that describe within-demographic group heterogeneity in preferences.

Table 4: Mean Preference Coefficients by Gender and Marital Stauts

	Gender and Marital Status			
Variable	Married Females	Married Males	Single Females	Single Males
Price	-1.36***	-1.42***	-1.88***	-1.64***
(tens of thousands of dollars)	(0.25)	(0.31)	(0.32)	(0.29)
Car	-1.76***	-1.57***	-1.82***	-1.67**
	(0.31)	(0.36)	(0.36)	(0.35)
Pickup	-2.62***	-1.20***	-0.56	-0.28
	(0.38)	(0.45)	(0.41)	(0.41)
Sporty Car	-1.04**	-1.54***	-0.87*	-0.52
	(0.41)	(0.48)	(0.46)	(0.47)
Curbweight	1.34***	1.21***	0.97**	1.32***
(thousands of pounds)	(0.33)	(0.41)	(0.44)	(0.38)
Number of Passengers	-0.23**	-0.57***	-0.28**	-0.49***
	(0.10)	(0.12)	(0.12)	(0.11)
Turning Radius	-4.60***	-3.83***	-3.76***	-4.38***
(feet)	(0.41)	(0.47)	(0.46)	(0.46)
Domestic	-1.28***	-0.59**	-1.93***	-0.67***
	(0.21)	(0.24)	(0.24)	(0.24)
Number of Observations	213	213	213	213

Instrumental variables regression of the mean preference of each group for each vehicle on vehicle characteristics. Instruments are functions of the vehicle attributes of competing vehicles, as in BLP. Weighted instrumental variables standard errors in parentheses, where the weights are equal to the number of observations for that demographic group that purchased that vehicle in the maximum likelihood stage. Significance level indicated by: *=10%, **=5%, ***=1%.

Table 5: Preference Heterogeneity Coefficients by Gender and Marital Status

		Gender and Marital Status			
Variable	Variable Type	Married Females	Married Males	Single Females	Single Males
Price	Std Dev	0.07**	0.06**	0.004	0.009
(tens of thousands of dollars)		(0.03)	(0.02)	(0.045)	(0.038)
	Divided by	-7.15***	-5.86***	-5.90***	-4.62***
	Income	(1.00)	(0.53)	(1.19)	(0.94)
SUV	Std Dev	1.47***	2.42***	0.03	1.76***
		(0.32)	(0.27)	(0.19)	(0.52)
Pickup	Std Dev	2.55***	3.78***	0.16	2.31***
		(0.56)	(0.37)	(0.36)	(0.65)
Van	Std Dev	1.83***	2.69***	1.59	0.88
		(0.48)	(0.40)	(1.06)	(1.72)
Car	Std Dev	2.90***	4.08***	2.80***	3.26***
		(0.41)	(0.35)	(0.61)	(0.75)
Horsepower	Std Dev	0.07	0.59***	0.37	0.39
(hundreds)		(0.20)	(0.11)	(0.31)	(0.24)
Fuel Use	Std Dev	0.10	0.26***	0.11	0.15**
(gallons per hundred miles)		(0.09)	(0.06)	(0.16)	(0.07)
Curb Weight	Std Dev	0.004	0.03	0.34	0.005
(thousands of pounds)		(0.091)	(0.09)	(0.21)	(0.031)
Outside Good	Std Dev	1.41***	0.46**	1.74***	0.82*
		(0.28)	(0.24)	(0.46)	(0.46)
	Divided by	-12.37***	-6.42***	-12.02***	-8.43***
	Income	(2.69)	(1.55)	(3.09)	(2.68)
Number of Halton Draws		100	80	100	100
Number of Observations		3092	5606	2400	2356

Standard errors in parentheses. Significance level indicated by: *=10%, **=5%, ***=1%. Coefficients estimated with maximum likelihood.

Table 6: Estimated Elasticities and Markups

		Gender and Marital Status			
Variable	Statistic	Married Females	Married Males	Single Females	Single Males
Elasticity	Mean	-10.10	-6.39	-17.82	-17.32
$ Markup	Mean	10,078	9,107	7,349	7,605
	Min	5,560	6,227	4,484	5,333
	Max	14,752	14,311	12,457	10,807
% Markup	Mean	35.25	32.78	25.58	27.17
	Min	21.19	15.88	16.64	14.58
	Max	57.75	60.73	45.61	49.79
Number of Vehicles		213	213	213	213

Descriptive statistics are over the 211 vehicles in the sample. All numbers are calculated using the demand coefficients presented in tables 4 and 5. Percent Markup is the markup divided by the average price paid for the vehicle by that demographic group.

Table 7: Correlations Between Price Paid and Demographic Group Controlling for Markups

	Dependent Variable	
	Price-markup	Price
Single Male	-411.58**	-209.09
	(161.01)	(170.94)
Married Female	-2955.33***	-1281.50***
	(132.31)	(332.97)
Married Male	-2135.81***	-903.92***
	(203.72)	(331.06)
Markup		0.3941***
		(0.1241)
Model fixed effects	Y	Y
F	171.52	5.64
Joint Significance	0.0000	0.0010
R^2	0.7548	0.8080

Regression of price paid minus optimal markup on consumer characteristics and vehicle fixed effects. Standard errors clustered by vehicle in parentheses. Coefficients marked with a *, **, or *** are significant at the 10%, 5%, or 1% level respectively. All regressions include 11115 observations and 213 vehicle fixed effects. The number of observations in these regressions is smaller than the number of observations in the discrete choice model because of the exclusion of consumers who purchase the outside good from these regressions.

Appendix

Ashley Langer
University of Michigan

October 2011

A Results with Income Group and Gender

A major assumption of the empirical model is that dealers are unable to price discriminate based on consumers demographics other than gender and marital status. This appendix explores the implication of that assumption by re-running the exact same empirical specification used in the text, but with demographic groups defined by gender and broad income category. In this specification, dealers are allowed to price discriminate based on whether the consumer comes from a household with income above or below $100k, but not at a finer level of household income. While the results are broadly consistent with those in the paper, a few additional interesting results do appear.

A.1 Mean Preferences

As in the text, the δ vector estimates each demographic group's mean preference for each vehicle. Again, there are examples of vehicles that certain gender and income groups find particularly distastefull on average and yet some people particularly like. A good example of such a scenario with gender and income groups occurs with muscle cars like the Chevrolet Corvette or the Pontiac GTO. While the average consumer of any demographic group would find such vehicles unattractive, there may be some consumers in each group with a high demand for vehicle performance and a low preference for fuel economy who find these cars attractive. The highest and lowest five mean preference vehicles for each demographic group are listed in Table A.1. Across all demographic groups, small and mid-sized SUVs such as the Jeep Liberty and the Honda CR-V frequently appear among the top 5 vehicles in terms of mean preference. Similarly, the Hummer H2 SUT and the Cadillac Escalade EXT[1] frequently appear among the bottom 5 vehicles, likely because of their unusual features and their extremely high price. It is additionally interesting to note some of the differences between different demographic groups that might conform to our priors. Low income women have strong average preferences for inported sedans such as the Nissan Altima and the Toyota Corolla that do not rank highly for other demographic groups. Additionally. the lower income groups, both men and women, have luxury vehicles among their bottom five while the high income groups do not.

Table A.2 reports the mean preference coefficients by income group and gender. The variables that are included are exactly the same as in the gender and marital status results in the text. The first thing to notice is that the low income demographic groups are substantially more price sensitive on average than the high income groups, as might be expected. Cars and pickup trucks are generally less attractive to consumers than SUVs and vans (the excluded categories along with the outside good), with the notable exception of low income womens' preference for cars. Sporty cars (which are generally quite small but with high horsepower, like the Acura RSX and the Volkswagen GTI) are even more unattractive to consumers than cars in general. Curb weight is particularly valued by low income consumers, while an increase in the the number of passengers a vehicle can carry, conditional on all other vehicle attributes, decreases consumers' value for the vehicle for all groups other than

[1] SUT stands for sport utility truck, which is effectively a large SUV with a pickup bed.

high income women. A high turning radius reduces consumer utility for all demographic groups, and domestic vehicles are less attrative to consumers than imports, although they are particularly unattractive to women of both income groups.

A.2 Preference Heterogeneity

Table A.3 presents the coefficient estimates for these consumer heterogeneity terms by income group and gender. Again, the specification is exactly the same as in the text. Although the general results are similar to those in the text, there are a few interesting differences. First, both high income men's and women's price sensitivity is much more respondent to differences in household income than low income groups (or than married or single men or women in the text). This could be because the high income group only includes to household income sub-categories: \$100k-\$150k and more than \$150k. This highest income group could have incomes that are many times as high as the lower income group, thus leading to greater differences in price sensitivity. Another interesting difference from the gender and marital status groups is that although single women did not exhibit a lot of preference heterogeneity in those results, low income women do exhibit substantial variation in their preferences for different types of vehicles here. If low income women include both single and married women, some of whom may have children, this could explain why there is greater variation in their preferences. Finally, the result that women are more heterogeneous in their preference for the outside good is repeated in these results.

A.3 Elasticities and Markups

Table A.4 gives the same descriptive statistics on own-price elasticities and markups across demographic groups as were shown in the text. Generally, the elasticities average from approximately 9 (in absolute value) for high income men and education groups to over 16 for low income women. Again, these elasticities are fairly large, but the markups are in line with conventional thinking. The average predicted markup across all demographic groups is 33.3% of the transaction price (again, as compared to the 31.5% used by the USEPA), again in keeping with the results in the text. Although some the markups for high income consumers are clearly quite high (at over 100% of the average transaction price), it is reasonable to expect that the wealthiest consumers will be willing to pay substantially more than the average consumer, and that these optimal markup differences may be mitigated by dealers' ability to negotiate sufficiently high prices for these high income consumers. The optimal markups for low income groups seem reasonable or slightly low, averaging between 7% and 43% of transaction prices.

A.4 Variation in Prices

Again turning to the predicted price in excess of optimal markups, I find that high income consumers appear to pay substantially lower prices for new vehicles, controlling for optimal markups, than low income consumers (as shown in Table A.5). This is in direct contrast to the simple correlations shown in the paper (Table 1) that indicate that the price paid for a new vehicle increases in household income. Those simple correlations include both the increased willingness-to-pay of high income consumers and the different approach to negotiation of high income consumers relative to low income consumers. High income households are better educated and perhaps better informed negotiators, which might lead them to pay less for a vehicle that they value equally to others. However, these households' high value of time might lead them to invest less time in the search and negotiation process, thus increasing the price they pay. These results seem to suggest that the increase in negotiating ability outweighs the decreased patience of high income consumers once preferences are controlled for. Of course, these results do not rule out that dealers are discriminating against low income consumers, or discriminating against a demographic characteristic that is associated with lower household incomes such as youth.

Another interesting similarity to the text is that while low income men and women pay similar prices for new vehicles (with men paying statistically insignificantly less), high income men pay *more* than high income women for the same vehicles, controlling for preferences. This again seems likely to be a selection story, where women from high income households who choose to go purchase a new vehicle for themselves are stronger negotiators than high income men. The magnitude of these results suggests that this selection may be even larger for high income women relative to high income men than it was for married women relative to married men.

A.5 Comparison to Main Results

Although the results presented in this appendix are generally in line with the main results in the text, they do highlight the importance of choosing the demographic groups. Since data constraints mean that dealers are likely able to price discriminate on more demographics than the econometrician is able to use in estimation, the definition of the demographic groups directly impacts the interpretation of the variation in prices controlling for markups. While the lower prices in excess of markups paid by married people relative to single people might particularly indicate consumers ability to send a household member and thus obscure their demographic group, the lower prices paid by higher income consumers seem to indicate that consumers from higher income households may be better negotiatiors (via increased education or general negotiating skills) than consumers from lower income households. In the gender and marital status results presented in the main text, price discrimination based on all other demographics that vary with gender and marital status (perhaps most importantly age and household income) must be thought of as entering into the estimates of the differences in the prices paid by different demographic groups. Similarly here, if high income consumers are more likely to be married, then the difference in prices paid by income must also be thought of an including differences in marital status. Ideally, a massive dataset would allow for estimation on enough dimensions to tell a complete story about consumer demand at the level of demographics that the dealer is able to observe, but I am unaware of such a dataset and, if it exists, it is unlikely to include additional features that the dealer uses to predict the optimal markup, such as the clothes that the consumer is wearing on the day she goes to purchase a new vehicle. By showing results for two different demographic group breakdowns, the goal here is to show that the general approach to thinking about the problem of price discrimination is robust, although the researcher must carefully weigh both the choice of the demographic groups to consider and the interpretation of the results.

Table A.1: Mean Preferences by Income Group and Gender

	Demographic Group	
	High Income	
	Women	Men
Top Five Vehicles:	Toyota Highlander	Jeep Wrangler
(in order, highest to lowest)	Ford Escape Hybrid	Ford Escape Hybrid
	Honda CR-V	Jeep Grand Cherokee
	Jeep Liberty	Saturn VUE
	Lexus RX 330	Honda CR-V
Bottom Five Vehicles:	Chevrolet Venture	Cadillac Escalade EXT
(in order, lowest to highest)	Pontiac GTO	Hummer H2 SUT
	Mercury Monterey	Pontiac GTO
	GMC Safari	Subaru Baja
	Dodge Sprinter	Mercury Monterey
	Low Income	
	Women	Men
Top Five Vehicles:	Nissan Altima	Dodge Caravan
(in order, highest to lowest)	Toyota Corolla	Chrysler Town & Country
	Toytoa Camry Sedan	Jeep Liberty
	Jeep Liberty	Honda CR-V
	Honda CRV	Jeeep Wrangler
Bottom Five Vehicles:	Cadillac Escalade EXT	Hummer H2 SUT
(in order, lowest to highest)	Chevrolet Corvette	Cadillac Escalade EXT
	Porsche Cayenne	Jaguar S-Type
	Hummer H2 SUT	Chevrolet Corvette
	Mercedes SLK-Class	Lexus LS 430

Vehicles are ordered by the estimated mean preference of each demographic group, controling for heterogeneity in group preference. The mean preference is estimated using the Berry (1994) inversion in the maximum likelihood estimation of the coefficients that describe within-demographic group heterogeneity in preferences.

Table A.2: Mean Preference Coefficients by Income Group and Gender

	Income Group and Gender			
Variable	High Females	High Males	Low Females	Low Males
Price	-0.75**	-0.97***	-2.42***	-2.35***
(tens of thousands of dollars)	(0.38)	(0.38)	(0.30)	(0.30)
Car	-1.22***	-1.64***	0.60*	-2.60**
	(0.42)	(0.41)	(0.35)	(0.33)
Pickup	-0.67	-4.26***	-1.09***	-2.22***
	(0.46)	(0.41)	(0.38)	(0.35)
Sporty Car	-1.50***	-1.59***	-1.09**	-0.99**
	(0.56)	(0.52)	(0.46)	(0.45)
Curbweight	0.18	0.63	2.09***	1.44**
(thousands of pounds)	(0.50)	(0.48)	(0.40)	(0.36)
Number of Passengers	-0.19	-0.58***	-0.38***	-0.50***
	(0.12)	(0.11)	(0.10)	(0.09)
Turning Radius	-4.48***	-3.64***	-4.98***	-3.16***
(feet)	(0.54)	(0.51)	(0.44)	(0.43)
Domestic	-1.65***	-0.45*	-1.44***	-0.39*
	(0.29)	(0.27)	(0.24)	(0.23)
Number of Observations	211	211	211	211

Instrumental variables regression of the mean preference of each group for each vehicle on vehicle characteristics. Instruments are functions of the vehicle attributes of competing vehicles, as in BLP. Weighted instrumental variables standard errors in parentheses, where the weights are equal to the number of observations for that demographic group that purchased that vehicle in the maximum likelihood stage. Significance level indicated by: *=10%, **=5%, ***=1%.

Table A.3: Preference Heterogeneity Coefficients by Income Group and Gender

		Income Group and Gender			
Variable	Variable Type	High Females	High Males	Low Females	Low Males
Price	Std Dev	0.05	0.07***	0.01	0.002
(10,000s)		(0.04)	(0.03)	(0.03)	(0.032)
	Divided by	-47.55***	-26.53***	-5.42***	-4.57***
	Income	(9.61)	(5.65)	(1.00)	(0.78)
SUV	Std Dev	1.83***	0.12	1.66***	0.15
		(0.50)	(0.12)	(0.37)	(0.11)
Pickup	Std Dev	1.96***	5.70***	1.85***	3.28***
		(0.80)	(0.65)	(0.60)	(0.44)
Van	Std Dev	4.70***	2.97***	2.45***	0.04
		(1.30)	(0.55)	(0.54)	(0.14)
Car	Std Dev	3.44***	4.09***	1.96***	3.51***
		(0.63)	(0.42)	(0.39)	(0.40)
Horsepower	Std Dev	1.19***	1.61***	0.18	0.86***
(100s)		(0.27)	(0.20)	(0.16)	(0.16)
Fuel Use	Std Dev	0.29***	0.01	0.05	0.18***
(gal. per 100 miles)		(0.11)	(0.06)	(0.08)	(0.06)
Curb Weight	Std Dev	0.01	0.66***	0.10	0.17**
(1000s of pounds)		(0.10)	(0.14)	(0.09)	(0.07)
Outside Good	Std Dev	1.39***	0.31	1.45***	0.92***
		(0.34)	(0.23)	(0.31)	(0.25)
	Divided by	-168.63***	-2.80	-8.31***	-2.71
	Income	(36.69)	(24.43)	(2.56)	(2.12)
# of Halton Draws		100	100	100	100
# of Obs					

Standard errors in parentheses. Significance level indicated by: *=10%, **=5%, ***=1%. Coefficients estimated with maximum likelihood.

Table A.4: Estimated Elasticities and Markups

		Income Group and Gender			
Variable	Statistic	High Females	High Males	Low Females	Low Males
Elasticity	Mean	-13.05	-8.98	-16.42	-13.28
$ Markup	Mean	15,393	11,886	5,156	5,282
	Min	10,748	9,972	3,945	4,224
	Max	37,322	16,792	6,885	7,427
% Markup	Mean	53.26	42.48	18.39	19.07
	Min	31.47	17.79	7.83	9.20
	Max	146.55	111.90	36.89	42.65
Number of Vehicles		211	211	211	211

Descriptive statistics are over the 211 vehicles in the sample. All numbers are calculated using the demand coefficients presented in tables A.2 and A.3. Percent Markup is the markup divided by the average price paid for the vehicle by that demographic group.

Table A.5: Correlations Between Price Paid minus Markup and Demographic Group

	(1)
Low Income Male	-199.48
	(133.87)
High Income Female	-9796.54***
	(356.39)
High Income Male	-6068.56***
	(304.87)
Model fixed effects	Y
F(3,210)	456.05
Joint Significance	0.0000
R^2	0.7477

Regression of price paid minus optimal markup on consumer characteristics and vehicle fixed effects. Standard errors clustered by vehicle in parentheses. Coefficients marked with a *, **, or *** are significant at the 10%, 5%, or 1% level respectively. All regressions include 11029 observations and 211 vehicle fixed effects. The number of observations in these regressions is smaller than the number of observations in the discrete choice model because of the exclusion of consumers who purchase the outside good from these regressions.