UNIVERSITY OF COPENHAGEN
FACULTY OF SCIENCE

Bsc in Computer Science and Economics

Analyzing Market Dynamics

Estimating Demand and Supply in the US Automobile Industry using the BLP method

Thomas Busk-Jepsen & Magnus Goltermann

Supervised by Bertel Schjerning

June 10, 2024

Thomas Busk-Jepsen & Magnus Goltermann
Analyzing Market Dynamics
Bsc in Computer Science and Economics, June 10, 2024
Supervisors: Bertel Schjerning

University of Copenhagen
Faculty of Science
Bachelor degree in Computer Science and Economics
Universitetsparken 1
2100 Copenhagen Ø

Abstract

We analyzed market dynamics in the U.S. car market, using sale data from 1980 to 2018, which includes prices, attributes, and units sold. To estimate demand, we employed the random coefficient logit model. On the supply side, we used the Bertrand-Nash assumption to derive marginal costs. Our analysis indicates that competition increased during the period and there was a decrease in markups. Furthermore, our results suggest that mergers before 2000 would have significantly reduced consumer surplus. However, mergers after 2010 might not adversely affect consumer surplus as much and could potentially improve it under certain conditions due to cost reductions.

Contents

1 Introduction

The aim of this paper is to analyze the U.S. car market by estimating both the demand and supply sides using the techniques first introduced by Berry, Levinsohn, and Pakes (1993) often referred to as the BLP estimation technique. We will focus on implementing these techniques ourselves. While we are not developing a Python library for future use, the model we create should be straightforward to apply in other oligopolistic markets.

We will be using data from Grieco *et al.* (2023) that has data on the cars sold from 1980-2018 in the US, giving us a total of 572.948.272 car purchases to analyze. This dataset has extensive data on each model sold, for example, but not limited to price, horsepower, miles per gallon, weight, and height.

By analyzing this dataset, we can infer how various parameters influence consumer utility. Although we cannot provide specifics, such as the exact increase in market share resulting from an increase in horsepower due to the complexity of the problem, we can estimate whether the parameters positively or negatively affect utility. Our results align well with intuition; for instance, we find that horsepower generally increases utility, while price decreases it. The only outlier in our estimation is miles per gallon, which is found to be statistically insignificant at the 95% confidence interval.

By utilizing the data to analyze the supply side, we can estimate the different producers marginal cost and their markups.

Utilizing our estimated demand and supply side we conduct counterfactuals. We will examine both hypothetical mergers and company breakups to analyze their impacts on market power, markups, and consumer surplus. Our findings indicate that from 1980 to 2000, stringent legislation and antitrust laws were crucial, as mergers during this period would have significantly reduced consumer surplus. For instance, we find that a merger of the top five largest companies in 1980 would have decreased consumer surplus by over 20%. Conversely, after 2010, more relaxed regulations could be advisable, as the negative impact of mergers on consumer surplus has diminished. In this later period, we find that the same merger would lead to a decrease in consumer surplus of under 5%.

Related literature

As mentioned earlier, this paper largely follows the method laid out in Berry, Levinsohn, and Pakes (1993) (referred to as the BLP method). This includes utilizing the same estimation techniques and approaches.

The BLP method has been applied in a wide array of contexts and fields. For instance, it has been used to analyze the airline industry and the entry costs associated with it (Berry *et al.*, 1996), to study market power and the effects of mergers on pricing in the ready-to-eat cereal market (Nevo, 2001) and (Nevo, 2000b), and to examine the introduction of new goods and their impact on the car market (Petrin, 2002).

While we do not use microdata in our analysis, it is important to note that incorporating microdata could significantly improve the estimates (Conlon and Gortmaker, 2023). Microdata allows for a more detailed understanding of individual consumer behavior and preferences, which enhances the accuracy and robustness of the demand estimation. An important example of this is provided by Berry *et al.* (2004) who included second choices in their model to capture substitution patterns more accurately.

Several packages for programming the BLP framework have been developed. Notably, the release of `PyBLP` in 2018, along with the accompanying article by Conlon and Gortmaker (2020b), outlines best practices for using the `PyBLP` library. While we do not exclusively use this library in our code, we rely on it for computationally intensive estimations due to its optimization for running large simulations. Our decision to program the BLP framework ourselves is based on the need to ensure transparency and control over the implementation, as these factors can be compromised when using libraries without careful consideration.

Theory

2

Demand & Supply

In this section, we will discuss the demand and supply framework for our estimations. We will explore three distinct logit models, each varying in flexibility and realism: The homogeneous logit, the nested logit, and the random coefficient logit. On the supply side, we assume oligopolistic competition with constant marginal cost.

Random Utility model

The consumer faces a discrete choice, and we assume that the consumer will behave in a utility-maximizing way when choosing an alternative. The exact utility is known by the individual but for us, some of the utility is unobserved. We can decompose the observed and the unobserved utility as follows $U_{ijt} = \delta_{ijt} + \epsilon_{ijt}$. Where U_{ijt} is the total utility consumer i gets from choosing car j at time t. δ_{ijt} represents the observable part of the utility and ϵ_{ijt} represents the error term that captures unobserved individual factors affecting utility. We assume a linear form of δ_{ijt} and can deconstruct the observed part of the utility term as follows:

$$\delta_{ijt} = \alpha p_{jt} + \boldsymbol{\beta}_i \boldsymbol{x}_{jt} + \xi_{jt} \tag{2.1}$$

Here α represents the price coefficient, p_{jt} is the price of car j at time t, $\boldsymbol{x}_{jt}$ is a vector of characteristics of car j, $\boldsymbol{\beta}_i$ is the individual specific parameters linked to the characteristics and ξ_{jt} is the unobserved product characteristics.
This formulation allows the product coefficients to vary at the individual level, although the price coefficient α remains constant across all consumers.

As we want to use the logit model, we assume that each ϵ_{ijt} is independently, identically distributed extreme value type 1 $\epsilon_{ijt} \sim$ IID Extreme Value type 1[1]. The derivation of the logit model is detailed in Train (2009), leading to the following function that describes

[1] Also known as the Gumbel distribution.

the probability of choosing car j conditioned on the preferences of consumer i at time t:

$$Pr(j|i,t) = \frac{\exp(\delta_{ijt})}{\Sigma_{k\in J_t}\exp(\delta_{ikt})} \tag{2.2}$$

Where J_t is all the cars in period t.

2.1 Demand

Modeling homogeneous consumers

Assuming homogeneous consumers, with only the error term differing among them, we can now express the utility function as $U_{ijt} = \delta_{jt} + \epsilon_{ijt}$, this simplifies equation (2.2) and (2.1) to the following:

$$Pr(j|t) = \frac{\exp(\delta_{jt})}{\Sigma_{k\in J_t}\exp(\delta_{kt})} \tag{2.3}$$

$$\delta_{jt} = \alpha p_{jt} + \boldsymbol{\beta}\boldsymbol{x}_{jt} + \xi_{jt} \tag{2.4}$$

To analyze how prices and product features influence consumer utility, and thereby estimating market shares, we need to determine the parameters α and $\boldsymbol{\beta}$ in equation (2.4).

Observations of market shares allow us to employ the inversion method initially introduced by Berry *et al.* (1993). This method entails analyzing the differences in the logarithm of market shares between each alternative and the outside good to infer the underlying utilities. For identification purposes, we set the utility of the outside good to zero, as only differences in utility matters. This implies $\delta_{0t} = 0 \forall t$, allowing us to solve the following equation:

$$\begin{aligned}
\log(Pr(j)) - \log(Pr(0)) &= \log\left(\frac{\exp(\delta_{jt})}{\sum_{k\in J_t}\exp(\delta_{kt})}\right) - \log\left(\frac{\exp(\delta_{0t})}{\sum_{k\in J_t}\exp(\delta_{kt})}\right) \\
&= \delta_{jt} - \log\left(\sum_{k\in J_t}\exp(\delta_{kt})\right) - \delta_{0t} + \log\left(\sum_{k\in J_t}\exp(\delta_{kt})\right) \\
&= \delta_{jt} - \underbrace{\delta_{0t}}_{=0} = \delta_{jt}
\end{aligned} \tag{2.5}$$

Since the observed market shares represent actual consumer choices, and given a sufficiently large sample, the law of large numbers ensures that the observed market shares converge to the true probabilities of choosing each alternative. Therefore, we can equate the observed market shares with the probabilities of choosing each car. Using (2.4) and

(2.5) we can formulate:

$$\log(\mathcal{S}_{jt}) - \log(\mathcal{S}_{0t}) = \delta_{jt} = \alpha p_{jt} + \boldsymbol{\beta x}_{jt} + \xi_{jt} \tag{2.6}$$

Where $\mathcal{S}_{jt}$ is the observed market share of car j at time t. We can estimate α and $\boldsymbol{\beta}$ in (2.6) with a linear regression model.

When calculating the elasticity of demand we can follow the same approach laid out in Rasmusen (2007) giving us the own price and cross-price elasticity as:

$$\mu_{jkt} = \frac{\partial \mathcal{S}_{jt}}{\partial p_{kt}} \cdot \frac{p_{kt}}{\mathcal{S}_{jt}} = \begin{cases} -\alpha p_{jt}(1 - \mathcal{S}_{jt}) & \text{if } j = k \\ \alpha p_{kt} \mathcal{S}_{kt} & \text{if } j \neq k \end{cases} \tag{2.7}$$

This equation gives us that if a car's price is increased then it will lose its market share equally to other competing cars. The cross-price elasticity of car j with respect to the price of cars k is $\alpha p_{kt} s_{kt}$, This means that the effect of a price increase in car k on the market shares of other cars is proportional to their existing market shares. Consequently, the cross-price elasticity appears the same for all cars relative to k, illustrating that the market share changes uniformly across competitors. We can further illustrate this point by looking at the substitution pattern of two cars:

$$\frac{Pr(j|t)}{Pr(l|t)} = \frac{\frac{\exp(\delta_{jt})}{\Sigma_{k \in J_t} \exp(\delta_{kt})}}{\frac{\exp(\delta_{lt})}{\Sigma_{k \in J_t} \exp(\delta_{kt})}} = \frac{\exp(\delta_{jt})}{\exp(\delta_{lt})} \tag{2.8}$$

The ratio only depends on attributes of alternative j and l, nothing about the rest of the choice set. This relative probability is therefore independent of irrelevant alternatives (IIA). This is a know issue with the homogeneous logit model, which we will address as we transition to more advanced models.

Modelling more complex substitution patterns with the Nested logit model

Due to the limitations of the homogeneous logit model regarding the IIA, we will now introduce the nested logit model. This model allow us to relax the IIA assumption as we group choices into nests. Within each group, the alternatives are allowed to be correlated, meaning that $Cov(\epsilon_{ij}, \epsilon_{ik}) \geq 0$ if j and k are in the same nest but still $Cov(\epsilon_{ij}, \epsilon_{ik}) = 0$ if j and k are in different nests.

While the IIA assumption still applies within each nest, the nested logit model introduces the concept of Independence of Irrelevant Nests (IIN), which allows for more complex

substitution patterns between cars in different nests.
We can split the probability of choosing a car into the probability of choosing the nest and the conditional probability of choosing the car given the specific nest is chosen. This gives the following equation:

$$Pr(j) = Pr(nest\ h_j) \cdot Pr(j|nest\ h_j) \tag{2.9}$$

Here h_j represents the nest that car j is in. We group our cars in the following nests: Sports car, van, truck, SUV, and other. This structure can be visualized in figure (2.1).

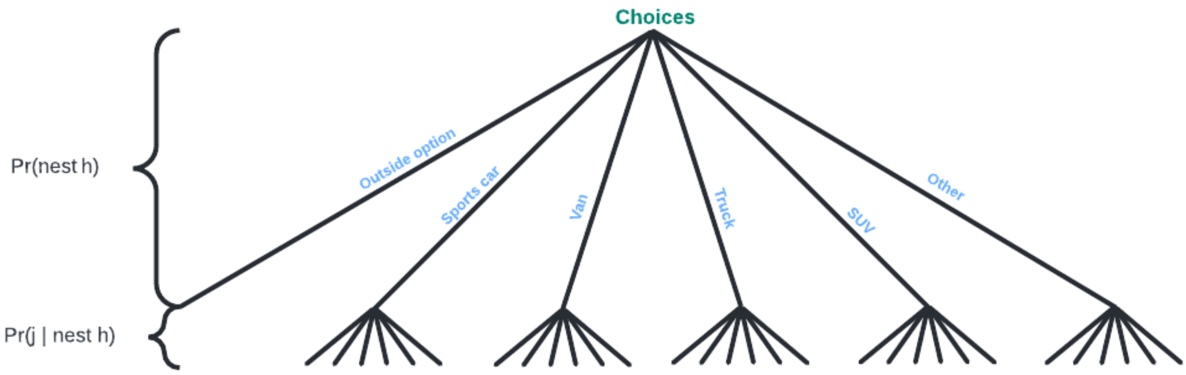

Figure 2.1: Tree diagram for illustrating nesting structure

We can now deconstruct the unconditional and conditional probability in equation (2.9) by following Østli *et al.* (2017) where the probability of the unconditional choice is given by the equation:

$$Pr(\text{nest } h_j) = \frac{\exp(I_{h_j t})}{\Sigma_{h\in H}\exp(I_{ht})} \tag{2.10}$$

Where the inclusive value is defined as $I_{ht} \equiv (1-\rho)\log \Sigma_{j\in J_{ht}}\exp\left(\frac{1}{1-\rho}\delta_{jt}\right)$, and ρ represents the nesting parameter, indicating the degree of correlation among unobserved factors within each nest. When $\rho = 0$, the Nested Logit model collapses to the homogeneous logit model. Conversely, when $\rho = 1$ all alternatives within a nest are perfectly correlated in terms of their unobserved utility components.
The probability of the conditional choice is then:

$$Pr(j|\text{nest } h_j) = \frac{\exp(\frac{1}{1-\rho}\delta_{jt})}{\Sigma_{k\in J_{ht}}\exp(\frac{1}{1-\rho}\delta_{kt})} \tag{2.11}$$

Putting it together we get:

$$Pr(j) = \left(\frac{\exp(I_{h_j t})}{\Sigma_{h\in H}\exp(I_{ht})}\right) \cdot \left(\frac{\exp(\frac{1}{1-\rho}\delta_{jt})}{\Sigma_{k\in J_{ht}}\exp(\frac{1}{1-\rho}\delta_{kt})}\right) \tag{2.12}$$

We can then calculate utility from the inversion step following Berry (1994):

$$\delta_{jt} = \log(S_{jt}) - \log(S_{0t}) - \rho \log(S_{j|ht}) \tag{2.13}$$

To get the elasticity of the nested logit model, we follow Ackerberg and Crawford (2009), and the nested models elasticities can be expressed as:

$$\mu_{jkt} = \begin{cases} -\alpha \mathcal{S}_{jt} \left(\frac{1}{1-\rho} - \frac{\rho}{1-\rho} \mathcal{S}_{jt|h} - \mathcal{S}_{jt} \right) & \text{if } j = k \\ \alpha \mathcal{S}_{kt} \left(\frac{\rho}{1-\rho} \mathcal{S}_{jt|h} + \mathcal{S}_{jt} \right) & \text{if } j \text{ and } k \text{ in same nest} \\ \alpha \mathcal{S}_{jt} \mathcal{S}_{kt} & \text{otherwise} \end{cases} \tag{2.14}$$

$\mathcal{S}_{jt|h}$ is the fractions of observed shares that car j has in nest h. From equation (2.14) we can see that we have two cases concerning the substitution pattern, one where j and k are in the same nest and one where they are not.

If j and k are in the same nest we get:

$$\frac{Pr(j)}{Pr(k)} = \frac{Pr(\text{nest } h_j) \cdot Pr(j|\text{nest } h_j)}{Pr(\text{nest } h_k) \cdot Pr(k|\text{nest } h_k))} = \frac{Pr(j|\text{nest } h_j)}{Pr(k|\text{nest } h_k)} = \frac{Pr(j)}{Pr(k)} \tag{2.15}$$

If j and k are in different nests we get:

$$\frac{Pr(j)}{Pr(k)} = \frac{Pr(\text{nest } h_j) \cdot Pr(j|\text{nest } h_j)}{Pr(\text{nest } h_k) \cdot Pr(k|\text{nest } h_k)} \tag{2.16}$$

From equation (2.15), it is evident that the IIA (Independence of Irrelevant Alternatives) assumption holds within nests, similar to what was observed in equation (2.8) in the homogeneous logit model. However, as shown in equation (2.16), the ratio of probabilities depends not only on the attributes of individual cars but also on the overall characteristics of the nest, illustrating the IIN (Independence of Irrelevant Nests) property. Notably, only the nests containing either j or k are considered in this context; nests that contain neither are not considered in the comparison.

The nested logit model significantly improves upon the homogeneous logit model by relaxing the IIA assumption, allowing for within-nest correlation, and thereby capturing more realistic substitution patterns. This leads to better model fit, more accurate predictions, and more reliable estimates of consumer behavior. Additionally, the hierarchical decision-making process and the flexibility in defining nests enable a more nuanced representation of consumer choices.
However, despite these improvements, the nested logit model still has some limitations.

Firstly, it assumes homogeneous preferences among consumers, meaning it does not account for individual-level variation. Secondly, while the IIA assumption is relaxed, the nested logit model still imposes a structured form of substitution that may not fully capture the complexity of consumer behavior.

Modelling heterogeneous consumers with the random coefficient logit model

To further relax the IIA/IIN assumption we can allow for individual taste variation using a random coefficients logit model, which allows for greater flexibility in consumer preference heterogeneity and can accommodate more complex substitution patterns among cars (Train, 2009). We will decompose the utility from the observed attributes into two components $\bar{\boldsymbol{\beta}}$ to denote the average utility for an attribute and $\tilde{\boldsymbol{\beta}}_i$ to denote the utility that varies for each individual (we will not let α vary), thus we can now write the decomposed utility into:

$$U_{ijt} = \alpha p_{jt} + \bar{\boldsymbol{\beta}}\boldsymbol{x}_{jt} + \tilde{\boldsymbol{\beta}}_i\boldsymbol{x}_{jt} + \xi_{jt} + \epsilon_{ijt} \tag{2.17}$$

Where ξ_{jt} again is the unobserved product-specific utility, and ϵ_{ijt} is unobserved individual-specific preferences. We can combine the mean utility into $\bar{\delta}_{jt} = \alpha p_{jt} + \bar{\boldsymbol{\beta}}\boldsymbol{x}_{jt} + \xi_{jt}$, thus rewriting (2.17) into

$$U_{ijt} = \bar{\delta}_{jt} + \tilde{\boldsymbol{\beta}}_i\boldsymbol{x}_{jt} + \epsilon_{ijt} \tag{2.18}$$

We again assume that ϵ_{ijt} is IID type I extreme value distributed and that $\tilde{\boldsymbol{\beta}}_i$ has the density $f(\tilde{\boldsymbol{\beta}}_i|\boldsymbol{\theta})$. Here $\boldsymbol{\theta}$ are parameters that define the the density of our random coefficients. From that, the choice probability of car j in time t for consumer i is given as:

$$Pr(j|\tilde{\beta}_i, t) = \frac{\exp\left(\bar{\delta}_{jt} + \tilde{\boldsymbol{\beta}}_i\boldsymbol{x}_{jt}\right)}{\sum_{k\in J_t}\exp\left(\bar{\delta}_{kt} + \tilde{\boldsymbol{\beta}}_i\boldsymbol{x}_{kt}\right)} \tag{2.19}$$

To get the aggregate choice probability, which is the market share of car j, we can then integrate over the density of $\tilde{\boldsymbol{\beta}}_i$ thus getting

$$s_{jt} = \int Pr(j|\tilde{\boldsymbol{\beta}}_i, t)\mathrm{d}f(\tilde{\boldsymbol{\beta}}_i|\boldsymbol{\theta}) \tag{2.20}$$

Where s_{jt} is the estimated market share of car j at time t. Which can be rewritten as a function of $\bar{\delta}_{jt}$ and $\boldsymbol{\theta}$ to

$$s_{jt}\left(\bar{\boldsymbol{\delta}}_t, \boldsymbol{\theta}\right) = \int \frac{\exp\left(\bar{\delta}_{jt} + \tilde{\boldsymbol{\beta}}_i\boldsymbol{x}_{jt}\right)}{\sum_{k\in J_t}\exp\left(\bar{\delta}_{kt} + \tilde{\boldsymbol{\beta}}\boldsymbol{x}_{kt}\right)} df(\tilde{\boldsymbol{\beta}}_i|\boldsymbol{\theta}) \tag{2.21}$$

To get the cross-price elasticity for the random coefficient model, we follow Nevo (2000a) and get the cross-price elasticity of the market, using s_{ijt} to be equivalent to the individual choice probability given in (2.19):

$$\mu_{jkt} = \begin{cases} -\frac{p_{jt}}{s_{jt}}\alpha \int s_{ijt}\left(1 - s_{ijt}\right) df(\tilde{\boldsymbol{\beta}}_i|\boldsymbol{\theta}) & \text{if } j = k \\ \frac{p_{kt}}{s_{jt}}\alpha \int s_{ijt}s_{ikt} df(\tilde{\boldsymbol{\beta}}_i|\boldsymbol{\theta}) & \text{if } j \neq k \end{cases} \tag{2.22}$$

We get a more realistic substitution pattern between the different cars. We however still have a problem, since we do not know $\boldsymbol{\theta}$, we will therefore simulate these choice probabilities using the BLP approach.

2.2 Estimating Demand

We will use an ordinary least squares (OLS) and instrument variables (IV) estimator for the homogenous model (2.6) and the nested model (2.13), and then use the BLP estimation technique that utilizes a random coefficient logit model.

Ordinary Least Squares

When using the OLS estimator, the key assumptions to note for us are exogeneity and normality. As we use the homogeneous logit model, we assume that all individuals have the same preferences and that the variance of the error term ϵ_{jt} is constant across individuals. We can also observe that equation (2.6) is linear, which makes it suitable for an OLS estimation.

In OLS, the normality of errors assumption states that the error terms should follow a normal distribution. However, in our logit model, the error term follows an extreme value type 1 distribution. This violation can lead to the standard errors estimated by OLS to be incorrect because the normality assumption helps correctly estimate the variance of the error terms.

While the normality assumption is not strictly required for the OLS estimators to be unbiased, it is important to construct confidence intervals.

The exogeneity assumption is crucial for the OLS estimators to be unbiased and consistent. This assumption requires that the error term ϵ_{jt} is uncorrelated with the independent variables p_{jt} and $\boldsymbol{x}_{jt}$. However, in our model, we can reasonably assume that the price of a car p_{jt} is correlated with the error term ϵ_{jt}. This correlation is likely due to unobserved characteristics that affect both the price and the utility of the car. To address this we can use the IV estimator.

Accounting for endogeneity with IV estimation

As mentioned in the previous section, we will utilize an IV regression to account for the endogeneity we have in our model using an instrument z_{jt}.

By using an instrument that is correlated with the price but not with the error term, we can isolate the exogenous variation in price from these endogenous influences. In the original BLP paper, the authors use the sum of the characteristics of competing models (competing models are defined as models sold in the same year) as instruments. However, Gandhi and Houde (2019) proposes a more effective instrument (among several), which they refer to as differentiation instrumental variables. The point of these instruments is to reflect degrees of differentiation between the products. The instrument can be made by comparing competing cars within a certain distance of their characteristics. We will focus on the instrument they refer to as the local difference, which Gandhi and Houde (2019) concludes tends to perform better than other instruments they investigated. This instrument counts for each characteristic, how many competing cars have characteristics close to it. The local difference instrument can be defined as

$$\boldsymbol{z}_{jt} = \left\{ \sum_{k \in J} \mathbf{1}\left(|x_{jt}^{(1)} - x_{kt}^{(1)}| < sd_1\right), \ldots, \sum_{k \in J} \mathbf{1}\left(|x_{jt}^{(m)} - x_{kt}^{(m)}| < sd_m\right) \right\} \tag{2.23}$$

Where $x_{jt}^{(m)}$ is the m'th attribute of the car j at time t, and sd_m denotes the standard deviation of $x_{jt}^{(m)}$. The instrument is a vector, where each element is the number of competing cars within its standard deviation.

Intuitively, it makes sense that a larger difference in characteristics between two models suggests less direct competition. A significant differentiation can considerably impact the pricing strategy of a model, as manufacturers might price a car differently depending on the degree to which it competes with models having similar characteristics. Thus we can more precisely instrument the effect of competition on price by focusing on how differentiated a model is in the market relative to its competitors. For example, a luxury sports car might not significantly alter its pricing based on the prices of low-end family cars, due to the large difference in their characteristics.

This addresses the endogeneity problem, but we still have IIA across the different cars, which we will address in the next sections using a nested logit model and the full BLP model.

BLP estimation framework

The BLP (Berry, Levinsohn, and Pakes) estimation technique, presented in Berry *et al.* (1993), is particularly useful when simultaneously dealing with both endogeneity and the IIA problem. This method incorporates the random coefficients logit model. First, since we do not have any demographical data, we will assume the distribution of the individual-specific preferences to be normal distributed. Since $\bar{\boldsymbol{\delta}}$ already contains the mean utility, we will assume $\tilde{\boldsymbol{\beta}}_i \sim \mathcal{N}(0, \sigma)$, thus we will simply have $\boldsymbol{\theta} = \boldsymbol{\sigma}$. We can rewrite $\tilde{\boldsymbol{\beta}}_i = \boldsymbol{\sigma v x}_{jt}$, where $\boldsymbol{v} \sim \mathcal{N}(0,1)$. The choice probability of a given car j at time t can be given as a function of $\bar{\boldsymbol{\delta}}$ and $\boldsymbol{\sigma}$:

$$Pr(\bar{\boldsymbol{\delta}}, \sigma) = \int \frac{\exp(\bar{\delta}_{jt} + \sigma_j \boldsymbol{v x}_{jt})}{\sum_{k \in J_t} \exp(\bar{\delta}_{kt} + \sigma \boldsymbol{v x}_{kt})} \phi(\boldsymbol{v}) d\boldsymbol{v} \tag{2.24}$$

The integral does not have a closed-form solution, so we will have to simulate it using the Monte Carlo approximation given as

$$\int \frac{\exp(\bar{\delta}_{jt} + \sigma_j \boldsymbol{v x}_{jt})}{\sum_{k \in J_t} \exp(\bar{\delta}_{kt} + \sigma \boldsymbol{v x}_{kt})} \phi(\boldsymbol{v}) d\boldsymbol{v} \cong R^{-1} \sum_{i=1}^{R} \frac{\exp(\bar{\delta}_{jt} + \sigma \boldsymbol{v}_i \boldsymbol{x}_{jt})}{\sum_{k \in J_t} \exp(\bar{\delta}_{kt} + \sigma \boldsymbol{v}_i \boldsymbol{x}_{kt})} \tag{2.25}$$

This means we draw $\{\boldsymbol{v}_r\}_{i=1}^{R}$ with $\boldsymbol{v}_i \sim \mathrm{IID}\, \mathcal{N}(0,1)$, thus the different $\boldsymbol{v}$'s represents the different individual's preferences, and $\boldsymbol{\sigma}$ represents how much the taste preference differs from the mean. This leads to the next problem, we do not know $\bar{\boldsymbol{\delta}}$ or $\boldsymbol{\sigma}$, thus they need to be estimated. Instead of estimating both $\boldsymbol{\delta}$ and $\boldsymbol{\sigma}$, which will be very computationally demanding, Berry *et al.* (1993) showed that for a given $\boldsymbol{\sigma}$ there is a unique $\bar{\boldsymbol{\delta}}$ that makes the estimated market share equate the observed market share, thus defining a contraction mapping algorithm to estimate the unique $\bar{\boldsymbol{\delta}}$ as:

$$\bar{\boldsymbol{\delta}}_t^{s+1} = \bar{\boldsymbol{\delta}}_t^{s} + \log \mathcal{S}_t - \log \mathrm{s}_j \left(\bar{\boldsymbol{\delta}}_t^{s}, \sigma \right) \tag{2.26}$$

which converges when $\bar{\boldsymbol{\delta}}_t^{s+1} \approx \bar{\boldsymbol{\delta}}_t^{s}$, which is when $\log \mathcal{S}_t \approx \log \mathrm{s}_j \left(\bar{\boldsymbol{\delta}}_t^{s}, \sigma \right)$, where $\mathrm{s}_j \left(\bar{\boldsymbol{\delta}}_t^{s}, \sigma \right)$ can be found by using the simulated choice probabilities in (2.25). When (2.26) converges, we get the estimated mean utility $\hat{\boldsymbol{\delta}}$. We can estimate α and $\bar{\boldsymbol{\beta}}$ using the same approach as in the IV estimation:

$$\hat{\boldsymbol{\delta}}_t = \alpha \boldsymbol{p}_t + \bar{\boldsymbol{\beta}} \boldsymbol{x}_t + \boldsymbol{\xi}_t \tag{2.27}$$

We save the residuals $\hat{\boldsymbol{\xi}}_t$, as we still need to find the best $\boldsymbol{\sigma}$. Finding the optimal $\boldsymbol{\sigma}$ can be done by minimizing the criterion function given by:

$$\min_{\theta} g(\sigma)' \mathrm{W} g(\sigma) \tag{2.28}$$

with

$$g(\sigma) \equiv \frac{1}{N} \sum_t \sum_j \hat{\xi}_{jt} Z_{jt} \tag{2.29}$$

Where W is the weighting matrix and Z_{jt} is the instrument. Since this approach requires so many steps, we have written a simplified pseudocode below

BLP Estimation Technique

Outer: $\min_\theta g(\sigma)' \mathrm{W} g(\sigma)$

$\delta \leftarrow$ contraction mapping(σ)

$\alpha, \bar{\beta} \leftarrow$ iv estimation(δ) *, save residuals in* $\hat{\xi}$

$g \leftarrow$ criterion$(\hat{\xi})$

2.3 Supply side

Next, we examine the supply side. We assume a multi-product oligopoly where firms independently and simultaneously set prices for each market. We assume that there are F firms where each firm produces some subset of the total amount of cars F_{jt}
We can then write the firm's profit per average consumer as:

$$\Pi_F = \sum_{j \in J_{ft}} (p_{jt} - c_{jt}) s_{jt} \tag{2.30}$$

Where s_{jt} is the market share of car j that can be approximated as a function of price of all cars, and c_{jt} is the marginal cost of producing car j at time t.

We can then get the FOC for each product j by following the same course of action from Conlon and Gortmaker (2020b):

$$\text{FOC} : s_{jt}(\mathbf{p}_t) + \sum_{k \in J_{ft}} \frac{\partial s_{kt}}{\partial p_{jt}} (p_{kt} - c_{kt}) = 0$$

Thus we have $J \times t$ first order conditions, which we can stack and we define $\Delta_t(\mathbf{p}_t) \equiv -\mathcal{H}_t \odot \nabla_{p_t} \mathbf{s}_t$, where $\nabla_{p_t} \mathbf{s}_t$ is the cross-price elasticity matrix calculated using (2.22), $\mathcal{H}$

is the ownership matrix, $\odot$ is the element-wise Hadamard product. This captures the internalization of cannibalization of a firm producing more than one car:

$$\mathbf{s}_t(\mathbf{p}_t) = \Delta_t(\mathbf{p}_t)(\mathbf{p}_t - \mathbf{c}_t)$$
$$\mathbf{p}_t = \mathbf{c}_t + \underbrace{[\Delta_t(\mathbf{p}_t)]^{-1}\,\mathbf{s}_t(\mathbf{p}_t)}_{=\eta_t(\mathbf{p}_t, s_t, \theta_2)}$$

Here η_t is the Bertrand markup, following from Magnolfi *et al.* (2022). Resulting in the price equilibrium:

$$\mathbf{p^*}_t = \mathbf{c}_t + \left[-\mathcal{H}_t \odot \nabla_{\mathbf{p^*}_t}\mathbf{s}_t\right]^{-1} \mathbf{s}_t(\mathbf{p^*}_t) \tag{2.31}$$

We can use equation (2.31) in several different ways, following Bresnahan (1989) the marginal cost cannot be directly or straightforwardly observed. Therefore we must use the estimation of the demand system along with the observed prices to estimate the firms' marginal cost using the observed prices and market share:

$$\hat{\mathbf{c}}_t = \mathbf{p^*}_t - \left[-\mathcal{H}_t \odot \nabla_{\mathbf{p^*}_t}\mathbf{s}_t\right]^{-1} \mathcal{S}_t \tag{2.32}$$

With the marginal cost now estimated, we can use the same approach as Nevo (2000b) and the assumption of a Bertrand-Nash equilibrium to predict the prices of a post-merger situation, giving us the following equation:

$$\mathbf{p}_t^{post} = \hat{\mathbf{c}}_t + \left[-\mathcal{H}_t^{post} \odot \nabla_{\mathbf{p}_t^{post}}\mathbf{s}_t\right]^{-1} \mathbf{s}_t(\mathbf{p}_t^{post}) \tag{2.33}$$

Here $\mathcal{H}^{post}$ is the ownership matrix after the merges, see figure 2.2 below for an example of how the ownership matrix changes, $\mathbf{p_t^{post}}$ is the post-merge price vector and $\hat{\mathbf{c}}_t$ is the marginal cost implied by our demand system in the pre-merger ownership structure. Having the marginal cost staying constant after merges makes several assumptions that might not hold in the real world. Firstly we assume that the company cannot cut costs when merging, and secondly, we assume no colluding in the pre-merger situation as we assume Bertrand-Nash.x

$$\begin{array}{lllll} & \text{Car}_{GM} & \text{Car}_{Toyota} & \text{Car}_{Ford} & \text{Car}_{Nissan} \\ \text{Car}_{GM} & 1 & 0 & 0 & 0 \\ \text{Car}_{Toyota} & 0 & 1 & 0 & 0 \\ \text{Car}_{Ford} & 0 & 0 & 1 & 0 \\ \text{Car}_{Nissan} & 0 & 0 & 0 & 1 \end{array} \xrightarrow{\text{Merge GM\&Ford}} \begin{array}{lllll} & \text{Car}_{GM} & \text{Car}_{Toyota} & \text{Car}_{Ford} & \text{Car}_{Nissan} \\ \text{Car}_{GM} & 1 & 0 & 1 & 0 \\ \text{Car}_{Toyota} & 0 & 1 & 0 & 0 \\ \text{Car}_{Ford} & 1 & 0 & 1 & 0 \\ \text{Car}_{Nissan} & 0 & 0 & 0 & 1 \end{array}$$

Figure 2.2: Example of how the ownership matrix changes in a merger. Left is $\mathcal{H}$ before the merge and right is $\mathcal{H}^{\text{post}}$ after GM and Ford merge.

To understand what drives prices, we can parameterize the marginal cost. This approach helps in identifying and quantifying the factors that impact cost and, consequently, the price the consumer faces.

$$\hat{\mathbf{c}}_t = \boldsymbol{x}_t\boldsymbol{\gamma} + \boldsymbol{\omega}_t \tag{2.34}$$

Where $\boldsymbol{x}_j$ is the cost-driving variables, and $\boldsymbol{\omega}_t$ is the error term that captures unobserved influences. The parameter $\boldsymbol{\gamma}$ reflects the sensitivity of the marginal cost to changes in $\boldsymbol{x}_t$.

As we have the prices in a post-merger situation, we would also be able to analyze the effect a merger would have on consumer welfare and how this merger puts the consumer. Here we again follow the approach and notation from Grieco *et al.* (2023) giving us the individual consumer surplus as:

$$\mathrm{CS}_{it} = \log\left(1 + \sum_{j \in J_t} \exp \delta_{ijt}\right) \cdot \frac{1}{-\alpha} \tag{2.35}$$

$$\mathrm{CS} = \sum_{i \in I_t} \mathrm{CS}_{it} \tag{2.36}$$

Since we can observe or estimate all variables in this equation, we can analyze the potential outcomes if companies were to merge and thus get more market power.

Data

3

The dataset used in this paper is composed of data from 1980 to 2018 with car characteristics, prices, market share, brand, etc. for each car and each year. The full list of characteristics gathered can be seen below in table 3.1.

Table 3.1: Characteristics of Variables

Characteristic	Description
Sales	How many units sold in the given year
Year	The year of the price, sales, etc.
Make	The brand of the producer
Price	Price of the vehicle (inflation adjusted to 2015 dollar)
Height	Physical height of the vehicle
Footprint	Ground area covered by the vehicle
Horsepower	Engine power output
Miles Per Gallon	Fuel efficiency
Curb Weight	Total weight of the vehicle
Number of Trims	Available trim options
Release Year	Year the model was released
Years Since Design	Years since the model was designed
Sport	If the vehicle is a sports model
Electric Vehicle	If the vehicle is electric
Truck	If the vehicle is a truck
SUV	If the vehicle is an SUV
Van	If the vehicle is a van
Make	A set of dummies indicating brand of the car
Year	A set of dummies indicating what year the car is sold

The data is collected from different sources by Grieco *et al.* (2023).

Exploring the data

The dataset contains 9694 rows from all 38 years (1980-2018), where every year has an average of 248.5 different models each year from 62 producers, whereas some have the same parent company, with a total of 33 parent companies. E.g the producers "AMC", "Jeep", and "Renault" are all owned by the parent company "Renault". A general summary of the data can be seen below in table 3.2.

Table 3.2: Descriptive Statistics of Vehicle Characteristics with Enhanced Decimal Precision

Statistic	Sales	Prices	Height	Footprint	HP	MPG	Curb Weight
Mean	59103.39	36.05	60.95	13392.63	192.18	20.94	3561.21
Std Dev	86940.25	17.13	8.41	1968.92	83.88	6.58	897.77
Min	10.00	11.14	43.50	6514.54	44.00	10.00	1113.00
25th Pctl	7990.50	24.08	54.70	12000.16	127.50	17.00	2925.00
Median	27394.00	31.82	57.70	13330.06	175.00	19.00	3470.50
75th Pctl	74560.50	43.40	67.10	14532.73	250.00	23.00	4045.50
Max	891482.00	99.99	107.50	21821.86	645.00	50.00	8550.00
Count	9694						

Every year an average of 59103 vehicles are sold with a mean price of $36,000$\$, but more interestingly the changes in these characteristics can be seen below in figure 3.1.

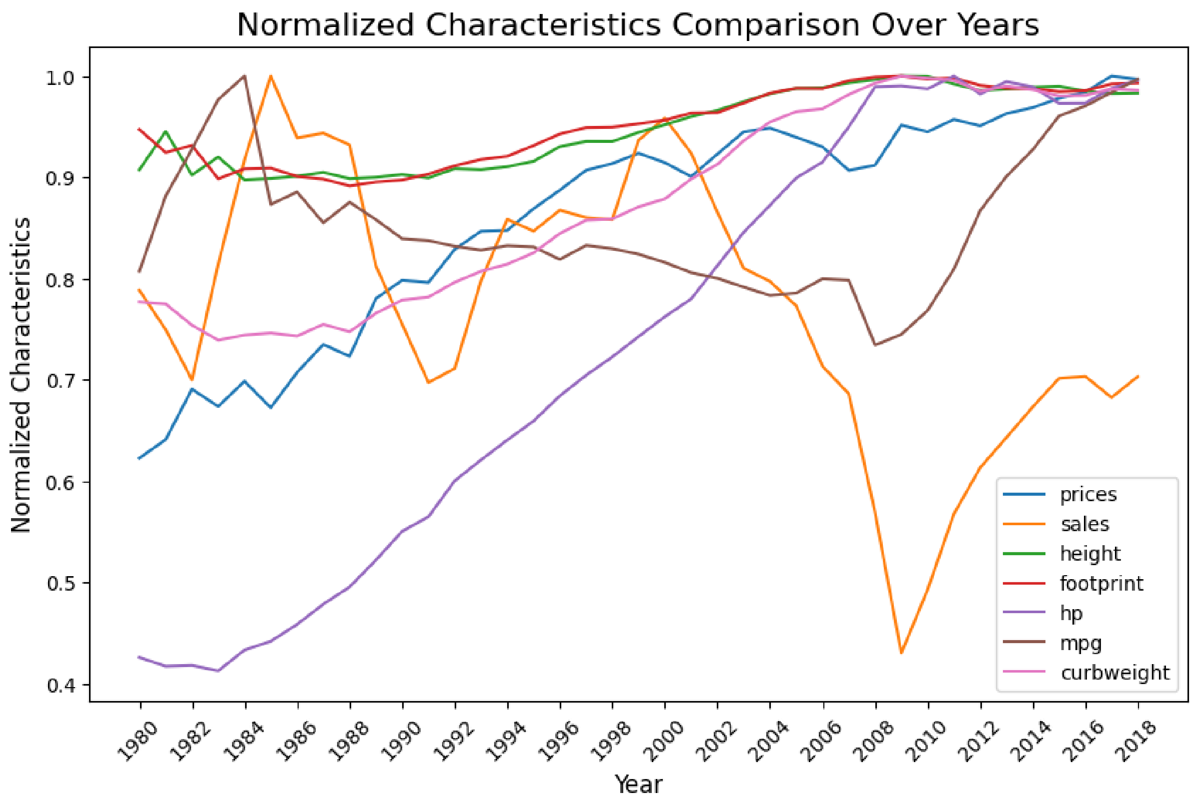

Figure 3.1: Mean of normalized characteristics over time

From the plot, we see that the cars in the market have become larger, heavier, and faster (with regards to horsepower), but also more expensive. The average car price in 1980 was $25,500$\$, but in 2018 it was $40,840$\$.

Market Shares

For the demand estimation, we need market share as a proxy for utility for each product. This is however not easily found, thus we need to construct this from the sales data and the market size. Since we do not know the exact market size either, we estimate it using the assumption that each household has 2 cars and buys one new car every 5th year. To reflect this, we divide the number of households each year by 2.5 (Grieco *et al.*, 2023). The market share for a given model i in year j is therefore

$$\text{market size}_j = \frac{\#\text{households}_j}{2.5}$$
$$\text{market share}_{ij} = \frac{\text{sales}_{ij}}{\text{market size}_j}$$

As we are interested in the evolution of market powers, an interesting statistic is how much the different parent companies control the market. The sum of market shares of all the cars the parent companies produce for each year is plotted below in figure 3.2.

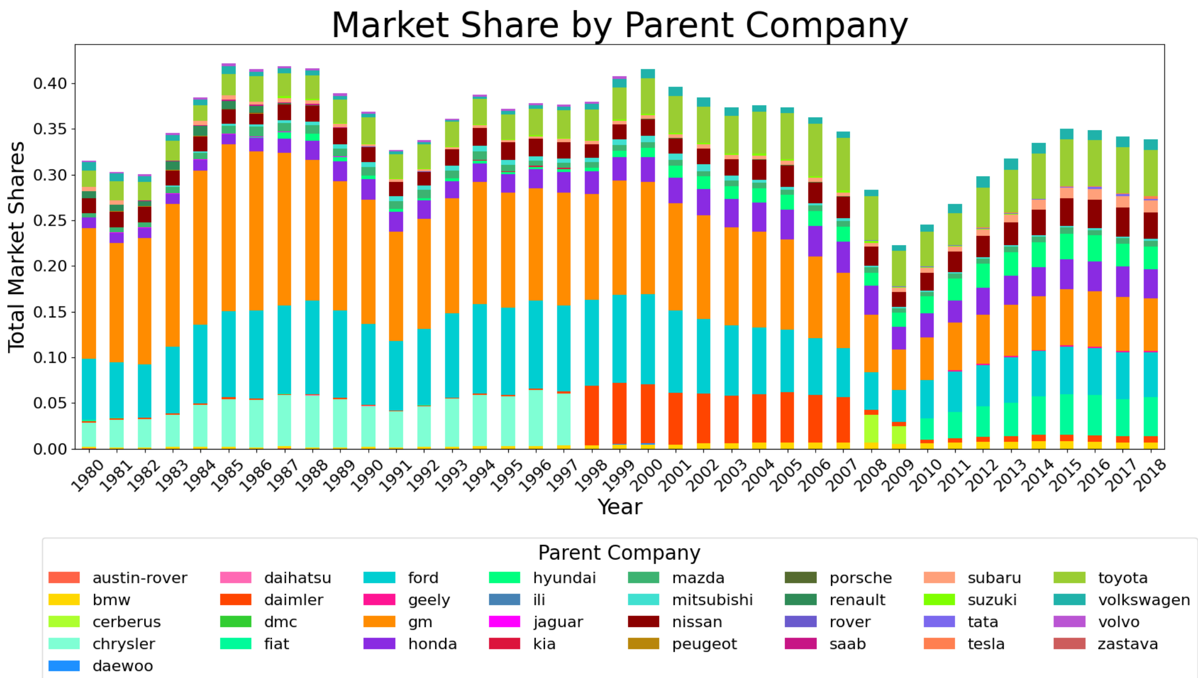

Figure 3.2: Market share for companies in the period 1980 to 2018.

The plot shows that the dominance of the largest companies has decreased, leading to a more evenly distributed market. This means that no single-parent company has a substantial hold on the market anymore. To further analyze this, we can plot the Herfindahl-Hirschman Index (HHI)[1], which is a common measure for market concentration (Herfindahl, 1950), see figure 3.3 below.

[1] Computed simply as the sum of squared market shares HHI $= \sum_{f \in F} \mathcal{S}_f^2$.

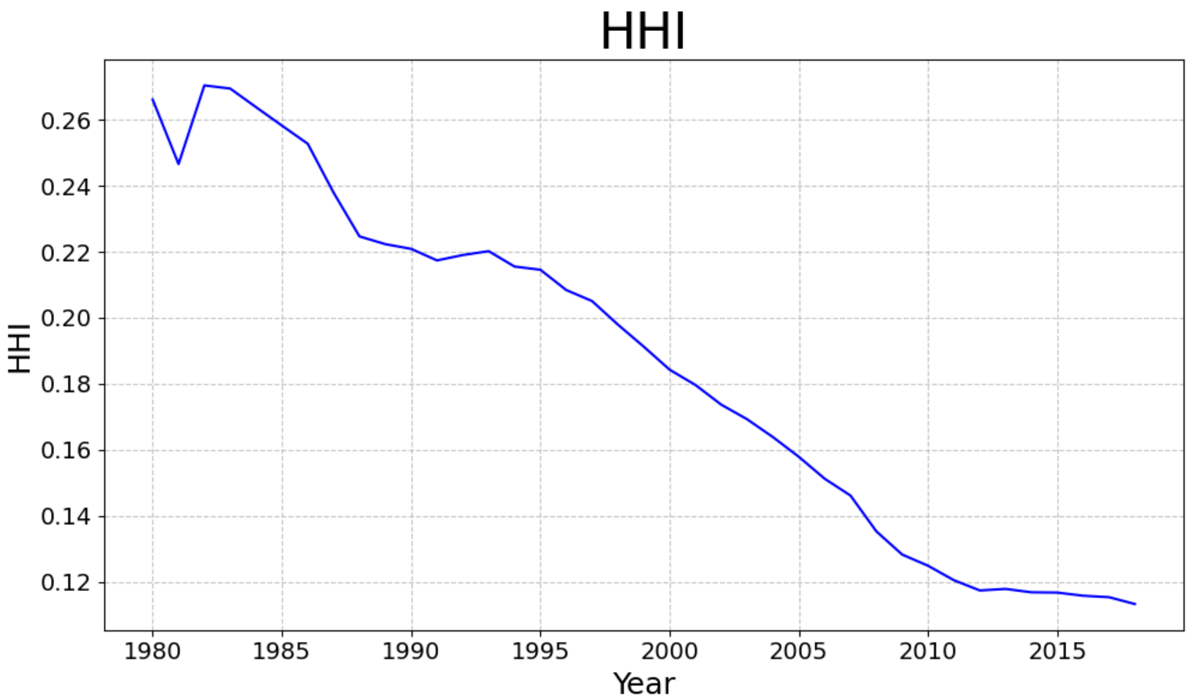

Figure 3.3: HHI of the car market during the period 1980-2018.

As expected we also see a general downward trend, thus we have a less concentrated and more competitive market in 2018 than in 1980. According to economic theory, this increased competition should lead to lower markups, as companies are forced to compete more aggressively for market share.

4 Results

All estimation techniques used in this study have been custom-programmed by us in Python, utilizing only fundamental mathematical and statistical libraries such as `Numpy`, `Pandas`, and `Statsmodels`. This approach ensures full transparency and control over the implementation process.

We first estimate the plain homogenous logit model using the OLS estimator to establish a baseline for our analysis. This makes it easier to understand basic relationships in our model, and serves as a useful benchmark when looking at the performance and improvements made by the more sophisticated models.

IV estimation results

We then estimate the homogenous logit model using an IV 2sls estimator using the local differentiation instrument mentioned in the theory section. We have also included the original BLP instrument of the sum of the characteristics, as well as the instrument used in Grieco *et al.* (2023)[1]. Estimation results for the IV regression using these instruments can be seen below in table 4.1.

[1]Lagged Real Exchange Rate for each automobile (RXR), defined as the price level of expenditure in the country of manufacture, reflecting the purchasing power parity exchange rate relative to the U.S. divided by the nominal exchange rate.

Table 4.1: IV Estimation Results

Variable	BLP-instrument	RXR-instrument	Local diff
First Stage			
R-squared	0.813	0.815	0.825
Second Stage			
Constant	-16.85 (2.00)	-41.10 (4.55)	-17.35 (2.38)
Price	-0.39 (0.05)	-1.69 (0.22)	-0.42 (0.08)
Log Height	-4.47 (1.28)	-14.65 (2.21)	-4.68 (1.44)
Log Footprint	19.74 (2.28)	8.23 (3.25)	19.50 (2.59)
Log Horsepower	2.99 (0.86)	17.23 (2.54)	3.28 (1.03)
Log MPG	-0.39 (0.42)	2.12 (0.65)	-0.33 (0.48)
Log Curbweight	-7.25 (0.77)	33.55 (7.39)	-6.40 (3.21)
Log Number of Trims	6.32 (0.12)	1.01 (0.04)	6.30 (0.13)
Release Year	-0.41 (0.05)	-0.52 (0.05)	-0.41 (0.05)
Years Since Design	-2.76 (0.15)	-0.11 (0.01)	-2.76 (0.14)
Sport	-0.52 (0.06)	0.14 (0.12)	-0.50 (0.07)
Electric Vehicle (EV)	-0.95 (0.21)	-0.09 (0.25)	-0.93 (0.21)
Truck	-0.66 (0.07)	-1.19 (0.11)	-0.67 (0.08)
SUV	0.50 (0.06)	0.36 (0.06)	0.50 (0.06)
Van	-0.08 (0.09)	-0.42 (0.10)	-0.08 (0.09)

Robust standard errors are reported in parentheses. Dummies are not included in the table for make and year. All logged values are standardized.

Given that Gandhi and Houde (2019) concludes, that the local differentiation instrument performed the best compared to the BLP instrument, and that the local differentiation instrument has a higher R^2 value than the RXR-instrument, we will proceed with the local differentiation instrument for further IV estimations.

When implementing the instruments, we aimed to address the bias in the price coefficient observed in our initial OLS estimates, which indicated a price coefficient of -0.34. The OLS method did not adequately account for the influence of unobserved factors, which are inherently linked to the price. To correct for this bias, we used the IV regression. This adjustment revealed a more accurate and pronounced negative relationship between price and demand, with the IV-adjusted price coefficient being -0.51. This finding confirms that the OLS method underestimated the true impact of price on demand due to the influence of unobserved competitive factors. The IV-adjusted coefficient thus provides a clearer and more reliable estimate of the price sensitivity in the market.

Nested logit results

As mentioned in earlier sections we introduce the use of the nested logit model to allow for more flexibility when modeling the demand, and also to relax the assumption of independence of irrelevant alternatives. To estimate the utility in the nested logit model, as shown in equation (2.13), it is necessary to estimate the nesting parameter ρ.
We will utilize the same estimator and instrument on the nested logit model as used on the homogeneous logit model. The estimation of ρ involves testing different values and minimizing the mean of the squared residuals, see figure 4.1 below where we estimate $\rho = 0.941$. This implies that there is a very large correlation of the error terms within the nests, meaning that the products in the same nests are very good substitutes for each other.

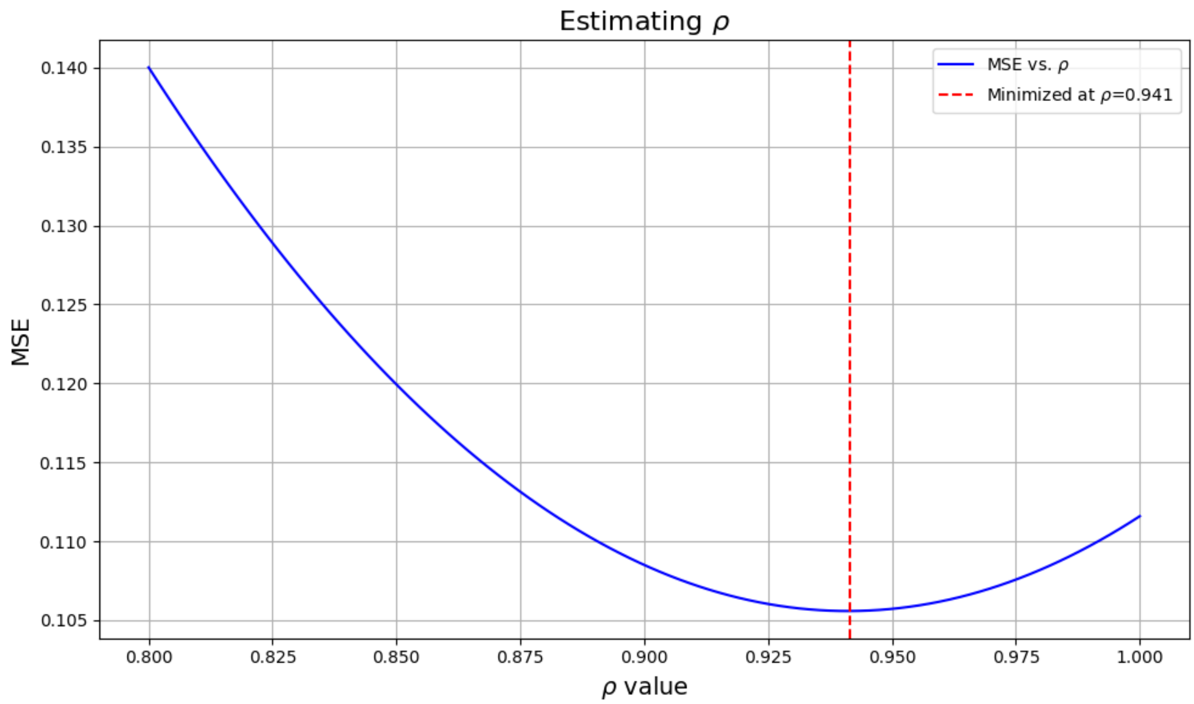

Figure 4.1: Finding the ρ values that minimizes the MSE

BLP results

For the more complex BLP model, we follow the pseudocode from 2.2, where we first have to find the σ that minimizes the criterion function. Since we are letting $\sigma_{\text{Log HP}}$ and $\sigma_{\text{Log MPG}}$ vary over individuals, we have performed a grid search to find suitable σ values. We find and hold $\sigma_{\text{Log HP}} = 2.4$ and then search for the optimal $\sigma_{\text{Log MPG}}$ around this value. We get the minima at $\sigma_{\text{Log MPG}} = 1.1$, which can be seen below in figure 4.2.

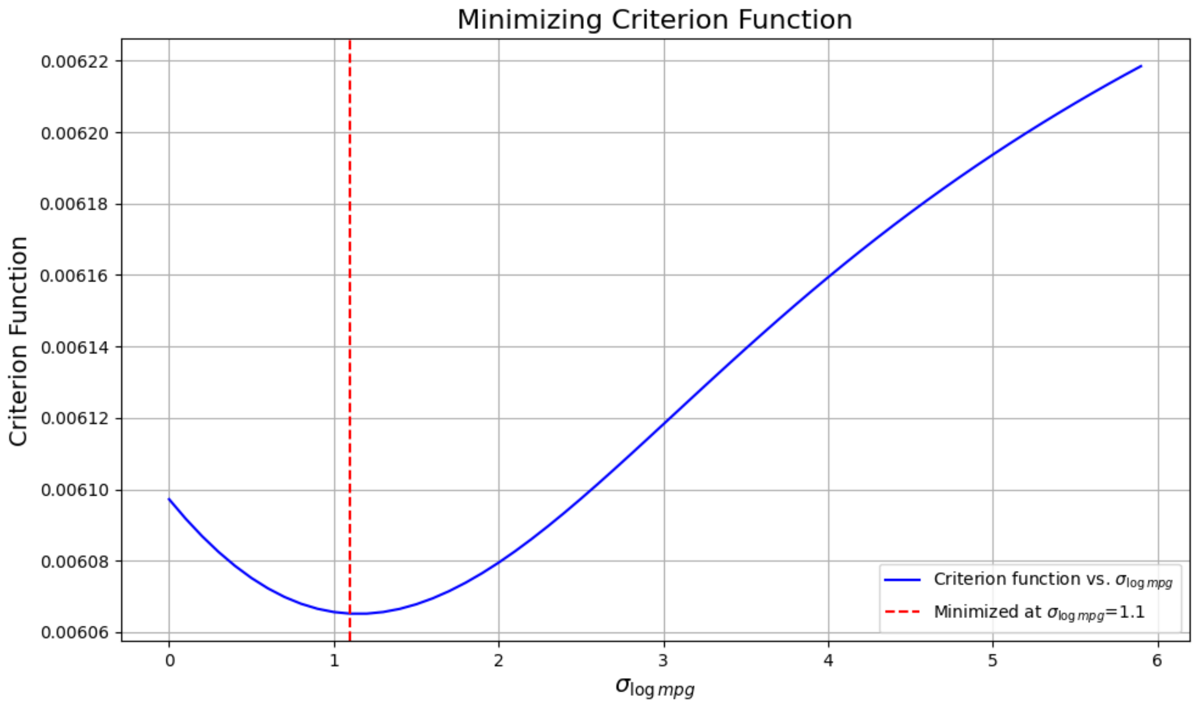

Figure 4.2: Minimizing the criterion function while $\sigma_{\text{Log HP}} = 2.4$

In the literature, there is not a consensus on whether this optimization problem is convex or not. Knittel and Metaxoglou (2014) struggles to make optimisation algorithms converge, whereas Conlon and Gortmaker (2020a) does not recognize this problem. As can be seen on figure 4.2 our estimation supports Conlon and Gortmaker (2020a). But as this is quite computational, we have not been able to establish a definitive conclusion whether this is indeed a global minima. Given that we need to perform many simulations for each σ-value, we only let $\log$ HP and $\log$ MGP vary, as letting another attribute vary, we add a quadratic number of needed simulations.

Table 4.2: Regression Results

Variable	OLS	IV	Nested IV	BLP
Constant	-15.90 (1.81)	-17.35 (2.38)	-6.93 (0.59)	-16.42 (2.38)
Price	-0.34 (0.02)	-0.42 (0.08)	-0.11 (0.02)	-0.43 (0.08)
Log Height	-0.87 (0.27)	-4.68 (1.44)	-3.97 (0.38)	-4.83 (1.44)
Log Footprint	2.02 (0.24)	19.50 (2.59)	9.61 (0.92)	19.48 (2.59)
Log Horsepower	0.38 (0.10)	3.28 (1.03)	0.13 (0.05)	1.51 (0.47)
Log MPG	-0.12 (0.11)	-0.33 (0.48)	0.52 (0.14)	-0.76 (0.48)
Log Curbweight	-0.98 (0.21)	-6.40 (3.21)	-1.24 (0.87)	-6.06 (3.21)
Log Number of Trims	1.19 (0.02)	6.30 (0.13)	0.09 (0.01)	6.30 (0.13)
Release Year	-0.41 (0.05)	-0.41 (0.05)	-0.02 (0.01)	-0.41 (0.05)
Years Since Design	-0.11 (0.01)	-2.76 (0.14)	-0.01 (0.002)	-2.76 (0.14)
Sport	-0.54 (0.05)	-0.50 (0.07)	-2.53 (0.02)	-0.50 (0.07)
Electric Vehicle (EV)	-0.98 (0.20)	-0.93 (0.21)	0.01 (0.04)	-0.93 (0.21)
Truck	-0.64 (0.07)	-0.67 (0.08)	-1.00 (0.02)	-0.68 (0.08)
SUV	0.51 (0.05)	0.50 (0.06)	-0.42 (0.02)	0.50 (0.06)
Van	-0.06 (0.08)	-0.08 (0.09)	-1.65 (0.02)	-0.09 (0.09)
Mean own price elasticity	-1.21	-1.50	-6.50	-3.74
ρ			-0.94	
$\sigma_{\text{Log HP}}$ & $\sigma_{\text{Log MPG}}$				2.4 & 1.1

Robust standard errors are reported in parentheses. Dummies are not included in the table for make and year. All logged values are standardized.

Most of our parameter estimates align with intuitive expectations. For instance, the price coefficient is negative, indicating that higher prices reduce utility. Similarly, the positive coefficient for horsepower suggests that more powerful cars are more desirable to consumers.

One parameter that stands out is the MPG (miles per gallon), while we would assume that better fuel efficiency would positively impact the utility, the coefficient is not significantly different from zero.

Supply side results

As discussed in the previous supply section, the price the firm sets is a combination of the cost of producing the car and the markup. We will analyze how much each attribute contributes to an increase in the marginal cost.

Given the estimated marginal costs using equation (2.32), we regress the estimated marginal cost on the car's characteristics from the model given in equation (2.34), using an OLS estimation. Estimation results can be seen below in table 4.3.

Table 4.3: Cost Estimation Results

Variable	γ
Constant	-18.66 (1.24)
Log Height	-7.12 (0.82)
Log Footprint	-10.12 (1.52)
Log Horsepower	10.87 (0.52)
Log MPG	1.83 (0.30)
Log Curbweight	30.96 (1.42)
Log Number of Trims	-0.72 (0.05)
Sport	0.50 (0.04)
Electric Vehicle (EV)	0.64 (0.12)
Truck	-0.41 (0.04)
SUV	-0.13 (0.03)
Van	-0.29 (0.04)

Robust standard errors are reported in parentheses. Dummies are not included in the table for make and year. All logged values are standardized.

As our intuition suggests when manufacturing a car with greater weight, higher horsepower, and improved fuel efficiency (measured in miles per gallon) gives higher costs. This is also supported by our data analysis, as illustrated in figure 3.1, which demonstrates a correlation between increasing vehicle prices and increases in these characteristics over time. Therefore, when evaluating consumer welfare in the car market, it is insufficient to attribute rising prices solely to producers' price increases, as their marginal cost would also have gone up due to the fact that the cars have become better.

Therefore we will analyse the evolution of the markup in the car market during the period. To calculate the markup we will follow (Grieco *et al.*, 2023) and use the Lerner index $\left(\frac{p-c}{p}\right)$. As can be seen on figure (4.3), we display the distribution of markups during the period.

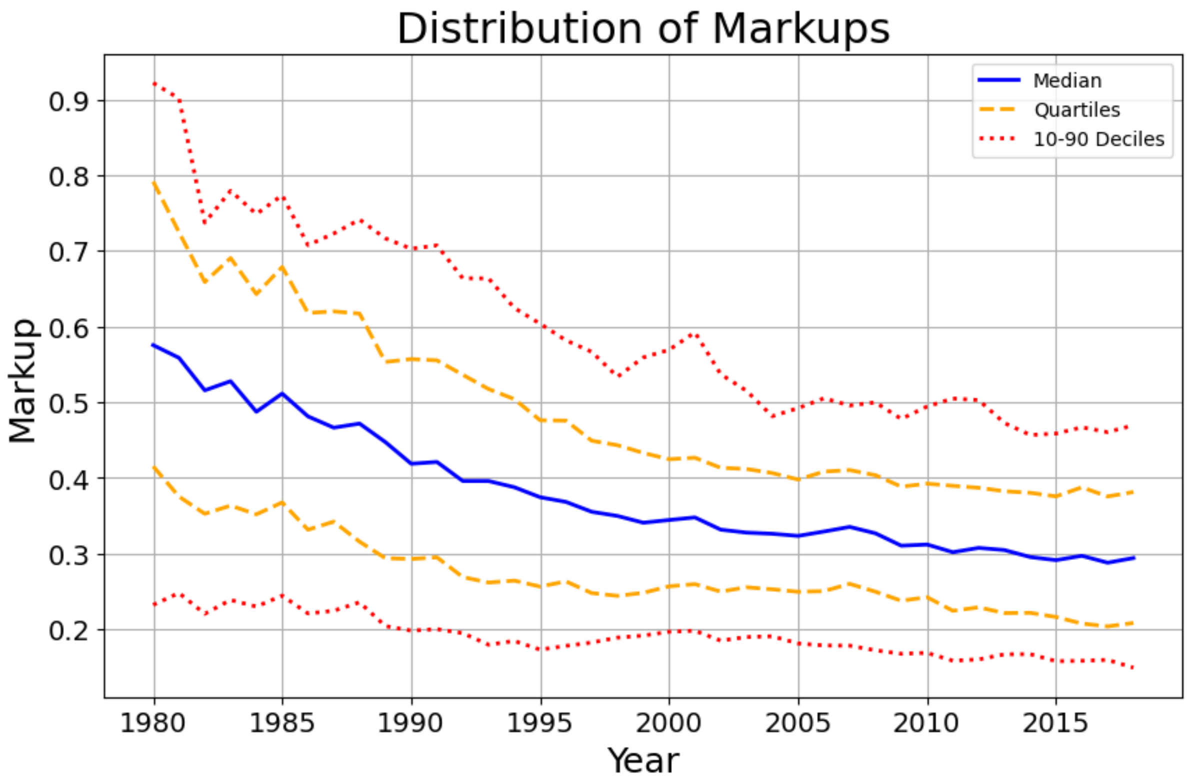

Figure 4.3: Distribution of markups over the years

We see that all markups are falling, both the more profitable cars with the higher markups in the 90% percentile and the less profitable cars in the 10% percentile, we calculate that our median markup falls from 0.59 in 1980 to 0.29 in 2018.

5 Counterfactuals

Having developed models for both the demand side, employing the full BLP model, and the supply side under the multi-product Bertrand-Nash framework, we can now make various counterfactual scenarios. We will use the estimations we computed, but use the `PyBLP` library to conduct the simulations. Our primary focus will be on mergers between the largest parent companies regarding total market share, examining the impacts on the market, particularly regarding consumer welfare. We will also shortly analyze the effect of breaking up companies.

From our estimation results and the interpretation of the cost parameterization of cars, we argued that prices alone are insufficient as proxies for consumer welfare. We will therefore analyze changes in market shares, markup, and consumer surplus to gain further insights into market dynamics and the shifts in market power resulting from mergers or breakups.

Merging the biggest companies

We will examine four merger scenarios and compare them to a baseline scenario where the parent companies operate independently. These mergers are constructed based on the average market shares of the various parent companies, specifically merging the top 3 to 6 companies. The details of the companies involved in each merger scenario are presented in table 5.1.[1]

Group	Companies
Top 3	GM, Toyota, Ford
Top 4	GM, Toyota, Ford, Fiat
Top 5	GM, Toyota, Ford, Fiat, Honda
Top 6	GM, Toyota, Ford, Fiat, Honda, Nissan

Table 5.1: Mergers of the Biggest Companies

[1] Fiat includes the ownership of Chrysler, Daimler, and Cerberus, reflecting the changes in ownership over the years.

We conducted a series of simulations to evaluate the impact of mergers among top car companies. First, we examine the changes in market concentration using the HHI, as depicted in figure 5.1.

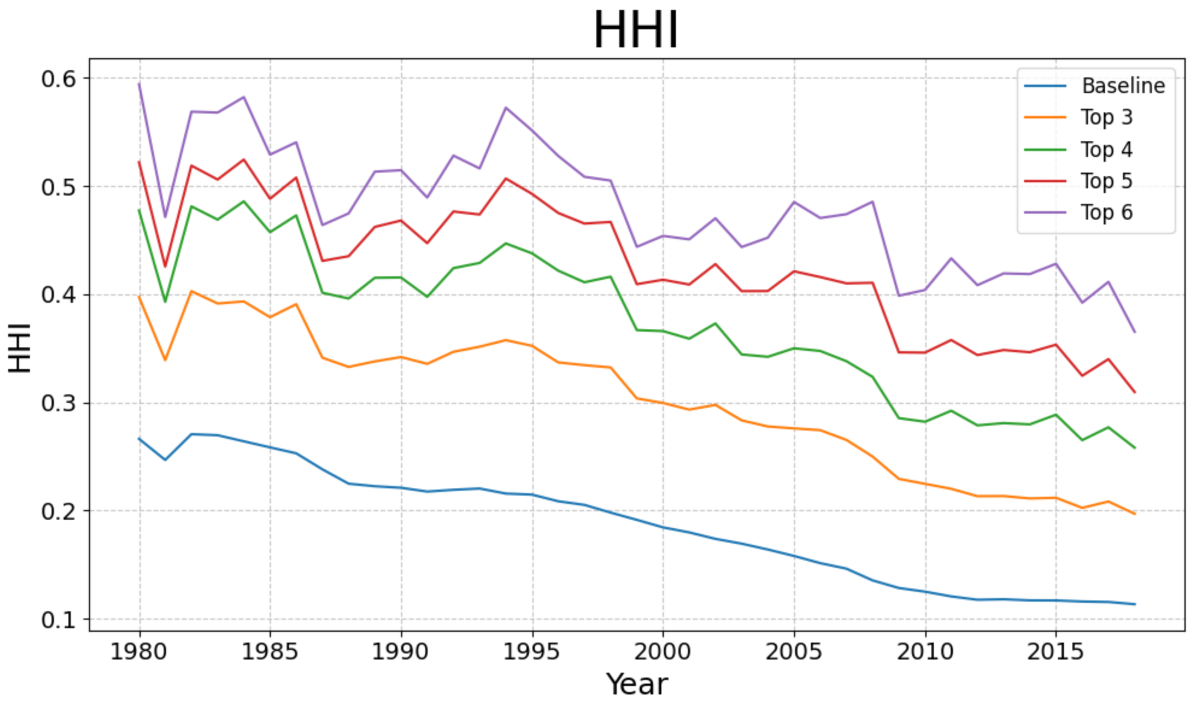

Figure 5.1: Herfindahl-Hirschman Index (HHI) for various mergers among top car companies.

As anticipated, the mergers led to a notable increase in market concentration, resulting in reduced competition. This heightened concentration increases the market power of the merged companies, but this concentration is still decreasing throughout the years. The subsequent sections analyze the broader economic implications of this increased market power.

The increase in market power leads to a rise in estimated mean markups (Figure 5.2). If we merge the six largest companies, the markups increase from the baseline of 0.59 to 0.67 in 1980, and from 0.29 to 0.40 in 2018. While there is still a drop in markups over the period, the merged markups remain higher than the baseline.
However, the rise in prices has an adverse effect on consumer welfare. Figure 5.3 illustrates the percentage change in consumer surplus, demonstrating a clear decline. The reduction in consumer surplus indicates that consumers are worse off post-merger, paying higher prices for the same goods. We estimate that mergers would significantly negatively affect consumer surplus from 1980 to 2000, with a smaller impact in subsequent periods, particularly after 2010.

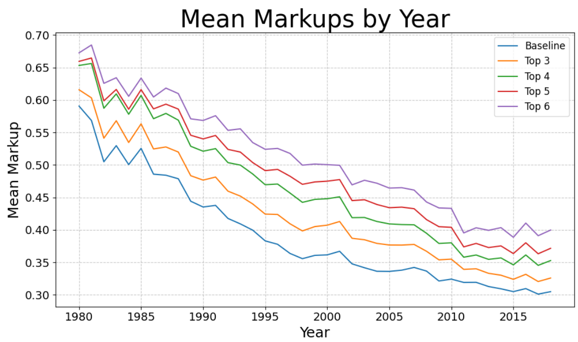

Figure 5.2: Mean Markups

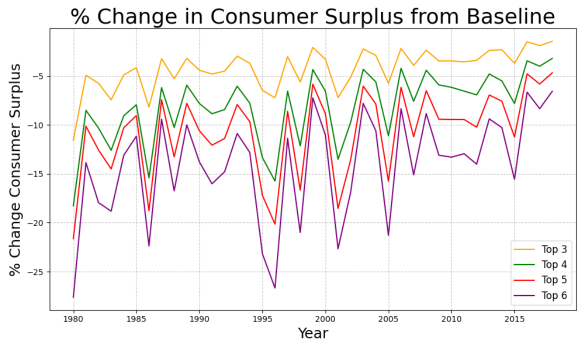

Figure 5.3: Mean change in consumer surplus

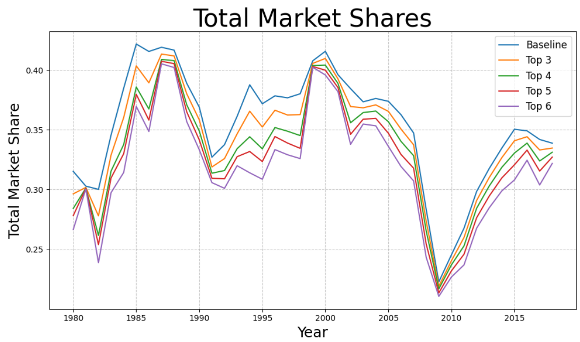

Figure 5.4: Total market shares

This decrease in consumer surplus also correlates with a reduction in total market shares (Figure 5.4), suggesting that the higher prices lead to a decline in the quantity of cars sold. This figure also illustrates the exogenous demand shocks. While we will not explicitly comment on their effects, it is important to note their presence as they impact demand.

An aspect not accounted for in our analysis is the potential reduction in marginal costs resulting from mergers. According to Devos *et al.* (2009), mergers often lead to cost reductions in addition to increasing market power. As our study focuses on the increase in market power, we assume constant costs. Although it is challenging to quantify the exact decrease in marginal costs, it is important to recognize that such reductions are likely to occur. With this cost reduction, it is likely that prices would also decrease. Under the right circumstances, this could result in a merger having a positive effect on consumer surplus.

Breaking corporations up

We will investigate two distinct types of corporate breakups: One involving the separation of parent companies such that each company operates independently, and another where each car is produced by its own unique manufacturer.

In the baseline scenario, the market comprises of 33 different companies. When parent companies are broken up, the market configuration changes to 62 independent companies. In the most extensive breakup scenario, where every car is manufactured by its own distinct producer, the market significantly expands to 1067 companies. First, we examine how this affects the market concentration using the HHI, as depicted below in figure 5.5.

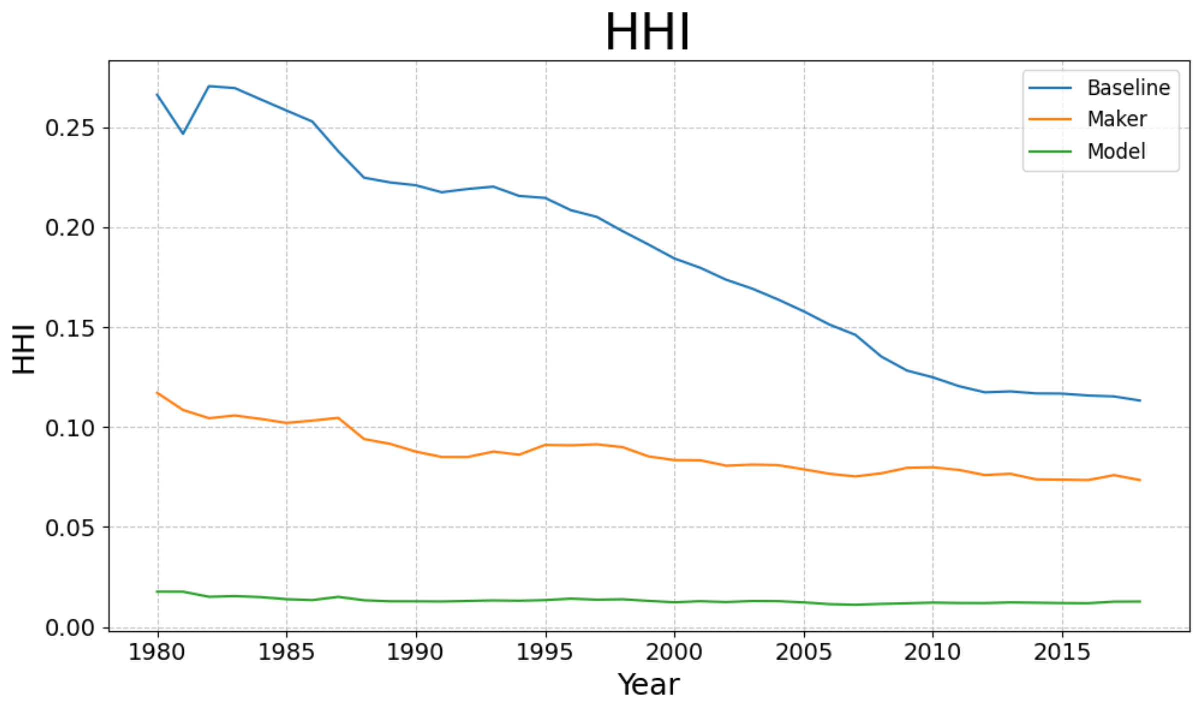

Figure 5.5: Herfindahl-Hirschman Index (HHI) for various breakups. Maker denotes the break up of the parent companies, and Model denotes when each car model is produced by its own company.

As anticipated, the breakups lead to a notable decrease in market concentration, creating a more competitive market. We can now analyze the implications of this increased competition on various market outcomes, similar to our analysis of mergers.

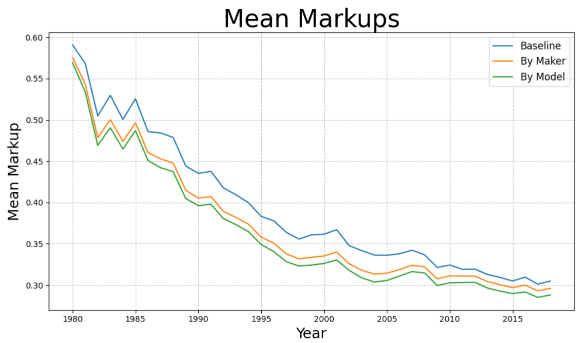

Figure 5.6: Mean Markups

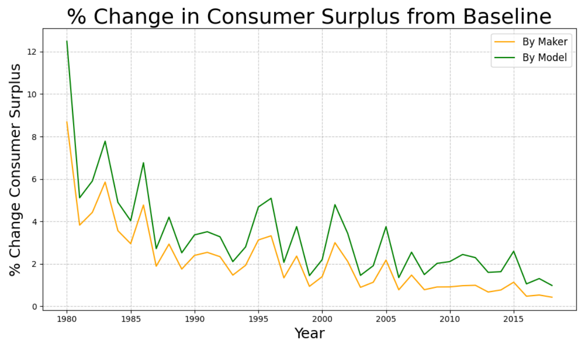

Figure 5.7: Mean change in consumer surplus

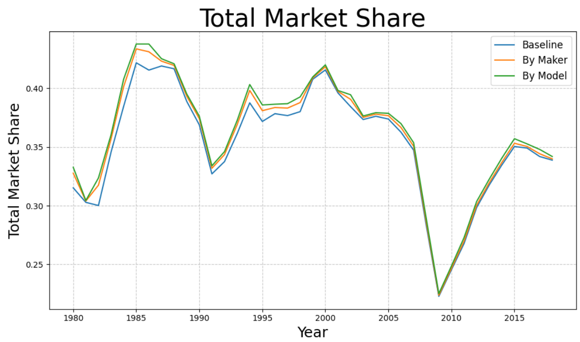

Figure 5.8: Total market shares

The results of the corporate breakups show trends that are the inverse of those observed in mergers. The breakup of large firms into smaller, independent entities typically leads to increased competition. This increased competition compels firms to lower their prices to remain competitive, resulting in a decrease in markups (Figure 5.6). However, our model still estimates that if each car model is produced by its own manufacturer, these producers would still make a profit. This suggests that while competition may be important for profit, other factors could also play a significant role. Further examination is needed to draw definitive conclusions on this matter.

The reduced prices result in an increase in consumer surplus, as depicted in figure 5.7. Notably, the consumer surplus increase is substantial, reaching up to 12% in 1980. However, in more recent years, the increase in consumer surplus has been relatively modest, ranging from 1-2%. Therefore, while consumers would have benefited significantly from the breakups in earlier periods, the gains in the current years are less noticeable.

6 Conclusion

By using data from 572,948,272 car purchases in the U.S. in the period from 1980 to 2018 we estimated the demand in a random coefficient logit model using the BLP framework. These estimations gave us the ability to tell how different attributes affect the consumers utility.
In addition to estimating the demand model, we analyzed the supply side under the multi-product Nash-Bertrand framework. Our findings indicate that the median markup decreased from 0.59 in 1980 to 0.29 in 2018. This trend aligns with our competition analysis, as evidenced by the Herfindahl-Hirschman Index, which dropped from 0.26 in 1980 to 0.11 in 2018, reflecting a significant increase in market competitiveness.
During the same period, we observed an increase in consumer surplus, which suggests a positive correlation between increased competition and consumer welfare. The increased competition forces firms to lower their prices and offer better value, which in turn increases the overall satisfaction and economic well-being of consumers in the market.

By leveraging both the demand and supply sides, we conducted counterfactual analyses to evaluate the potential impacts of mergers and breakups in the market. This approach allowed us to simulate and understand the consequences of changes in market dynamics.
Our counterfactual analysis results align with our expectations. In scenarios where companies merge, we observed an increase in markups accompanied by a decrease in consumer surplus. Conversely, when companies were broken up, markups tended to decrease, leading to an increase in consumer surplus.
One of our most significant findings is that increased competition alone does not fully explain the decrease in markup during the period. Even in a scenario where the six largest companies merge, their markup would still fall significantly from 0.67 to 0.40 over the period. This indicates the need for a more detailed analysis to identify other factors that contributed to the decrease in markup and the corresponding increase in consumer welfare during this time.

Regarding the role of policymakers, our research suggests that strict regulations against mergers of large companies, particularly before 2000, were justified. However, mergers in the period post-2000, and especially post-2010, may not have as detrimental an effect on consumers. Therefore, more relaxed legislation could be considered during these later periods. This is especially true if companies can provide compelling evidence that mergers would reduce their marginal costs, thereby increasing their profits while potentially improving overall consumer welfare. Recognizing the limitations of our models and estimations, further research should be conducted to ensure the robustness of our results.

Bibliography

7

Ackerberg, Daniel A and Gregory S Crawford (2009). *Estimating price elasticities in differentiated product demand models with endogenous characteristics*. Tech. rep. working paper.

Berry, Steven, James Levinsohn, and Ariel Pakes (2004). „Differentiated products demand systems from a combination of micro and macro data: The new car market". In: *Journal of political Economy* 112.1, pp. 68–105.

Berry, Steven T (1994). „Estimating discrete-choice models of product differentiation". In: *The RAND Journal of Economics*, pp. 242–262.

Berry, Steven T, Michael Carnall, and Pablo T Spiller (1996). *Airline hubs: costs, markups and the implications of customer heterogeneity*.

Berry, Steven T, James A Levinsohn, and Ariel Pakes (1993). *Automobile prices in market equilibrium: Part I and II*.

Bresnahan, Timothy F (1989). „Empirical studies of industries with market power". In: *Handbook of industrial organization* 2, pp. 1011–1057.

Conlon, Christopher and Jeff Gortmaker (Nov. 2020a). „Best practices for differentiated products demand estimation with PyBLP". In: *The RAND Journal of Economics* 51.

Conlon, Christopher and Jeff Gortmaker (2020b). „Best practices for differentiated products demand estimation with pyblp". In: *The RAND Journal of Economics* 51.4, pp. 1108–1161.

Conlon, Christopher and Jeff Gortmaker (2023). *Incorporating Micro Data into Differentiated Products Demand Estimation with PyBLP*. Tech. rep. National Bureau of Economic Research.

Devos, Erik, Palani-Rajan Kadapakkam, and Srinivasan Krishnamurthy (2009). „How do mergers create value? A comparison of taxes, market power, and efficiency improvements as explanations for synergies". In: *The review of financial studies* 22.3, pp. 1179–1211.

Gandhi, Amit and Jean-François Houde (Oct. 2019). *Measuring Substitution Patterns in Differentiated-Products Industries*. Working Paper 26375. National Bureau of Economic Research.

Grieco, Paul LE, Charles Murry, and Ali Yurukoglu (2023). „The evolution of market power in the us automobile industry". In: *The Quarterly Journal of Economics*, qjad047.

Herfindahl, O.C. (1950). *Concentration in the Steel Industry*. Columbia university.

Knittel, Christopher R. and Konstantinos Metaxoglou (Mar. 2014). „Estimation of Random-Coefficient Demand Models: Two Empiricists' Perspective". In: *The Review of Economics and Statistics* 96.1, pp. 34–59. eprint: `https://direct.mit.edu/rest/article-pdf/96/1/34/1974554/rest\_a\_00394.pdf`.

Magnolfi, Lorenzo, Daniel Quint, Christopher Sullivan, and Sarah Waldfogel (2022). „Differentiated-products Cournot attributes higher markups than Bertrand–Nash". In: *Economics Letters* 219, p. 110804.

Nevo, Aviv (2000a). „A Practitioner's Guide to Estimation of Random-Coefficients Logit Models of Demand". In: *Journal of Economics & Management Strategy* 9.4, pp. 513–548. eprint: `https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1430-9134.2000.00513.x`.

Nevo, Aviv (2000b). „Mergers with differentiated products: The case of the ready-to-eat cereal industry". In: *The RAND Journal of Economics*, pp. 395–421.

Nevo, Aviv (2001). „Measuring market power in the ready-to-eat cereal industry". In: *Econometrica* 69.2, pp. 307–342.

Østli, Vegard, Lasse Fridstrøm, Kjell Werner Johansen, and Yin-Yen Tseng (2017). „A generic discrete choice model of automobile purchase". In: *European Transport Research Review* 9, pp. 1–20.

Petrin, Amil (2002). „Quantifying the benefits of new products: The case of the minivan". In: *Journal of political Economy* 110.4, pp. 705–729.

Rasmusen, Eric Bennett (Dec. 2007). „The BLP Method of Demand Curve Estimation in Industrial Organization". In: *Translated into Japanese in Gendai Keizaigaku 1, mikuro-bunseki, Isao Miura, Tohru Naito, eds., Tokyo: Keiso shobo Publisher, 2006*. Available at SSRN: `https://ssrn.com/abstract=890923`.

Train, Kenneth E. (2009). *Discrete choice methods with simulation*. Cambridge University Press.

8 Appendix

Symbol	Label	Symbol	Label
j	Car	μ	Elasticity
k	Alternative car	s	Estimated market share
J	Set of cars	$\mathcal{S}$	Observed market share
N	Number of cars	p	Price
t	Year	η	Bertrand markup
i	Consumer	h	Group of the nest
U_{ij}	Utility	I	Inclusive value
ϵ	Idiosyncratic preferences	ρ	Nesting parameter
δ	Mean utility	θ	Parameter
α	Price coefficient	σ	Magnitude of taste variation
β	Characteristics coefficient	ν	Individual taste variation
x	Observed car characteristics	g	Criterion function
ξ	Unobserved car characteristic	Z	Instrument
γ	Character coefficients for cost model	F	Firm
c	Cost	Π	Profit
$\mathcal{H}$	Ownership matrix	$\nabla_{p_t}\mathbf{s}_t$	Cross price elasticity matrix
CS	Consumer surplus		