Price Discrimination and Online Sales in the Automobile Industry

Xavier D'Haultfœuille* Isis Durrmeyer†
Jean-François Fournel‡ Alessandro Iaria§

November 7, 2024

Latest version available here.

Abstract

We investigate the welfare consequences of introducing an online distribution channel in the French car market, where most of the sales currently take place in-person through car dealers relying on third-degree price discrimination. We estimate a structural model of demand with unobserved third-degree price discrimination and transportation costs related to visiting car dealers. In a series of counterfactuals, we introduce an online distribution channel in which prices are uniform and consumers benefit from lower transportation costs. When both distribution channels are simultaneously available, competition from the online channel reduces the extent of third-degree price discrimination in the in-person channel. Introducing the online distribution channel leads to higher profits and an increase in aggregate consumer surplus. Despite aggregate surplus gains, the costs and benefits of the online channel are unevenly distributed among consumers, with the less internet-savvy consumers bearing more of the costs and obtaining fewer of the benefits.

*CREST–ENSAE. E-mail: xavier.dhaultfoeuille@ensae.fr.
†Toulouse School of Economics and CEPR. E-mail: isis.durrmeyer@tse-fr.eu.
‡Toulouse School of Economics. E-mail: jean-francois.fournel@tse-fr.eu.
§University of Bristol and CEPR. E-mail: alessandro.iaria@bristol.ac.uk
We would like to thank Gaston Illanes for his helpful comments and suggestions. We would also like to thank the participants of various seminars and conferences. We acknowledge financial support from the European Research Council under grant ERC-2019-STG-852815 "PRIDISP" and the Agence Nationale de Recherche under grant ANR-17-EUR-0010 (Investissements d'Avenir program).

1 Introduction

Recent technological progress has been facilitating online transactions for a wide variety of products and services. Doing business online has several advantages for firms and consumers. On the one hand, firms may gain access to a larger consumer base and may save the costs of establishing and maintaining a dense network of physical stores. Consumers, on the other hand, may benefit from having access to a wider variety of products and services and avoiding the potentially significant transportation costs associated with visiting physical stores for their purchases.

As consumers and firms are getting used to online marketplaces, there is some evidence suggesting that larger and more expensive products, such as cars, will also be traded online. A pioneering example is Tesla, which operates almost exclusively online. Ordering the vehicle, signing the contract, and making the payment are all done through the company's website, and the car is delivered to the buyer's doorsteps at no extra cost provided they live within 354 km (220 miles) of a Tesla distribution center. Along the same lines, Ford's CEO Jim Farley announced in 2022 a plan to move part of the downstream company's activity online, ending the traditional dealership model and selling directly to consumers at a fixed price.[1] Moving sales online and simplifying the transaction process is part of a larger plan to enforce price transparency and improve consumer convenience and overall purchase experience.[2] Other manufacturers are expected to follow suit if the example of Tesla and Ford prove successful.

It is well documented that, very often, consumers obtain discounts over the posted prices when buying a new car in person at a car dealer. Through personal interactions, the car dealer observes the consumer's characteristics, forms an expectation about their preferences, and then offers a discount over posted prices (or valuable advantages like free upgrades or an extended warranty). We interpret the discounts over the posted prices as a form of third-degree price discrimination. In addition, purchasing a car in person entails transportation costs for consumers, typically associated with traveling a certain distance to reach the car dealer and the associated opportunity cost of time.

In this context, a hypothetical online distribution channel introduces a trade-off for consumers. On the one hand, by choosing the online distribution channel, consumers lose a potential discount and are bound to pay the uniform online price. We impose this assumption to match our observation of the car industry's intentions towards online

[1] Source: Phoebe Ward Howard, "Ford CEO Farley says electric vehicles will be sold 100% online, have non-negotiable price", Detroit Free Press. The full article is available here.

[2] Additional information can be found on Ford's website, see here.

sales.[3] On the other hand, concluding the transaction entirely online entails a reduction in transportation costs since individuals do not have to visit the car dealer to take possession of their new vehicle as it is delivered to their doorsteps.

Using French data for the years 2009–2021, a period in which online car sales were essentially absent, we estimate an equilibrium model of car pricing and sales. We explicitly account for the locations of car dealers and consumers, and their equilibrium effects on firms' pricing and consumers' purchasing decisions. To achieve this, we rely on a novel dataset of car dealer locations and consumers' driving distances in France. We combine it with car registrations, by age and municipality, and various municipal-level demographics. We categorize consumers in broad demographic groups, based on their age and the median income in their municipality of residence. We assume that these groups are observable by car dealers, who then use this information (along with personal interactions) to engage in third-degree price discrimination.

As in D'Haultfœuille et al. (2019), our model enables us to estimate unobserved price discounts for groups of consumers based on demographic characteristics. This method extends the standard demand estimation approach developed by Berry et al. (1995) to account for unobserved price discrimination. We extend D'Haultfœuille et al. (2019) to infer the extent of third-degree price discrimination in the presence of transportation costs. We model transportation costs similarly to Nurski and Verboven (2016), who assume that they are a function of the driving distance between consumers and the closest car dealer of each car model. We take advantage of the granularity of the data to augment the standard demand- and supply-side moments with micro moments that match observed average distances with the corresponding model predictions for each demographic group. There are two key advantages to using these additional micro moments. First, we expect these micro moments to be informative about transportation costs. Second, thanks to these micro moments, our estimation method is robust to the potential endogeneity of distance without requiring additional instrumental variables. Distance could be endogenous if, for example, firms took unobserved components of preferences into account at the time of opening new car dealers. In fact, our estimation method is valid under any assumption regarding the game of entry underlying the

[3] Firms could in principle price discriminate online—for example, by leveraging consumers' browsing histories to "guess" their preferences and their price sensitivity, see Shiller (2020). This behavior does not match our observation of the French car market—all of the car manufacturers that offer an "Order online" option, as of 2024, sell at the uniform price advertised on the website. This makes sense as one advantage of selling online is the possibility to completely streamline and automate the transaction process, which limits human interactions, reduce the scope for price discrimination, and augments the consumer's experience. Nevertheless, our methodology is flexible enough to encompass the case where firms price discriminate both online and offline, see Appendix C.

observed network of car dealers.

Given our estimates of preference parameters and transportation costs, we simulate the introduction of an online distribution channel and study its interactions with brick-and-mortar car dealers. In counterfactual experiments, we assume that firms charge a uniform price online while still offering discounts for in-person transactions. We also assume that consumers face reduced transportation costs when buying online. Since car dealers are important for after-sale services, maintenance, and repairs, we believe that transportation costs may still matter when shopping online, but to a lesser extent than for in-person transactions. We consider various levels of transportation cost reductions in the counterfactual experiments. In the most extreme case, transportation costs are eliminated, implicitly assuming that consumers do not expect future interactions with car dealers. Throughout the analysis, we account for consumer heterogeneity in their propensity to shop online relying on a survey of attitudes toward online purchases by demographic group.

Our results can be summarized by two key findings. Our first finding is that introducing online sales at a uniform price reduces price dispersion in the offline channel. When all consumers are unrestricted to use either channel, it is optimal for firms to set a uniform price in both channels. This makes the online channel unambiguously better for all consumers, since transportation costs are lower online. We observe a large transfer from the in-person to the online channel and a market expansion as a result. When, instead, some consumers are restricted in their ability to shop online, the results are different. We observe two forces at play. On the one hand, firms want to set discriminatory prices offline to extract more surplus from captive consumers. On the other hand, the competitive pressure from the online channel leads firms to set a uniform price for consumers who have access to both channels. When the online channel provides a small reduction in transportation costs, the second effect dominates and firms set an almost uniform price in both channels. In contrast, when the transportation cost reduction is large, the online channel becomes overwhelmingly better for unrestricted consumers, and firms continue to price discriminate in the offline channel. In some sense, the level of transportation cost reduction accruing from shopping online dictates firms' ability to separate the market between captive consumers who always shop in person and non-captive consumers who can use both channels.

Our second finding relates to the welfare effects of introducing an online channel. We find that price discrimination, taken in isolation, is beneficial only to some consumers. The aggregate effect on consumers is a small loss in surplus, and the aggregate effect on profits is a small increase in industry profits, in the range of one percent. Transportation costs are detrimental to all consumers. We find that reducing transportation costs by 25% generates welfare effects in the same range as those obtained by eliminating price discrimination. Eliminating entirely transportation costs yields a much larger effect, roughly by one order of magnitude, meaning that price discrimination is a second order effect compared to transportation costs. For this reason, online transactions should be very attractive for consumers and firms. However, we find that some groups of consumers experience a decrease in surplus when online sales are introduced. These consumers are characterized by high price sensitivity and a low sensitivity to transportation costs. Without online sales, they obtain large discounts over posted prices. When both sales channels are available, firms offer significantly smaller discounts to these consumers, and the reduction in transportation costs they get for shopping online is not important enough to offset the increase in prices.

A limitation of this work is that we maintain the dealership network fixed when simulating the introduction of the online distribution channel. In this sense, our main results should be interpreted as a collection of short-run responses to the introduction of the online distribution channel. To provide some insight about possible long-run responses, we consider an additional experiment in which the 10% lowest performing car dealers of each brand exit the market. We find that our main results overstate the gains in consumer surplus and profits by 4% and 9%, respectively, compared to the case with exit of car dealers.

Related literature. Our research contributes to several branches of the literature. First, it relates to a growing literature on the welfare effects of e-commerce, such as Brown and Goolsbee (2002) on the impact of comparison websites on insurance prices in the US and Morton et al. (2001) on car referral websites (a precursor of online sales in the car industry). Similarly to Pozzi (2013), Fan et al. (2018), and Forman et al. (2009), our study shows that the coexistence of in-person and online distribution channels can generate welfare gains through both increased price competition and reductions in transportation costs. Along the lines of Huang and Bronnenberg (2023) and Brynjolfsson et al. (2003), our analysis also illustrates that these forces are closely related to the amount of product variety available to consumers. Our paper contributes to this literature with novel empirical evidence on the ways in which an online distribution channel can limit the ability to price discriminate of brick-and-mortar stores.

Second, we contribute to the literature on price discrimination and price dispersion in retail markets. Seminal works by Corts (1998) and Thisse and Vives (1988) and recent work by Iaria and Wang (2021) and Rhodes and Zhou (2024) provide evidence that price discrimination can intensify competition, benefit consumers (in the aggregate), and decrease profits in oligopolistic industries. Previous works on the car industry have studied empirically price discrimination based on consumer demographics, see Ayres and Siegelman (1995), Goldberg (1996), Harless and Hoffer (2002), and Chandra et al. (2017). These works find contrasting evidence linking price dispersion to demographics (typically gender and race). We provide novel empirical evidence on the relationship among transaction prices, income, and age (we find no relationship with gender), and more broadly on the relationship between price dispersion and spatial differentiation (for related evidence on a homogeneous product, see Miller and Osborne, 2014).

Third, our paper is closely related to recent work studying price discrimination through the lens of structural models, such as D'Haultfœuille et al. (2019) on the French car industry and Sagl (2024) on the trucking industry in Texas. Sagl (2024) finds that most of the observed price dispersion can be explained by consumer unobservables (or soft information), as opposed to demographics. His analysis leverages consumer-level transaction prices and repeated purchases over time. D'Haultfœuille et al. (2019) instead rely on list prices and recover unobserved transaction prices resulting from third-degree price discrimination based on consumer demographics. We contribute to this literature by proposing a unified framework that incorporates (potentially unobserved) transaction prices and spatial differentiation, and that can be used to investigate the relationship between the two in oligopolistic industries with differentiated products.

Fourth, our work relates to recent papers studying price personalization in online markets. Shiller (2020) studies price personalization for Netflix subscription plans based on consumers' browsing histories, while Dubé and Misra (2023) study price personalization for a digital firm based on detailed observable consumer characteristics. These studies leverage the vast amount of information available online to investigate the consequences of price personalization for a firm. Instead, we focus on the consequences of online sales for an oligopolistic industry with differentiated products and a long tradition of sales through brick-and-mortar stores. Car manufacturers use the internet to enforce price transparency and uniform prices rather than personalization, allowing us to deepen our understanding of oligopolistic pricing behavior when both an online and offline distribution channels coexist.

Fifth, we contribute to the literature investigating the role of dealer networks in the car industry. A growing strand of this literature takes the perspective of costly search (e.g., Moraga-González et al., 2023; Murry and Zhou, 2020; Yavorsky et al., 2021), where consumers need to personally visit car dealers to learn about some of the features of car models (or about their very existence), essentially adding them to their consideration sets, and where search costs depend on the distance to car dealers. For the separate identification of search from utility (necessary for counterfactual analyses), the implementation of these structural models typically rules out price discrimination, especially in cases such as ours in which consumer-level transaction prices are not observed (Moraga-González et al., 2023). As the main objective of this paper is to investigate the relationship between price discrimination and spatial differentiation, we instead follow the route of augmenting the structural model of unobserved price discrimination by D'Haultfœuille et al. (2019) with transportation costs, leaving the important question of also incorporating a search dimension into the framework for future research (for evidence on the relevance of each of these dimensions, see Scott Morton et al., 2011).

Finally, our work is also closely related to Duch-Brown et al. (2023), who study the interaction between online and offline sales in the portable PC industry in Europe. In their application, price dispersion occurs in the online market as a result of geoblocking restrictions on cross-border transactions (self-imposed by firms). They show that banning these restrictions results in the convergence of prices to a unique European-level price for each product sold online. In the context of the French car industry, we uncover a similar mechanism: the introduction of an online distribution channel with uniform prices leads to an attenuation in price dispersion for in-person transactions and a convergence towards the online uniform prices. We contribute to Duch-Brown et al. (2023) by specifically investigating the roles of spatial differentiation and transportation costs in the transition to market integration promoted by an online distribution channel.

The rest of the paper is organized as follows. Section 2 discusses the model and estimation. Section 3 discusses the data and presents estimation results. Section 4 presents the augmented model with online dealers. The main results from the counterfactual analysis are presented in Section 5. Section 6 provides concluding remarks.

2 Model

2.1 Demand

We incorporate transportation costs in the model of (unobserved) third-degree price discrimination by D'Haultfœuille et al. (2019). Throughout, we assume that consumers belong to one of D mutually exclusive groups based on their observed demographics, and that firms price discriminate by offering different prices to consumers from different groups. We enrich the model by explicitly considering that consumers are spatially distributed across municipalities and face heterogeneous distances to car dealers. This implies heterogeneous transportation costs when purchasing a car of a given brand.

Consider consumer i belonging to demographic group d (e.g., a profile of age and income) and living in municipality m (e.g., a town in France). We omit the time subscript for simplicity. Their indirect utility from purchasing car model $j = 1, ..., J$ is

$$U_{ijdm} = \underbrace{X_j'\beta_d + \alpha_d p_{jd} + \xi_{jd}}_{\delta_{jd}} + \underbrace{\gamma_d \text{dist}_{jm}}_{\mu_{jdm}} + \epsilon_{ijdm}, \tag{1}$$

where X_j is a vector of observed car characteristics that is invariant across groups and municipalities (e.g., horsepower), p_{jd} is the (unobserved) transaction price faced by group d, and ξ_{jd} captures the average indirect utility of the car characteristics unobserved by the econometrician. Note that both p_{jd} and ξ_{jd} do not vary across municipalities. Variable dist_{jm} is the driving distance from municipality m to the closest dealer selling car model j. Variable ϵ_{ijdm} is an idiosyncratic error term assumed to be distributed extreme value type I. The demand parameters $(\beta_d', \alpha_d, \gamma_d)$ are allowed to be heterogeneous across demographic groups, but they are constant within group.

Indirect utility (1) is similar to that in D'Haultfœuille et al. (2019), with the exception of the additional term $\mu_{jdm} \equiv \gamma_d \text{dist}_{jm}$, representing the transportation cost for consumers of group d living in municipality m traveling to the closest car dealer selling j. Conditional on distances from the closest car dealers of all car models, the probability that a consumer in group d and municipality m purchases car model j is

$$s_{jdm}(\text{dist}_{1m}, ..., \text{dist}_{Jm}) = \frac{\exp(\delta_{jd} + \gamma_d \text{dist}_{jm})}{1 + \sum_{k=1}^{J} \exp(\delta_{kd} + \gamma_d \text{dist}_{km})}. \tag{2}$$

As in D'Haultfœuille et al. (2019), we assume that each transaction price p_{jd} is chosen at the national level (more details on this below). To obtain the national-level market share of car model j for group d, we integrate (2) over municipalities,

$$s_{jd}(\delta_d, \mu_d) = \sum_{m \in \mathcal{M}} w_{dm} \cdot s_{jdm}(\text{dist}_{1m}, ..., \text{dist}_{Jm}), \tag{3}$$

where $\mathcal{M}$ collects all municipalities, $\delta_d = (\delta_{1d}, ..., \delta_{Jd})$, and $\mu_d = (\mu_{1dm}, ..., \mu_{Jdm})_{m \in \mathcal{M}}$. The variable $w_{dm} = M_{dm}/M_d$ is the weight observed for the specific municipality of the group, where M_{dm} and M_d are the populations of the group d in the municipality m and throughout the country, respectively.

2.2 Supply

We consider a Bertrand-Nash price-setting game in which firms are able to implement third-degree price discrimination by setting different prices for each demographic group $d = 1, ..., D$. Each firm $f = 1, ..., F$ selects a menu of prices $p_j = (p_{j1}, ..., p_{jD})$ for each car model j they sell by maximizing the national-level profit function

$$\pi_f = \sum_{d=1}^{D} \phi_d \sum_{j \in \mathcal{J}_f} s_{jd}(p_d) \cdot (p_{jd} - c_{jd}), \tag{4}$$

where $\mathcal{J}_f$ is the collection of car models sold by firm f, p_d is the vector of all prices for group d, and c_{jd} the marginal cost of car model j for group d. $\phi_d = M_d/M$ is the observed group-specific population weight, where M_d is the national population of group d and M is the national population (over all demographic groups).

The system of J first-order conditions associated with demographic group d are

$$p_d = c_d - \widetilde{\mathcal{D}}_d(p_d, \mathcal{H})^{-1} s_d, \tag{5}$$

where c_d is the vector of group-specific marginal costs of all car models, s_d is the vector of group-specific market shares for all car models, $\widetilde{\mathcal{D}}_d(p_d) = \mathcal{H} \odot \mathcal{D}_d(p_d)$, $\mathcal{H}$ is the ownership matrix, and $\mathcal{D}_d(p_d)$ is the matrix of derivatives of s_d with respect to p_d, with typical element (j, k) equal to $\partial s_{kd}/\partial p_{jd}$.

2.3 Identification and estimation

Identification of the unobserved transaction prices. Compared to standard demand models, the identification of $\theta = (\theta_1, ..., \theta_D)$, where $\theta_d = (\beta_d', \alpha_d, \gamma_d)$, presents the additional complication that the group-specific transaction prices $p(\theta) = (p_1, ..., p_D)$ are not observed by the econometrician. Following D'Haultfœuille et al. (2019), we address this complication by relying on both demand and supply restrictions to jointly identify preference parameters θ and transaction prices $p(\theta)$.[4]

We make the following assumptions, which are sufficient for the validity of the estimator by D'Haultfœuille et al. (2019).

A1. Observability of group-specific market share s_{jd} for each j and d.

A2. Constant marginal costs across demographic groups, $c_{jd} = c_j$ for any j and d.

[4]Note that, even if transaction prices were observed, the prices that consumers face for the car models they do not purchase would remain unobserved.

A3. Observability of list prices. For each car model j, we observe the list price $\bar{p}_j = \max\{p_{j1}, ..., p_{jD}\}$, which is the highest transaction price any demographic group can be asked to pay at any car dealer.

Intuitively, assumptions A1-A3 allow us to reconstruct, for any given value of the demand parameters θ, the transaction prices $p(\theta)$ that rationalize both demand and supply. With these, we are back to a standard model in which all prices are observed and we can identify and estimate θ following the methodology of Berry et al. (1995).

First, given assumption A1 and following Berry (1994), for a given group d and some value of γ_d, we obtain $\delta_d(\gamma_d)$ by inverting the system of J market share equations given by (3) with $j = 1, ..., J$. Second, we obtain the transaction prices corresponding to θ. To this end, note that the first-order conditions (5) and assumptions A2-A3 imply

$$\bar{p}_j = c_j - \min\left\{\left[\widetilde{\mathcal{D}}_1{}^{-1} s_1\right]_j, ..., \left[\widetilde{\mathcal{D}}_D{}^{-1} s_D\right]_j\right\}, \tag{6}$$

with $[x]_j$ denoting the j-th element of vector x. Then, for given value of θ, each transaction price can be obtained as

$$p_{jd}(\theta) = \bar{p}_j + \min\left\{\left[\widetilde{\mathcal{D}}_1^{-1} s_1\right]_j, ..., \left[\widetilde{\mathcal{D}}_D^{-1} s_D\right]_j\right\} - \left[\widetilde{\mathcal{D}}_d^{-1} s_d\right]_j,$$

since one can check that the matrix of derivatives $\widetilde{\mathcal{D}}_d$ does not depend on transaction prices $p_{d'}$ from groups $d' \neq d$.

Demand-side moments. We compute the empirical counterpart of the following moment conditions

$$\mathbb{E}\left[Z'_{jd}\xi_{jd}\right] = 0, \quad d = 1, ..., D, \tag{7}$$

with Z_{jd} a group-specific vector of instruments. To do this, we first compute $\xi_{jd}(\theta) = \delta_{jd}(\gamma_d) - X'_j\beta_d - \alpha_d p_{jd}(\theta)$ and then consider the empirical moment condition $g_1(\theta) = (g_{11}(\theta), ..., g_{1D}(\theta))$, where

$$g_{1d}(\theta) = \frac{1}{J}\sum_{j=1}^{J} Z'_{jd}\xi_{jd}(\theta). \tag{8}$$

Moment conditions (7) hold for a vector of valid instruments Z_{jd}. As usual, while transaction prices are endogenous by construction, we assume the observed characteristics X_j to be exogenous. Valid instruments (in addition to X_j) can then be obtained as functions of the exogenous characteristics of other car models but j, i.e., X_k for $k \neq j$.

Micro moments. We complement (7) with micro moments that help identify the distance coefficient γ_d. We take advantage of the fact that we observe market shares at the demographic group-by-municipality level. We construct micro moments by matching the observed and predicted average distance between consumers and car dealers of the purchased car models. In particular, we specify $g_3(\gamma) = (g_{31}(\gamma_1), ..., g_{3D}(\gamma_D))$, with

$$g_{3d}(\gamma_d) = \frac{\sum_{m\in\mathcal{M}}\sum_{j=1}^{J} M_{dm}\cdot s_{jdm}(\gamma_d)\cdot \text{dist}_{jm}}{\sum_{m\in\mathcal{M}}\sum_{j=1}^{J} M_{dm}\cdot s_{jdm}(\gamma_d)} - \frac{\sum_{m\in\mathcal{M}}\sum_{j=1}^{J} M_{dm}\cdot s_{jdm}\cdot \text{dist}_{jm}}{\sum_{m\in\mathcal{M}}\sum_{j=1}^{J} M_{dm}\cdot s_{jdm}}, \quad (9)$$

where s_{jdm} is the observed market share for product j for group d in municipality m. The corresponding market share predicted by the model given γ_d instead is

$$s_{jdm}(\gamma_d) = \frac{\exp(\delta_{jd}(\gamma_d) + \gamma_d \text{dist}_{jm})}{1 + \sum_{k=1}^{J}\exp(\delta_{kd}(\gamma_d) + \gamma_d \text{dist}_{km})},$$

where we obtain $(\delta_{kd}(\gamma_d))_{k=1,...,J}$ by inverting the system of J market shares. Note that by construction $s_{jd}(\gamma_d) = s_{jd}$, but in general $s_{jdm}(\gamma_d) \neq s_{jdm}$.

Three points are worth noting. First, if we only have one instrument for prices in (7), we lack a moment condition to identify γ_d. Then, additional moments, such as the micro moments we propose, are necessary to identify γ_d. Generally, instruments Z_j may lack the power to estimate with sufficient precision γ_d. Instead, we expect the above micro moments to be informative about γ_d since, intuitively, the observed average distance between consumers and the car dealers of their purchased car models decreases with γ_d. In Section 3.3, we use our data to illustrate the monotonicity of $g_{3d}(\gamma_d)$ in γ_d.

Second, these micro moments remain valid even if distance dist_{jm} is endogenous. By this, we mean that distances could be correlated with ξ_{jd}: for example, if firms partially or fully observed $(\xi_{jd})_{j,d}$ at the moment of deciding where to locate their car dealers. To see this, note that at the true value of γ_d, say γ_d^0, $\delta_{jd}(\gamma_d)$ will also be equal to its true value $\delta_{jd}(\gamma_d^0) = \delta_{jd}$, and thus $s_{jdm}(\gamma_d^0) = s_{jdm}$. This only follows from Berry (1994)'s demand inverse and therefore holds irrespective of any dependence between the distances and the unobserved terms $(\xi_{jd})_{j,d}$.

Third, assuming that s_{jd} is measured without error is reasonable (and standard) given the large number of consumers in each demographic group throughout the country. However, the assumption that we also perfectly observe the municipality-level market shares s_{jdm} may be strong. We only observe a proportion on a finite sample instead of the true purchase probability (say, s^*_{jdm}), and the corresponding sample is small for small municipalities. However, we still have

$$\mathbb{E}\left[\sum_{m\in\mathcal{M}}\sum_{j=1}^{J} M_{dm}\cdot s_{jdm}\cdot \text{dist}_{jm}\middle|(\text{dist}_{jm})_{j,m},(\xi_{jd})_{j,d}\right]=\sum_{m\in\mathcal{M}}\sum_{j=1}^{J} M_{dm}\cdot s^*_{jdm}\cdot \text{dist}_{jm}.$$

Hence, even if s_{jdm} is measured with error (but maintaining, as usual, that the $(s_{jd})_{j=1,...,J}$ are measured without error), we still have that at the true value γ_d^0, $\mathbb{E}[g_{3d}(\gamma_d^0)]=0$. In other words, these micro moments are robust to this concern about measurement error. For the same reason, with these micro moments, "zeros" in the observed market shares at the municipality level do not raise any concern.

Supply-side moments. We assume that the marginal costs c_j can be computed as a linear combination of observed car characteristics X_j, cost shifters W_j, and an unobserved cost shock ω_j, such that

$$c_j = X_j'\lambda_1 + W_j'\lambda_2 + \omega_j. \tag{10}$$

We compute the empirical counterpart of the following moment conditions

$$\mathbb{E}\left[Z_{jS}'\omega_j\right]=0, \tag{11}$$

with Z_{jS} a vector of supply-side instruments. The associate supply-side moment conditions are

$$g_2(\theta)=\frac{1}{J}\sum_{j=1}^{J} Z_{jS}'\omega_j(\theta)=0. \tag{12}$$

Again, valid instruments can be obtained as functions of the exogenous characteristics and cost shifters of other products. Note that including cost shifters to the supply side (i.e., excluded from the demand side) is crucial for identifying the price sensitivities.

Estimation. The model is estimated by generalized method of moments, following Berry et al. (1995) and subsequent best practices. We stack the moment conditions $g(\theta)=(g_1(\theta)', g_2(\theta)', g_3(\gamma)')'$ and construct the following estimator

$$\widehat{\theta}=\arg\min_{\theta} \quad g(\theta)'\mathbf{W}g(\theta) \tag{13}$$

where $\mathbf{W}=\text{diag}(\mathbf{W}_{11},...,\mathbf{W}_{1D},\mathbf{W}_2,\mathbf{W}_3)$ is a block-diagonal symmetric weighting matrix. The linear parameters $(\beta_1',...,\beta_D',\lambda_1',\lambda_2')$ are concentrated out of the estimation.

Additional computational details are reported in Section B.2.

3 Data and estimation results

3.1 Demographic groups definition

We divide consumers into three age categories (below 40, 40 to 59, and above 60) and two income categories (low and high income) to form six demographic groups. These groups are easily observable by car dealers and presumably associated with heterogeneous preferences, thus forming a basis for price discrimination. We do not observe consumers' income directly in the car registration data, so we assign an income category based on their municipality of residence and age. More specifically, within each age category, we divide municipalities evenly into low- and high-income classes based on their median income. As a consequence, consumers living in the same municipality and of the same age are assigned to the same group. However, a given municipality could be considered high-income in one age class and low-income in another.

To fully characterize the potential car buyers, we assemble a rich dataset of consumer-level demographics by demographic group. All are obtained from the National Institute of Statistics and Economic Studies (INSEE). Our data includes yearly population censuses, income data by age category, and a survey of consumers' attitudes towards online purchases.[5] This survey collects information on a representative sample of French individuals about their private use of information and communication technologies, including their usage of online sales platforms. The provided individual-level demographics allow us to link each survey entry to our demographic groups. We note that consumers are highly heterogeneous in their propensity to use online sales platforms along the age and income dimensions. We refer to Table A.1 for a summary of these consumer demographics by group.

3.2 Evidence of price dispersion

We want to provide some preliminary evidence of price dispersion with respect to our chosen demographic groups, based on income and age. In theory, car dealers could price discriminate based other observed and unobserved characteristics of buyers beyond their age and their income level. To that end, we combine two waves of a French survey of consumers expenditures, which contain both demographic characteristics of consumers

[5] Source: "lil-1407 : Technologies de l'information et de la communication auprès des ménages (TIC) - 2019 (2019, INSEE)", accessed from Progedo Adisp.

and the transaction price of their most recent car purchase.[6]

In these surveys, car purchases are split into new and second hand vehicles, and we can distinguish sales that occurred at a dealership versus sales that occurred between consumers. Whenever a consumer resold his old car in the same year, the trade-in value is also recorded. We correlate the transaction price paid by consumers who purchase directly from a dealership with demographics characteristics that are available in our main data set. The results are presented in Table 1. We focus on two different measures for the transaction price: the transaction price paid by the consumer and the transaction price net of the buyback value. Finally, we include a rich set of fixed effects to control for product characteristics, and seasonality.

Our finding indicates that income positively correlates with the transaction price. Since our specification includes model by engine fixed effects, this means that high income consumers pay more on average for the same car model and engine type. We cannot rule out that the effect is partly driven by the choice of additional options as these are unobserved to us. It is not clear in which direction this bias could go. On the one hand, the price dispersion could be explained by the fact that richer individuals purchase vehicles with more expensive options. On the other hand, price dispersion could be under-evaluated if car dealer provides options at no cost to low income consumers and both groups buy similar vehicles with similar options.

Our results suggests that age also correlates positively with the transaction price, although the effect is not significant if we focus only on new vehicles. We find an insignificant effect for other demographic characteristics, namely gender, household size, and the level of urbanity in the consumer's municipality of residence.[7] The fact that gender is insignificant is not surprising: in most cases, purchasing a car is a decision that is taken at the level of the household and both partners could have shopped for the car together. In this case, transaction prices are not expected to correlate with the gender of the car owner.

Now that we have established that income and age correlate positively with the transaction price, we use the available demographics in the expenditure survey to reconstruct our chosen demographic groups. We split consumers by age (three group, defined as above) and income (wages income above or below the median by age group) and correlate demographic group dummies with the transaction price. Again, we control for

[6]Source: "Enquête Budget des Familles (BDC) - 2011–2017 (2011–2017, INSEE)".

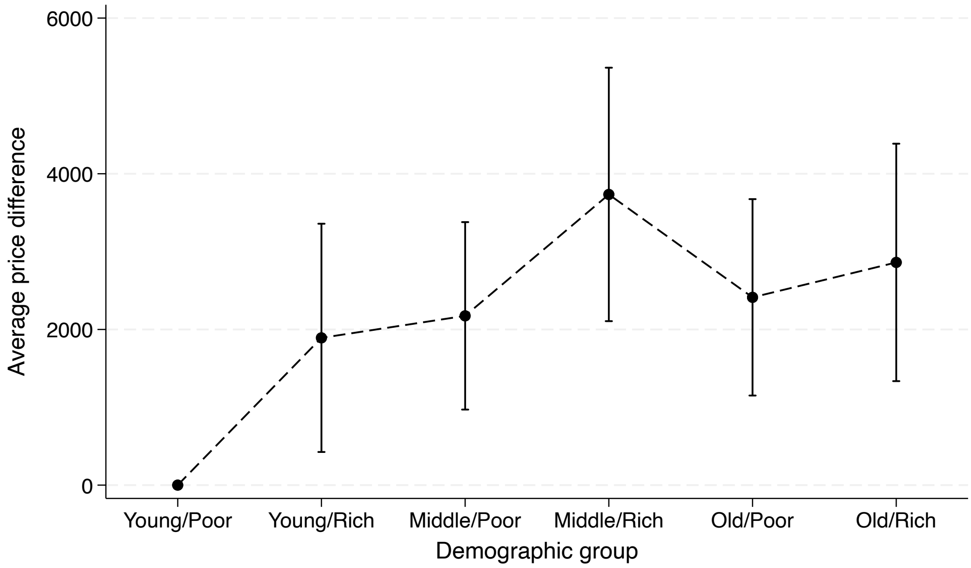
[7]All regressions also include indicator variables for the country of origin (omitted), also mostly insignificant.

Table 1: Evidence of price dispersion

	Transaction price		Transaction price - buyback value	
	(1)	(2)	(3)	(4)
Income	40.497***	25.916**	39.253***	28.970**
	(13.112)	(11.958)	(13.760)	(14.533)
Age	67.387***	20.156	47.276***	2.754
	(15.828)	(19.359)	(17.284)	(25.626)
Female	851.802	141.661	1,051.810	604.796
	(1,078.583)	(1,463.399)	(1,147.382)	(1,893.736)
Age × female	-17.770	-5.874	-25.134	-13.574
	(19.998)	(26.556)	(21.727)	(32.152)
Value of down payment	0.006	0.005	0.013	0.011**
	(0.005)	(0.003)	(0.008)	(0.005)
Household: 2 pers.	-182.571	-250.977	-478.245	-707.159
	(393.500)	(633.025)	(500.774)	(815.847)
Household: 3 pers.	-688.958	-786.539	-954.493	-1,117.172
	(620.113)	(925.880)	(709.379)	(1,208.438)
Household: 4 pers.	-644.082	-1,357.238*	-1,464.877**	-2,485.373**
	(626.979)	(802.681)	(655.590)	(1,044.736)
Household: 5 pers.	-3,000.296***	-2,397.637**	-2,987.593***	-4,892.123***
	(882.021)	(1,109.743)	(962.957)	(1,589.441)
Household: 6+ pers.	-202.262	-929.100	1,490.901	-399.893
	(2,294.165)	(2,114.791)	(2,113.088)	(2,009.281)
Urban area: less than 15,000	-825.524	2,275.435	-1,235.378	2,558.358
	(1,425.966)	(1,552.297)	(2,290.406)	(3,842.866)
Urban area: 15,000–24,999	345.570	548.107	1,610.750	2,790.948
	(1,717.314)	(1,827.368)	(1,458.979)	(2,818.982)
Urban area: 25,000–34,999	-1,588.274	1,024.648	-1,243.062	1,283.947
	(1,377.100)	(2,077.463)	(1,659.565)	(3,108.214)
Urban area: 35,000–49,999	-1,733.160	-6.787	-1,418.662	1,805.762
	(1,095.743)	(1,002.154)	(1,220.085)	(1,598.266)
Urban area: 50,000–99,999	-1,316.561	-462.050	-1,999.194**	-1,382.434
	(815.740)	(1,136.611)	(785.564)	(1,128.750)
Urban area: 100,000–199,999	-822.825	195.033	-198.663	222.065
	(791.007)	(930.849)	(714.154)	(1,036.763)
Urban area: 200,000–499,999	-1,093.716	182.223	-864.898	-6.343
	(697.664)	(876.896)	(608.700)	(1,023.145)
Urban area: 500,000 or more	-1,143.714*	-366.690	-795.605	197.814
	(657.948)	(888.946)	(643.131)	(882.634)
Urban area: Paris greater metro area	-900.305	-523.017	-121.970	189.700
	(731.188)	(1,140.885)	(676.618)	(990.022)
New vehicles only	No	Yes	No	Yes
Fixed effects				
Car model × engine ×new	Yes	Yes	Yes	Yes
Year × month of purchase	Yes	Yes	Yes	Yes
Country of origin of buyer	Yes	Yes	Yes	Yes
Observations	1,283	698	1,283	698
R-squared	0.742	0.795	0.600	0.617

Notes: This Table represents the result of a regression of transaction prices on demographic characteristics of buyers, based on a survey of consumers' expenditure. We have excluded observations where the car was purchased following an insurance claim (i.e., the replacement of a damaged vehicle). Columns (1), (3), and (5) include sales of both new and used cars, purchased at a car dealer. Columns (2) and (4) include only new cars. the buyback value represents the payment that was received by the consumer for trading in his old car. The F-statistic tests for the hypothesis that the coefficients on the group dummies are jointly zero. Standard errors in parenthesis are clustered at the car model × engine × new/used level. Significance: * $<$ 0.10, ** $<$ 0.05, *** $<$ 0.01.

Figure 1: Evidence of price dispersion among demographic groups



Notes: This Figure represents the result of a regression of transaction prices on demographic group dummies, based on a survey of consumers' expenditure (see Table A.2, column (1)). We have excluded observations where the car was purchased following an insurance claim (i.e., the replacement of a damaged vehicle). Includes demographic characteristics (gender, household size, urbanity), car model × engine × new/used and year × month fixed effects. The brackets represent the 95% confidence interval, clustered at the car model × engine × new/used level.

gender, household size, the level of urbanity, and a rich set of fixed effects.

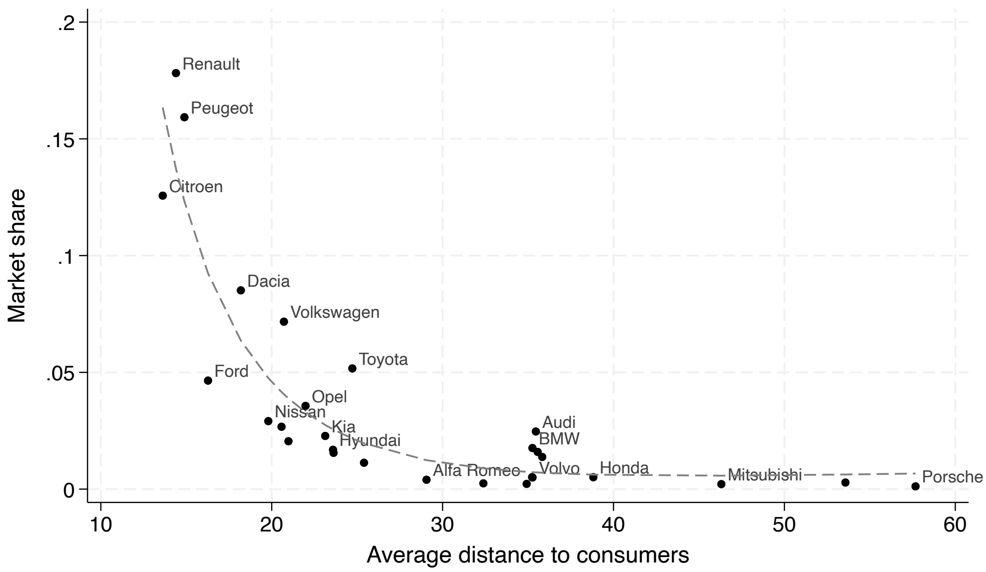
We present the coefficients associated to the demographic groups dummies in Figure 1. We relay the detailed regression results to the Appendix, Table A.2. Our results reveal the presence of price dispersion at the level of our chosen demographic groups. We find that middle aged consumers with high income face the highest transaction prices, followed by old consumers with high income. Young consumers, and poor consumers pay on average less for the same car.

3.3 Data and descriptive statistics

Dealership locations. We construct a novel dataset of car dealer locations in France. The data was obtained from AutoConcession and VendiAuto in early 2020.[8] Both are online registry of car dealers in France. Each entry in our dealership dataset describes a unique car dealer. It contains the name of the dealership, its associated brands, and its address. The data include 4,649 dealer-brand combinations.

[8]Sources: `https://www.auto-concession.fr/` and `https://www.vendiauto.fr/`.

Figure 2: Market presence by brand and dealer proximity



Notes: This figure represents the relationship between brands' total market share and their market presence. The market share of each brand is computed as the ratio of its total sales to the total sales of all brands between 2009 and 2021. Market presence is represented as the average distance to consumers over the same period. Lexus is excluded from the graph (average distance to consumers = 111 km, market share < 0.001). The dashed line represents the fitted values of a fractional polynomial regression.

We augment our car dealer dataset with the driving distance from each car dealer to each municipality in Metropolitan France. Distances are computed from 969,455 queries on TomTom's API.[9] To limit the number of required queries on TomTom's API, we consider each car dealer to be located at the centroid of their zipcode, and consumers to be located at the centroid of their municipality of residence. Zipcodes and municipalities are small geographical units in France.

We provide an overview of the market presence of each firm in Figure 2. We plot the aggregate market share of each brand against the average distance to consumers.[10] Brands with a large share of the market are typically located closer to consumers (i.e., they operate at more locations which improve proximity to potential buyers). For completeness, we provide additional information on the market presence of firms in Table A.3

Car registrations. We obtain a dataset of all new car registrations in France, between 2009 and 2021, from AAA Data.[11] The data is aggregated at the municipal-level and by

[9]See https://developer.tomtom.com/.
[10]The graph is similar if we instead plot the market share against the number of car dealers.
[11]Source: https://www.aaa-data.fr/.

age classes (in 5 years increments). Municipalities in France correspond to a very fine geographical unit. There are on average 1,350 inhabitants per municipality and there are 35,296 municipalities in Metropolitan France (mainland European France).

Within municipality and age group, sales are recorded at the level of the brand (29 brands), model (372 models), engine type (gas, diesel, electric, plug-in, hybrid), and trim (sedan, convertible, station wagon). It includes common car attributes such as horsepower, weight, CO_2 emissions, and fuel consumption, as well as the list price. These car attributes are collected by AAA Data from car manufacturer catalogues. We complement the dataset with annual average fuel and electricity prices to construct a measure of driving cost (in euros per 100 km). Finally, we recover the market segment (e.g., subcompact, compact, SUV, etc.) of each model from Jato Dynamics.[12]

We define a product to be a combination of a brand, a model, an engine type, and a trim. After aggregating by product, demographic group, and year, the final dataset includes 4,975 observations over 13 years. Whenever the data are available at a more disaggregated level than our product definition, we keep the characteristics of the most frequently purchased option. List prices are adjusted to be net of fees and rebates tied to the French Feebate Program.[13] Both the net list prices and the driving costs are deflated to 2018 euros. We encountered several missing observations in key car characteristics (namely, horsepower for electric vehicles). Whenever this happened, we filled in the missing value with additional data from the French National System of Vehicle Registration (SIV).

Descriptive statistics on car sold are presented in Table 2. In the first panel, we offer a breakdown of sales by product and demographic group. Group 4 and 6, representing high income consumers, aged 40 years old and above, purchase on average more than twice the number of vehicles than other groups.

The second panel reports statistics related to how far car dealers are located from consumers. Distances are computed as the driving distance from the centroid of the municipality of consumers to the centroid of the zipcode of the closest dealer of each brand. These statistics are not weighted by group-specific sales. Instead, we weight them by brand importance (i.e., total sales of each brand) and group-specific municipal-level populations. We do this to preserve comparability across groups. One notable observation is that consumers belonging to group 5 (low income, aged 60 and above) live significantly

[12]Source: https://www.jato.com.

[13]The French Feebate Program offers incentives to promote low-emission vehicles, based on engine type and tailpipe emissions.

Table 2: Car characteristics

Description	Mean	Std. dev.	10th pct.	Median	90th pct.	Observations
Sales						
Group 1: Age < 40, Inc. $=$ Low	272	635	8	62	710	4,975
Group 2: Age < 40, Inc. $=$ High	252	578	10	64	630	4,975
Group 3: Age $\in [40, 60)$, Inc. $=$ Low	381	821	21	108	934	4,975
Group 4: Age $\in [40, 60)$, Inc. $=$ High	719	1,491	46	215	1,836	4,975
Group 5: Age ≥ 60, Inc. $=$ Low	262	698	9	50	609	4,975
Group 6: Age ≥ 60, Inc. $=$ High	765	1,817	34	179	1,758	4,975
Distance to dealers, km						
Group 1: Age < 40, Inc. $=$ Low	18.8	21.3	0.2	11.2	47.3	492,623
Group 2: Age < 40, Inc. $=$ High	17.2	18.3	2.2	12.1	38.2	440,104
Group 3: Age $\in [40, 60)$, Inc. $=$ Low	21.6	22.8	0.3	14.3	51.4	463,420
Group 4: Age $\in [40, 60)$, Inc. $=$ High	19.1	19.7	2.7	13.7	41.0	485,663
Group 5: Age ≥ 60, Inc. $=$ Low	25.2	21.8	2.3	21.5	52.7	452,574
Group 6: Age ≥ 60, Inc. $=$ High	17.7	20.0	0.5	11.8	40.8	494,653
Car characteristics						
Net list price, in €	22,640	9,581	12,718	20,400	33,591	4,975
Horsepower, in kW	75.5	22.9	51.0	70.0	103.0	4,975
Weight, in kg	1,736	274	1,418	1,700	2,080	4,975
Fuel cost, in €/100km	6.5	1.6	4.8	6.5	8.4	4,975
Fuel consumption, in L/100km	4.7	0.9	3.7	4.7	5.9	4,975
Gasoline	0.49	0.50	0	0	1	4,975
Diesel	0.45	0.50	0	0	1	4,975
Electric	0.02	0.12	0	0	0	4,975
Plug-in hybrid	0.00	0.07	0	0	0	4,975
Hybrid	0.04	0.19	0	0	0	4,975
Sedan	0.73	0.44	0	1	1	4,975
Convertible	0.01	0.08	0	0	0	4,975
Station wagon	0.27	0.44	0	0	1	4,975

Notes: Sales are aggregated at the national level, by product, year, and demographic groups. The associated statistics are unweighted. Distance to dealer is the driving distance to the closest dealer of each brand by demographic group, and the associated statistics are weighted by brand importance and municipal-level populations. All other statistics are weighted by sales. All monetary values are converted to 2018 euros.

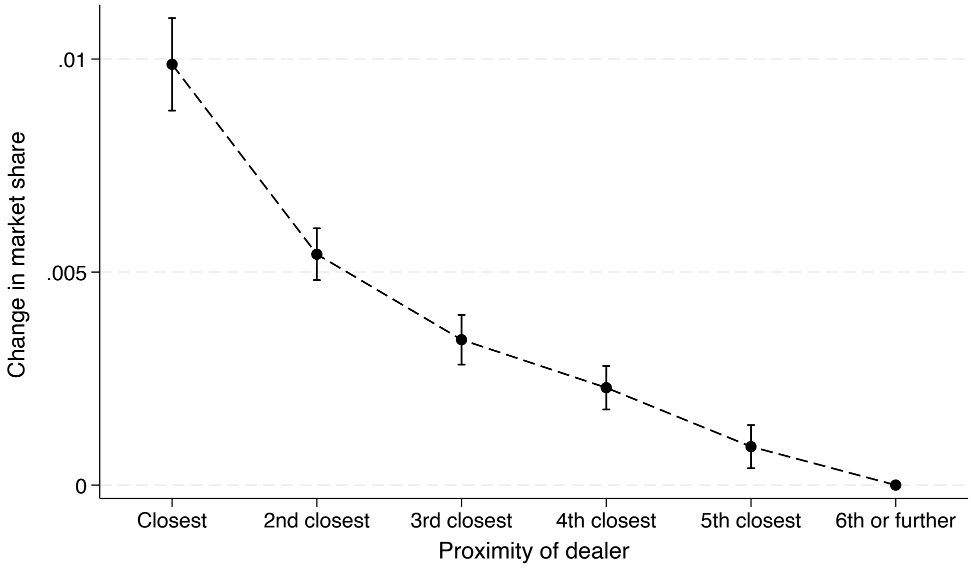
further away from car dealers than other groups. This is explained by the fact that a large share of these consumers live in rural areas compared to other groups. A second observation is that car dealers are on average slightly closer to high income municipalities than low income ones.

Finally, the third panel offers a summary of the car characteristics included in the utility specification. Since all consumers face the same set of products, we provide a common set of statistics, weighted by total sales.

3.4 Evidence of transportation costs

We provide some initial evidence that the distance to the car dealers, which we use as a proxy for transportation costs, matters for consumers. To achieve this, we correlate the market share of each brand at the municipal level with proximity dummies for car

Figure 3: Market share advantage from dealer proximity



Notes: This figure represents the coefficient from a regression of market shares at the level of the brand, year, and municipality, on brand proximity indicator variables. Includes municipality, and brand × year fixed effects. The brackets represent the 95% confidence interval, clustered at the municipality level. The calculation of the market share excludes the outside option and pools sales from all products within the same brand and all demographic groups.

dealers. We would expect for example that the market share of Renault is larger in municipalities where it is the closest dealer versus the second closest dealer, the second closest dealer versus the third, the third closest dealer versus the fourth, and so on. The results of the regression of the market shares on proximity dummies are depicted in Figure 3. We observe a clear decreasing pattern, indicating a clear correlation between proximity and sales.

3.5 Estimation results

We proceed with the estimation of our structural demand model. As stated previously, we define products as brand, model, engine type, and body trim combinations. We consider each year to be a different market and set the potential market for each demographic group to be one quarter of the number of households in that group, by year. We include the list price (net of rebates and fees), the horsepower (in 100 kW), the weight (in 1,000 kg), the fuel cost (in euros per 100 km), and fixed effects for the various engine types and body trims. Consumers' preferences for these characteristics are allowed to vary across demographic group. We also include brand and year fixed effects that are common across groups to capture brand perceptions and market conditions not accounted for by the model. Finally, we include the driving distance to the closest brand of each dealer (in 10 km).

We estimate both the demand parameters and the marginal cost equations simultaneously. Our cost specification includes the horsepower, the weight, the fuel consumption (in litters per 100 km), fixed effects for the engine types and trims, and a time trend. We include two cost shifters. First, we interact the average yearly price of several key inputs (steel, iron, plastics, and aluminium) with the car's weight to compute a single input price index, similarly to D'Haultfœuille et al. (2019). We assume cars are made of 56% steel, 8% iron, 8% plastics, 10% aluminium, and 18% other materials not captured by the index. Second, we follow Grieco et al. (2023) and use the real exchange rate interracted with the car's country of origin as an additional cost shifter.[14] The real exchange rate is meant to capture differences in the cost of labor for each brand. Finally, we lag both cost shifters by one year to reflect planning horizons.

We deal with the endogeneity of prices and the market shares using BLP-type instruments, constructed from exogenous car characteristics. On the demand side, the same set of instruments is used for all demographic groups (such that $Z_{jd} = Z_j, \forall d = 1, ..., D$). The instruments are constructed as the sum of exogenous characteristics of competitors' products. The chosen characteristics are the horsepower, the weight, and the fuel cost. We also include the number products sold by competitors, the number of products sold by competitors that have the same engine type, and the number of products sold by competitors that have the same body trim.

We construct instruments for the supply side in a similar fashion, using the horsepower, the weight, the fuel consumption, both cost shifters, the number of competing products, the number of competing products that have the same engine type, and the number of competing products that have the same body trim.

As discussed in Section 2.3, our estimator is robust to potential endogeneity of the observed distances without requiring additional intruments. Additional details on estimation can be found in the Appendix B.2.

We present the estimated parameters in Table 3. Our first observation is that there is significant heterogeneity in price sensitivities across demographic groups. The price sensitivities vary between -1.861 to -1.448, and the associated own-price elasticities range from -3.92 to -3.30, see Table 4. Two patterns emerge with respect to these price sensitivities. First, consumers from high-income groups are less price sensitive than

[14]The real exchange rate is taken from Penn World Tables 10.0, `pl_con`. See Feenstra et al. (2015).

Table 3: Parameter estimates

	Demand parameters						Cost function
	Group 1	Group2	Group 3	Group 4	Group 5	Group 6	c_j
Price (α_d)	-1.861	-1.772	-1.656	-1.448	-1.788	-1.567	
	(0.059)	(0.065)	(0.061)	(0.075)	(0.070)	(0.061)	
Distance (γ_d)	-0.218	-0.144	-0.171	-0.166	-0.322	-0.204	
	(0.066)	(0.062)	(0.076)	(0.063)	(0.106)	(0.100)	
Horsepower	3.230	3.053	2.822	2.473	3.212	2.810	0.086
	(0.111)	(0.106)	(0.101)	(0.092)	(0.113)	(0.105)	(0.063)
Weight	1.201	1.413	1.269	1.002	1.390	1.206	0.147
	(0.119)	(0.112)	(0.108)	(0.100)	(0.123)	(0.114)	(0.069)
Fuel cost	-0.211	-0.188	-0.149	-0.139	-0.171	-0.172	
	(0.022)	(0.020)	(0.019)	(0.018)	(0.020)	(0.019)	
Fuel consumption							-0.080
							(0.084)
Diesel	0.286	0.056	0.451	0.180	0.014	-0.381	-0.411
	(0.084)	(0.079)	(0.074)	(0.071)	(0.078)	(0.075)	(0.060)
Electric	-0.087	0.171	0.160	0.269	-0.464	-0.556	-0.344
	(0.232)	(0.215)	(0.204)	(0.188)	(0.210)	(0.197)	(0.126)
Plug-in hybrid	-0.743	-0.711	-0.344	-0.283	-0.481	-0.587	-0.227
	(0.205)	(0.196)	(0.186)	(0.173)	(0.197)	(0.184)	(0.062)
Hybrid	0.217	0.093	0.388	0.237	0.481	0.262	-0.137
	(0.110)	(0.104)	(0.101)	(0.096)	(0.106)	(0.102)	(0.068)
Convertible	-0.733	-0.665	-0.491	-0.372	-0.959	-0.770	-0.391
	(0.096)	(0.096)	(0.084)	(0.078)	(0.094)	(0.083)	(0.089)
Station wagon	0.105	0.195	0.031	0.102	0.011	0.072	0.058
	(0.048)	(0.046)	(0.045)	(0.043)	(0.050)	(0.048)	(0.058)
Input price index							-0.297
							(0.067)
Real exchange rate							-0.339
							(0.061)
Trend							-0.225
							(0.064)
Willingness to pay	117.2	81.1	103.2	114.9	180.3	130.1	
	(33.6)	(33.4)	(44.1)	(40.5)	(57.5)	(62.1)	
Observations	4,975	4,975	4,975	4,975	4,975	4,975	4,975

Notes: The demand-side specification includes (non-group specific) brand and year fixed effects. Price is in €10,000, Horsepower is in 100kW, Weight is in 1,000kg, Fuel cost is in €/100km, and Fuel consumption is in L/100km. Distance is the driving distance to the nearest retailer of each brand, in 10 km. Willingness to pay, in €/km, is computed as γ_d/α_d $(\times 1,000)$ for each demographic group. Its standard error is computed using the delta method. Robust standard errors in parenthesis.

Table 4: Estimated own-price elasticities

Group definition	Mean	Std. dev.	10th pct.	Median	90th pct.	Observations
Group 1: Age < 40, Inc. $=$ Low	-3.92	1.81	-6.01	-3.52	-2.03	4,975
Group 2: Age < 40, Inc. $=$ High	-3.78	1.72	-5.78	-3.40	-1.99	4,975
Group 3: Age $\in [40, 60)$, Inc. $=$ Low	-3.62	1.60	-5.48	-3.26	-1.94	4,975
Group 4: Age $\in [40, 60)$, Inc. $=$ High	-3.30	1.40	-4.93	-2.99	-1.83	4,975
Group 5: Age ≥ 60, Inc. $=$ Low	-3.83	1.73	-5.84	-3.45	-2.01	4,975
Group 6: Age ≥ 60, Inc. $=$ High	-3.50	1.51	-5.27	-3.17	-1.89	4,975

Notes: To maintain comparability, statistics are computed using a set of uniform weights $w_j = \sum_d \phi_d s_{jd} / \sum_j \sum_d \phi_d s_{jd}$ that are common across demographic groups.

low-income consumers. Second, price sensitivities seem to follow a strict ranking with respect to age (within income categories): middle aged individuals have the lowest price sensitivities (group 3 and 4), followed by older consumers (group 5 and 6), and younger consumers are the most price sensitive (group 1 and 2). Comparing both dimensions, we find that age is a more important determinant of price sensitivity than income.

Interestingly, these patterns match our preliminary evidence from Figure 1, computed from an alternative data source. Only the magnitude of the effect differs: for example, our estimated parameters imply that young, low income individuals receive an average discount of €1,813 over middle aged, high income consumers (see Table 5) while our preliminary evidence suggested an average discount in the range of €3,622. These findings are not completely inconsistent, if part of the price dispersion from our preliminary analysis is due to additional options which we do not control for.

Our second observation relates to the disutility of distance, which we associate to transportation costs. We find that consumers are also heterogeneous in that regard: the estimated parameters range from -0.322 to -0.144. As for price sensitivities, we find that consumers from high-income groups seem less sensitive to distance than low-income consumers. This could happen if low-income consumers face constraints that make traveling prohibitive. We do not recover a clear ranking with respect to age, apart from older consumers (group 5 and 6) having the strongest distaste for traveling distance within income categories.

We combine our price and distance parameters to compute consumers' willingness to pay to reduce traveling distance by one kilometer. Our estimated willingness to pay range from €81.1 to €180.3. This is in line with previous results from Nurski and Verboven (2016) which estimate willingness to pay to be €112 for the Belgian car market. Our estimates, while being of a similar magnitude, provide new evidence that consumers have heterogeneous transportation costs. While these estimates seem large at first glance, we

Table 5: Estimated transaction prices

Description	Mean	Std. dev.	10th pct.	Median	90th pct.	Observations
Transaction price (€)						
Group 1: Age < 40, Inc. $=$ Low	21,070	9,711	10,901	18,944	32,318	4,975
Group 2: Age < 40, Inc. $=$ High	21,378	9,705	11,228	19,245	32,627	4,975
Group 3: Age $\in [40, 60)$, Inc. $=$ Low	21,887	9,676	11,720	19,756	33,085	4,975
Group 4: Age $\in [40, 60)$, Inc. $=$ High	22,883	9,656	12,768	20,797	34,087	4,975
Group 5: Age ≥ 60, Inc. $=$ Low	21,476	9,666	11,272	19,324	32,690	4,975
Group 6: Age ≥ 60, Inc. $=$ High	22,389	9,649	12,119	20,354	33,694	4,975
Discount (€)						
Group 1: Age < 40, Inc. $=$ Low	1,813	195	1,581	1,800	2,101	4,975
Group 2: Age < 40, Inc. $=$ High	1,505	171	1,304	1,504	1,759	4,975
Group 3: Age $\in [40, 60)$, Inc. $=$ Low	996	82	887	991	1,111	4,975
Group 4: Age $\in [40, 60)$, Inc. $=$ High	0	0	0	0	0	4,975
Group 5: Age ≥ 60, Inc. $=$ Low	1,408	83	1,332	1,368	1,526	4,975
Group 6: Age ≥ 60, Inc. $=$ High	495	91	352	510	597	4,975
Discount (%)						
Group 1: Age < 40, Inc. $=$ Low	9.27	3.82	4.90	8.69	14.69	4,975
Group 2: Age < 40, Inc. $=$ High	7.71	3.20	4.02	7.20	12.29	4,975
Group 3: Age $\in [40, 60)$, Inc. $=$ Low	5.07	2.01	2.77	4.82	7.93	4,975
Group 4: Age $\in [40, 60)$, Inc. $=$ High	0	0	0	0	0	4,975
Group 5: Age ≥ 60, Inc. $=$ Low	7.15	2.78	4.07	6.75	11.24	4,975
Group 6: Age ≥ 60, Inc. $=$ High	2.50	1.08	1.35	2.29	4.02	4,975

Notes: All monetary values are converted to 2018 euros. The pivot group is Group 4 for all products. Statistics are computed using a set of uniform weights $w_j = \sum_d \phi_d s_{jd} / \sum_k \sum_d \phi_d s_{kd}$ that are common across demographic groups.

must keep in mind that they encompass all visits to the dealers. In particular, they can include visits prior to the purchase (e.g., the customer went for a test drive) and expected future visits (e.g., maintenance, after-sale services, etc.).

Finally, we find that consumers are also heterogeneous in their preferences for other car attributes, however no clear pattern emerges. We find that consumers in general prefer powerful cars, heavier vehicles (weight is a proxy for security), and dislike fuel costs. Our cost estimates suggest that increasing horsepower or decreasing fuel consumption is costly for firms.

Our approach allows us to recover transaction prices paid by different consumer groups, resulting from price discrimination. We report on these estimated transaction prices in Table 5. For each group, we provide the average transaction price, its standard deviation, as well as key percentiles of the distribution. To avoid group-specific sales driving the results, we compute all statistics using a single set of weights, based on the total sales of each model (i.e., aggregated over groups).

We want to highlight some key facts regarding transaction prices. First, we find that group 4 (aged between 40 and 59, high income) is the pivot group for all products, hence it always pays the observed list price. This follows from the fact that it has the lowest price sensitivity of all groups. Second, we find that discounts can be significant: the average discount ranges from 2.5% to 9.3%, corresponding to €495 and €1,813 respectively. In extreme cases, discounts can reach 14.7% or €2,101 (e.g., see group 1 at the 90th percentile of the discount distribution).

3.6 Disentangling the effects of price discrimination and transportation costs

Before we move on to the main part of the analysis, which involves the introduction of an additional online sales channel, we perform several counterfactual experiments to understand how price discrimination and transportation costs interact with each other. The results are presented in Table 6 and Table 7.

We focus on three broad sets of counterfactual experiments. First, we consider the case where firms cannot price discriminate consumers. Second, we consider the case where price discrimination is possible, but consumers face reduced transportation costs. We allow for various levels of transportation cost reductions which allow consumers to still have some value for dealer proximity. Finally, we consider the case where both price discrimination is not possible and consumers face reduced transportation costs. This last counterfactual coincides with our description of a world where all sales occur online.

We highlight several key observations. The first observation relates to price discrimination. We focus the first two rows of Table 7 and compare a counterfactual without price discrimination to the baseline. Most consumers (mostly low-income groups) benefit from a discount over the non-discriminatory prices. This increases their total purchases and their gain in consumer surplus ranges between €36.5 and €102.6 per consumer per year. Consumers from group 4 and 6 (high-income, aged 40 and above) pay higher prices under price discrimination and reduce their purchases of all car models. Their loss of consumer surplus is significantly larger, €168.2 and €64.6 per consumer per year. Overall, price discrimination decreases consumer surplus by €7 per consumer per year. Whether or not price discrimination is profitable remains an empirical question in oligopoly settings. We find that industry profits increase by less than one percent, suggesting that price discrimination is not particularly profitable for car manufacturers in the French market. Since the gains from price discrimination are small for firms, and the loss are small on average for consumers, we conclude that price discrimination is mostly redistributive (shifting surplus from rich to poor).

Our second observation concerns transportation costs and prices. We compared the set of counterfactuals with price discrimination and reduced transportation costs to the baseline. We notice that reducing transportation costs (with or without price discrimination) does not seem to particularly affect pricing decisions (see Panel 1 of Table 6). In some sense, when consumers face reduced transportation costs, firms are not able to capture the associated gains through increased prices. Instead, large profits are accrued from a large market expansion in all demographic groups. This suggests that the pass-through of transportation costs to prices is small.

Our third observation concerns consumers' response to a reduction in transportation costs. We again contrast the set of counterfactuals with price discrimination and reduced transportation costs to the baseline. We notice that, as we reduce transportation costs, consumers switch gradually to more expensive cars (see Panel 2 of Table 6) from dealers located on average further away (see Panel 4 of Table 6). In some sense, consumers save on transportation costs, and "reinvest" part of these gains by spending more on cars and shopping at dealers located further away. Our model predicts that removing transportation costs entirely can lead to an increase in consumer surplus and profits in the range of 30%.

Finally, we comment on the magnitude of the observed effects. We notice that even modest reductions in transportation costs (e.g., a 25% decrease) result in an effect of roughly the same magnitude as removing price discrimination. Eliminating transportation costs entirely leads to an effect more than one order of magnitude larger than removing price discrimination. Therefore, our estimated preference parameters imply that price discrimination is a second order effect compared to transportation costs. This provides some additional evidence that firms would benefit from moving their business entirely online, with or without price discrimination.

Table 6: Effect of price discrimination and transportation costs on consumers' purchases

	Baseline	No discr.	Reduced transportation costs				No discr., reduced transportation costs			
			90%	75%	50%	No cost	90%	75%	50%	No cost
Transaction prices, uniform weights										
Group 1: Age < 40, Inc. $=$ Low	21,825		21,826	21,828	21,832	21,843				
Group 2: Age < 40, Inc. $=$ High	22,119		22,120	22,122	22,126	22,136				
Group 3: Age $\in [40, 60)$, Inc. $=$ Low	22,611		22,613	22,618	22,626	22,650				
Group 4: Age $\in [40, 60)$, Inc. $=$ High	23,583		23,587	23,593	23,606	23,636				
Group 5: Age ≥ 60, Inc. $=$ Low	22,222		22,227	22,235	22,255	22,322				
Group 6: Age ≥ 60, Inc. $=$ High	23,118		23,123	23,130	23,145	23,182				
Uniform		22,857					22,858	22,860	22,865	22,879
Transaction prices, sales-weighted										
Group 1: Age < 40, Inc. $=$ Low	20,585		20,610	20,652	20,737	21,002				
Group 2: Age < 40, Inc. $=$ High	22,476		22,493	22,522	22,577	22,726				
Group 3: Age $\in [40, 60)$, Inc. $=$ Low	22,555		22,592	22,653	22,776	23,145				
Group 4: Age $\in [40, 60)$, Inc. $=$ High	24,736		24,771	24,827	24,936	25,255				
Group 5: Age ≥ 60, Inc. $=$ Low	20,727		20,769	20,843	21,006	21,675				
Group 6: Age ≥ 60, Inc. $=$ High	23,129		23,164	23,223	23,345	23,765				
Uniform		22,583					22,612	22,662	22,762	23,099
Sales, in units										
Group 1: Age < 40, Inc. $=$ Low	101,254	-15,689	+2,630	+6,896	+15,009	+36,453	-13,435	-9,772	-2,784	+15,807
Group 2: Age < 40, Inc. $=$ High	71,907	-7,488	+1,303	+3,347	+7,016	+15,543	-6,302	-4,440	-1,093	+6,692
Group 3: Age $\in [40, 60)$, Inc. $=$ Low	135,790	-4,313	+3,214	+8,351	+17,859	+41,324	-1,158	+3,892	+13,264	+36,496
Group 4: Age $\in [40, 60)$, Inc. $=$ High	232,803	+19,300	+4,740	+12,178	+25,535	+56,637	+24,402	+32,401	+46,745	+79,976
Group 5: Age ≥ 60, Inc. $=$ Low	104,283	-9,262	+4,764	+12,671	+28,235	+71,566	-4,812	+2,611	+17,361	+59,356
Group 6: Age ≥ 60, Inc. $=$ High	314,559	+9,555	+6,665	+17,207	+36,400	+82,626	+16,478	+27,430	+47,367	+95,277
Average distance to dealers, in km										
Group 1: Age < 40, Inc. $=$ Low	12.43	+0.04	+0.48	+1.27	+2.81	+7.09	+0.52	+1.32	+2.87	+7.22
Group 2: Age < 40, Inc. $=$ High	13.76	+0.04	+0.28	+0.72	+1.54	+3.61	+0.32	+0.76	+1.59	+3.70
Group 3: Age $\in [40, 60)$, Inc. $=$ Low	15.97	+0.01	+0.53	+1.39	+2.99	+7.16	+0.54	+1.40	+3.01	+7.21
Group 4: Age $\in [40, 60)$, Inc. $=$ High	15.54	+0.03	+0.36	+0.93	+2.01	+4.97	+0.38	+0.95	+2.03	+4.96
Group 5: Age ≥ 60, Inc. $=$ Low	16.27	-0.08	+0.76	+2.01	+4.42	+11.52	+0.69	+1.94	+4.35	+11.45
Group 6: Age ≥ 60, Inc. $=$ High	13.52	-0.02	+0.41	+1.07	+2.35	+6.21	+0.39	+1.04	+2.31	+6.13

Notes: All counterfactual experiments are computed using the 2019 data only. Unless indicated otherwise, in-person sales imply price discrimination and transportation costs and online sales imply a uniform price and reduced transportation costs. Transaction prices are in 2018 euros. "Uniform weights" are constructed using the total sales of each product in the baseline scenario, hence are fixed across demographic groups and counterfactual experiments. "Sales weights" use realized sales for each demographic group and counterfactual experiment. For sales and average distances, we report the values at baseline in the first column, and the change from baseline in the other columns.

Table 7: Effect of price discrimination and transportation costs on welfare

Counterfactual	Δ Consumer surplus, per capita							Δ Profits
	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6	All	Total
Baseline	411.2	589.7	1,053.5	1,757.6	1,051.7	1,823.7	1,136.1	6,229.8
No discrimination	-65.8	-64.3	-36.5	+168.2	-102.6	+64.6	+4.2	-53.6
Reduced transportation costs								
• 90% transportation costs	+11.0	+11.1	+26.6	+40.0	+51.0	+43.4	+29.6	+151.9
• 75% transportation costs	+28.8	+28.7	+69.5	+103.3	+136.6	+112.9	+77.3	+395.3
• 50% transportation costs	+63.0	+60.3	+149.9	+218.7	+309.4	+241.8	+167.2	+848.1
• No transportation costs	+154.7	+134.8	+354.8	+497.4	+827.0	+567.3	+401.3	+1,986.8
No discr. and reduced transportation costs								
• 90% transportation costs	-56.5	-54.2	-10.5	+212.2	-55.7	+110.2	+34.0	+96.5
• 75% transportation costs	-41.3	-38.4	+31.5	+281.9	+23.4	+183.1	+81.9	+337.0
• 50% transportation costs	-12.2	-9.9	+110.3	+409.1	+184.2	+318.5	+172.2	+784.5
• No transportation cost	+66.1	+57.3	+311.8	+715.7	+674.3	+660.3	+407.3	+1,911.6

Notes: All counterfactual experiments are computed using the 2019 data only. Unless indicated otherwise, in-person sales imply price discrimination and transportation costs and online sales imply a uniform price and reduced transportation costs. We report the values at baseline in the first row, and the change from baseline in the other rows.

4 Model with the online sales channel

4.1 Introducing online car dealers

We extend our model to investigate a set of counterfactual scenarios in which cars can be purchased either in person at the closest dealer of the chosen brand (as observed in the data) or online directly from a firm's website (a channel currently not observed in the data). By completing the transaction online and having the car delivered to their doorsteps, consumers would face a lower transportation cost since fewer or no visits to the car dealer are required, and would pay the non-discriminatory uniform price advertised on the website. Concluding the transaction in person instead implies that the consumer would physically travel, potentially multiple times, to a car dealer which would then offer them a personalized price different from the uniform online price.

Throughout, we are agnostic about the extend to which buying online reduces consumers' transportation costs. In extreme cases where consumers do not travel to the dealer for a test drive and do not value after-sale services, purchasing the car online could remove transportation costs entirely. In more realistic cases, some transportation costs may remain if the consumer expects to go to the dealer for maintenance in the future. In the absence of clear guidance on this subject, we allow for various levels of transportation cost reductions, captured by the parameter $\tau \in [0, 1)$.

We also focus our attention on the case where car dealers do not price discriminate against consumers who purchase online. We do this to match the observed industry practice, which leans heavily towards price transparency and the streamlining of the transaction process. Our model is flexible enough to allow for pricing discrimination in both sales channel. We focus on the case where firms use different price regimes offline and online to match the observed industry practice. For completeness, we evaluate a set of counterfactuals with price discrimination in both sales channels. The results are relayed to Appendix C.

Lastly, we assume that introducing an online channel does not change the observed configuration of stores (e.g., no entry or exit of dealers) or the vertical relations between the dealers and the manufacturers (e.g., the manufacturer cannot eliminate the middleman). We evaluate the robustness of our results to these assumptions in Section 5.5 and leave the study of these additional channels for future research.

4.2 Demand

We extend our model from Section 2.1 in the simplest way possible that allows us to capture key features of the online channel. For clarity, we define p^P_{jd} (previously p_{jd}) as the discriminatory in-person price paid for product j by consumers of group d, and p^O_j as the uniform online price paid for product j by all consumers. We maintain the assumption that all car dealers operate in both the in-person and online channels, and that all car models are available in both channels.

We allow for the possibility that consumers belonging to different demographic groups have different propensities toward online shopping. For example, younger and wealthier consumers may be at ease with the option of purchasing a car from a firm's website, while older and less affluent consumers may not be willing or able to do so. We rely on a national survey on attitudes toward online shopping to estimate the probability with which consumers belonging to each demographic group consider the online sales channel in their choice set, denoted ψ_d. We report on these probabilities in Table A.1.

A share $1 - \psi_d$ of consumers do not have access to the online sales channel and their indirect utility is given by (1). For the remaining share of ψ_d consumers who can access both channels, their indirect utility is

$$U_{ijdm} = \underbrace{X_j'\beta_d + \alpha_d p^P_{jd} + \xi_{jd}}_{\delta_{jd}} + \max\Big\{\underbrace{\gamma_d \text{dist}_{jm}}_{\mu^P_{jdm}}, \underbrace{\alpha_d(p^O_j - p^P_{jd}) + \tau\gamma_d \text{dist}_{jm}}_{\mu^O_{jdm}}\Big\} + \epsilon_{ijdm}, \quad (14)$$

where X_j is a vector of observed car characteristics, ξ_{jd} is a vector of group-specific preferences for unobserved car characteristics, dist_{jm} is the distance between municipality m and the closest dealer of car model j, and $\tau \in [0, 1]$ is a parameter to be calibrated that controls the transportation cost reduction in the online channel. Note that μ^P_{jdm}

(previously μ_{jdm}) represents the transportation cost of purchasing a car in-person as before, while μ^O_{jdm} represents the trade-off faced by consumers who shop online (i.e., reduced transportation costs versus losing a potential discount).

Given this specification, the probability that consumers in demographic group d and municipality m to purchase car model j is

$$\begin{aligned} s_{jdm}(\text{dist}_{1m}, ..., \text{dist}_{Jm}) = \psi_d \cdot & \frac{\exp\left(\delta_{jd} + \max\{\mu^P_{jdm}, \mu^O_{jdm}\}\right)}{1 + \sum_{k=1}^J \exp\left(\delta_{kd} + \max\{\mu^P_{kdm}, \mu^O_{kdm}\}\right)} \\ & + (1 - \psi_d) \cdot \frac{\exp\left(\delta_{jd} + \mu^P_{jdm}\right)}{1 + \sum_{k=1}^J \exp\left(\delta_{kd} + \mu^P_{kdm}\right)}. \end{aligned} \tag{15}$$

Note that we can equivalently express (15) as $s_{jdm} = s^P_{jdm} + s^O_{jdm}$, distinguishing between the share of in-person purchases (denoted by superscript P) and the share of online purchases (superscript O), that is,

$$\begin{aligned} s^P_{jdm}(\text{dist}_{1m}, ..., \text{dist}_{Jm}) = \;& \psi_d \cdot \frac{\exp(\delta_{jd} + \mu^P_{jdm}) \mathbb{1}\left\{\mu^P_{jdm} \geq \mu^O_{jdm}\right\}}{1 + \sum_{k=1}^J \exp(\delta_{kd} + \max\{\mu^P_{kdm}, \mu^O_{kdm}\})} \\ & + (1 - \psi_d) \cdot \frac{\exp\left(\delta_{jd} + \mu^P_{jdm}\right)}{1 + \sum_{k=1}^J \exp\left(\delta_{kd} + \mu^P_{kdm}\right)} \end{aligned} \tag{16}$$

$$s^O_{jdm}(\text{dist}_{1m}, ..., \text{dist}_{Jm}) = \; \psi_d \cdot \frac{\exp(\delta_{jd} + \mu^O_{jdm}) \mathbb{1}\left\{\mu^P_{jdm} < \mu^O_{jdm}\right\}}{1 + \sum_{k=1}^J \exp(\delta_{kd} + \max\{\mu^P_{kdm}, \mu^O_{kdm}\})}. \tag{17}$$

Averaging (16) and (17) over municipalities, we obtain the national-level market shares of group d for car model j from sales channel $\ell \in \{P, O\}$,

$$s^\ell_{jd}(\delta_d, \mu^P_d, \mu^O_d) = \sum_{m \in \mathcal{M}} w_{dm} \cdot s^\ell_{jdm}(\text{dist}_{1m}, ..., \text{dist}_{Jm}), \tag{18}$$

where $\mu^\ell_d = (\mu^\ell_{1dm}, ..., \mu^\ell_{Jdm})_{m \in \mathcal{M}}$, $\ell \in \{P, O\}$.

4.3 Supply

Similar to the model described in Section 2.2, we consider a Bertrand-Nash price-setting game in which every firm f chooses a menu of transaction prices for in-person sales $p^P_j = (p^P_{j1}, ..., p^P_{jD})$ and the non-discriminatory online price p^O_j for each j they sell by maximizing the national-level profit function

$$\pi_f(p_1^P, ..., p_D^P, p^O) = \sum_{d=1}^{D} \phi_d \sum_{j \in \mathcal{J}_f} s_{jd}^P (p_{jd}^P - c_j) + \sum_{d=1}^{D} \phi_d \sum_{j \in \mathcal{J}_f} s_{jd}^O (p_j^O - c_j), \tag{19}$$

where $p_d^P = (p_{1d}^P, ..., p_{Jd}^P)$ and $p^O = (p_1^O, ..., p_J^O)$.

4.4 Solving the model

The model with online sales presented in the previous subsection is hard to solve in practice. This is due to the maximum operator in indirect utility function (14), which leads to discontinuities in the resulting purchase probabilities (16)-(17). This causes traditional price optimization routines for the maximization of (19) to fail, as small price changes may cause discontinuous changes to the system of first-order conditions. A similar numerical problem was highlighted by Duch-Brown et al. (2023).

To circumvent this issue, we implement a methodology similar to Duch-Brown et al. (2023). The idea is to approximate the multinomial logit model implied by indirect utility (14) by a nested logit formulation in which each j belongs to a nest and where each of these J nests includes the two sales channels: in person (j, P) and online (j, O). In other words, consumers first choose which of the J car models (or the outside option) to purchase (i.e., the choose the "nest"), and then choose whether to purchase that car model in person or online. In this case, the indirect utility of purchasing car model j from $\ell \in \{P, O\}$ does not involve any maximum operator, and we have that

$$U_{ijdm}^{\ell} = X_j' \beta_d + \alpha_d p_{jd} + \xi_{jd} + \mu_{jdm}^{\ell} + \zeta_{ijdm} + (1 - \sigma)\epsilon_{ijdm}^{\ell}, \tag{20}$$

where both ϵ_{ijdm}^{ℓ} and $\zeta_{ijdm} + (1-\sigma)\epsilon_{ijdm}^{\ell}$ are distributed extreme value, ζ_{ijdm} is common to both sales channels $\ell \in \{P, O\}$ of car model j, and parameter $\sigma \in [0, 1)$ (Cardell, 1997). Importantly and as pointed out by Duch-Brown et al. (2023), when σ tends to 1, the two sales channels become perfect substitutes for each other, and the nested logit market share implied by indirect utility (20) converges to that of the multinomial logit implied by indirect utility (14).

We solve for the optimal price vector $p^* = (p_1^{P*}, ..., p_D^{P*}, p^{O*})$ by adapting the ζ-markup algorithm suggested by Morrow and Skerlos (2011). In practice, we cannot maximize profit function (19) at $\sigma = 1$ as the nested logit market shares are not defined. We instead compute market shares for values of σ numerically close to 1 and rely on extrapolation to approach the limit as $\sigma \to 1$. The computational details are reported in Appendix

B.3.

5 Counterfactual simulations

We focus in this section on our main counterfactual analysis, where we introduce a new online sales channel to the existing French car market. All counterfactual experiments are conducted on our 2019 data. Unless indicated otherwise, consumers shopping in person at the dealership receive a discount as a result of price discrimination, and incur full transportation costs. Consumers shopping online pay instead a uniform price advertised on the manufacturer's website, and reduce their transportation costs by a factor of $1-\tau$.

We consider four different levels of transportation cost reductions, that is $\tau = \{0.90, 0.75, 0.50, 0\}$. Additionally, we consider two cases for consumers' attitude towards online shopping. First, we posit that everyone can use either sales channel without restrictions. We use this as a benchmark to understand the forces at play. Then, we consider a more realistic scenario where some consumers never shop online. The propensity to shop online, denoted ψ_d, is calibrated using a survey of consumers' attitudes towards online shopping, as described in Section 3.1.

5.1 Unrestricted access to online dealers

The results from our counterfactuals with an unrestricted access to the online channel are presented in Table 8. We focus our attention on three key observations.

Our first observation concerns sales. We note that introducing the online channel creates a large market expansion, driven mostly by consumers in group 4 and 6 (high-income, aged above 40). The reason is that adding this option provides them with an opportunity to reduce both the price they pay for all products (the uniform online price is lower than the discriminatory price they previously paid in person), and to reduce their transportation costs. Therefore, the observed effect is mechanical: the increase in utility generates substitution away from the outside option. For other consumer groups, the price effect and the reduction in transportation costs go in opposite directions, so the impact on sales is smaller in magnitude (and can be negative for some groups).

Our second observation concerns consumers' shopping patterns and their attitude towards the online channel. We notice that once the online channel is available, most sales are diverted away from the traditional in-person channel. This is true even for small decreases in transportation costs. As an example, decreasing transportation costs by

Table 8: Effect of online channel with unrestricted access

	In-person only	Both channels, unrestricted online access			
	Baseline	$\tau = 0.90$	$\tau = 0.75$	$\tau = 0.50$	$\tau = 0$
Transaction prices, uniform weighted					
Group 1: Age < 40, Inc. $=$ Low	21,825	22,853	22,861	22,858	22,875
Group 2: Age < 40, Inc. $=$ High	22,119	22,862	22,863	22,865	22,875
Group 3: Age $\in [40, 60)$, Inc. $=$ Low	22,611	22,859	22,860	22,864	22,878
Group 4: Age $\in [40, 60)$, Inc. $=$ High	23,583	22,854	22,857	22,865	22,882
Group 5: Age ≥ 60, Inc. $=$ Low	22,222	22,861	22,859	22,861	22,877
Group 6: Age ≥ 60, Inc. $=$ High	23,118	22,857	22,859	22,865	22,879
Online		22,857	22,859	22,864	22,879
Transaction prices, sales-weighted					
Group 1: Age < 40, Inc. $=$ Low	20,585	21,122	20,843	20,648	20,985
Group 2: Age < 40, Inc. $=$ High	22,476	22,831	22,451	21,326	20,822
Group 3: Age $\in [40, 60)$, Inc. $=$ Low	22,555	22,218	21,936	21,472	21,819
Group 4: Age $\in [40, 60)$, Inc. $=$ High	24,736	23,427	22,832	21,182	22,948
Group 5: Age ≥ 60, Inc. $=$ Low	20,727	20,826	20,407	20,384	20,839
Group 6: Age ≥ 60, Inc. $=$ High	23,129	22,408	22,046	21,757	22,220
Online		22,773	22,788	22,840	23,149
Sales, in units					
Group 1: Age < 40, Inc. $=$ Low	101,254	-14,260	-10,698	-3,343	+15,506
Group 2: Age < 40, Inc. $=$ High	71,907	-6,950	-5,217	-1,640	+6,393
Group 3: Age $\in [40, 60)$, Inc. $=$ Low	135,790	-2,136	+2,912	+12,652	+36,173
Group 4: Age $\in [40, 60)$, Inc. $=$ High	232,803	+22,979	+30,792	+45,848	+79,629
Group 5: Age ≥ 60, Inc. $=$ Low	104,283	-5,729	+1,932	+16,986	+59,121
Group 6: Age ≥ 60, Inc. $=$ High	314,559	+14,675	+25,537	+46,367	+94,889
Prop. of online sales					
Group 1: Age < 40, Inc. $=$ Low	0	0.673	0.829	0.913	0.935
Group 2: Age < 40, Inc. $=$ High	0	0.655	0.830	0.953	0.985
Group 3: Age $\in [40, 60)$, Inc. $=$ Low	0	0.729	0.854	0.926	0.952
Group 4: Age $\in [40, 60)$, Inc. $=$ High	0	0.769	0.913	0.985	0.986
Group 5: Age ≥ 60, Inc. $=$ Low	0	0.840	0.931	0.963	0.974
Group 6: Age ≥ 60, Inc. $=$ High	0	0.757	0.891	0.955	0.963
Average distance to dealers, in km					
Group 1: Age < 40, Inc. $=$ Low	12.43	+0.51	+1.32	+2.87	+7.22
Group 2: Age < 40, Inc. $=$ High	13.76	+0.32	+0.76	+1.59	+3.70
Group 3: Age $\in [40, 60)$, Inc. $=$ Low	15.97	+0.54	+1.40	+3.01	+7.21
Group 4: Age $\in [40, 60)$, Inc. $=$ High	15.54	+0.38	+0.95	+2.03	+4.96
Group 5: Age ≥ 60, Inc. $=$ Low	16.27	+0.69	+1.94	+4.35	+11.45
Group 6: Age ≥ 60, Inc. $=$ High	13.52	+0.39	+1.04	+2.31	+6.13

Notes: All counterfactual experiments are computed using the 2019 data only. Unless indicated otherwise, in-person sales imply price discrimination and transportation costs and online sales imply a uniform price and reduced transportation costs. Transaction prices are in 2018 euros. “Uniform weights” are constructed using the total sales of each product in the baseline scenario, hence are fixed across demographic groups and counterfactual experiments. “Sales weights” use realized sales for each demographic group and counterfactual experiment. For sales and average distances, we report the values at baseline in the first column, and the change from baseline in the other columns.

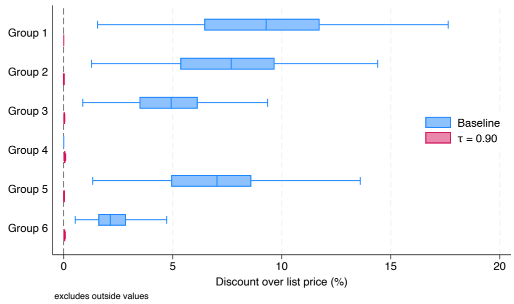
10% leads to over 70% of consumers to shop online. If transportation costs are eliminated, then more than 95% of consumers shop online (the only consumers shopping in person are those with a distance to dealer very close or equal to zero). Similar to our preliminary evidence in Section 3.6, we find that consumers opt for dealers that are located further away from where they live. This is attributed both to the decrease in transportation costs and changes in the price schedule.

Our last observation concerns price dispersion in the traditional sales channel. We find that once the online channel with a uniform price becomes available, dealers have the incentive to end price discrimination in the offline channel (see Panel 1 of Table 8). This is a result of the competitive pressure that the online channel exerts on the traditional sales channel. In some sense, because of the lower transportation costs, firms benefit by redirecting most consumers to that channel which increases sales. However, online prices are restricted to be uniform. In that context, firms attract consumers to the online channel by setting the same uniform price in both channels, which makes the online channel unambiguously better.

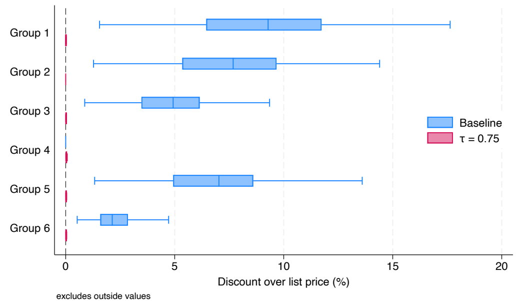
Figure 4 provides the distribution of discounts when the online channel is absent versus when it is available. We see clearly that discounts are eliminated at the product level, even for very small reductions in transportation costs.

We believe that this finding has important implications. For example, it shows that the mere availability of the online channel with a fixed price is sufficient to enforce a fixed price across the industry and in both sales channel. If for example price discrimination arose from price manipulation by the car dealers, then our results suggest that the car manufacturer can discipline these dealers by selling directly online at a fixed price (a form of resale price maintenance).

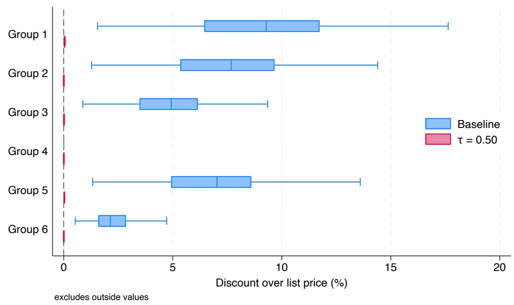
Figure 4: Price dispersion with unrestricted online access



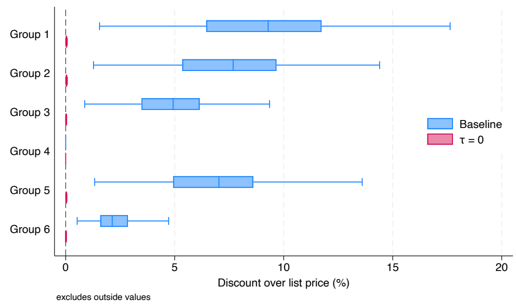
(a) Transportation costs = 75%



(b) Transportation costs = 50%



(c) Transportation costs = 25%



(d) No transportation cost

Notes: These figures represent the price dispersion in the in-person channel when consumers have a restricted access to the online channel, for varying transportation cost savings as per Table 8. Price dispersion is represented as a discount over the list price.

5.2 Restricted access to online dealers

We now turn to our preferred specification. Since we cannot estimate a preference parameter for the online channel in the utility specification (as there are no online sales at the estimation stage), we enrich our model with online sales by restricting the access to the online channel for a subset of consumers. We believe that this approach constitute a valid, feasible alternative in this context, given that several consumers could be averse to purchasing such an important product online. Table 9 reports the chosen propensity to shop online by demographic group, which we calibrated using a survey of consumers attitudes towards online shopping. We constructed these probabilities as the proportion of consumers of each demographic group who purchased anything online in the year prior to the survey. A share ψ_d of consumers have access to both online and offline sales. For the remaining $(1 - \psi_d)$ consumers, we consider them to be captive to the traditional sales channel.

Table 9: Propensity to shop online, in 2019

Demographic group	Propensity (ψ_d)	Observations
Group 1: Age < 40, Income $=$ Low	0.770	9,545
Group 2: Age < 40, Income $=$ High	0.894	12,131
Group 3: Age $\in [40, 60)$, Income $=$ Low	0.505	14,098
Group 4: Age $\in [40, 60)$, Income $=$ High	0.770	19,614
Group 5: Age ≥ 60, Income $=$ Low	0.146	20,034
Group 6: Age ≥ 60, Income $=$ High	0.464	17,724

We conduct a similar set of counterfactual experiments, with varying levels of transportation reductions accruing from the online channel, and where consumers are restricted to shop in-person according to Table 9. The results of these counterfactual experiments are presented in Table 10. We notice that restricting consumers limits the expansion in sales observed in the previous set of counterfactuals, driven by a decrease in online sales from the most restricted groups. For this reason, the proportion of consumers who shop online also decreases significantly. For example, less than 25% of consumers who belong to group 5 (low-income, 60 years and older) purchase a car online. For groups 3 and 6, less than 60% of consumers buy their cars online. As a result, the average distance to dealers does not increase as much, since many consumers are restricted to the traditional sales channel.

Our next comment relates to prices and price dispersion. In this case, we can see that introducing an online channel with uniform prices does not completely iron out price dispersion, although consumers receive much smaller discounts. We notice an interesting pattern: some consumers actually must pay a premium over the online uniform price to shop in person (e.g., group 4 and 6). One way to interpret this is that firms set a list price above the online price, such that these consumers get a discount when shopping online. This results from the fact that some consumers are captive to the offline channel.

We highlight two opposing forces at play. On the one hand, the competitive pressure from the online channel onto the traditional channel leads firms to set a uniform price everywhere. This makes the online channel unambiguously better and directs most consumers towards that channel. With reduced transportation costs, the resulting market expansion increases profits for all firms. The second force concerns captive consumers. For these consumers, dealers would still like to price discriminate and extract more surplus. Since unrestricted consumers can shop both in person and online, firms settle for a compromise: they set a uniform online price, and discriminate slightly offline.

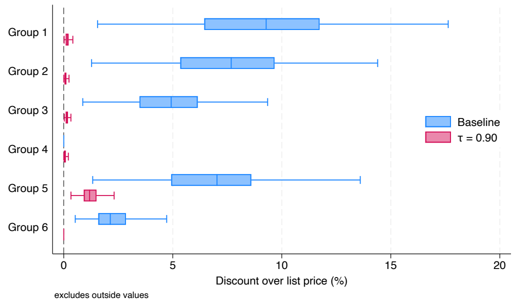
We plot the distribution of discounts offered as a function of the cost reduction in Fig-

Table 10: Effect of online channel with restricted access

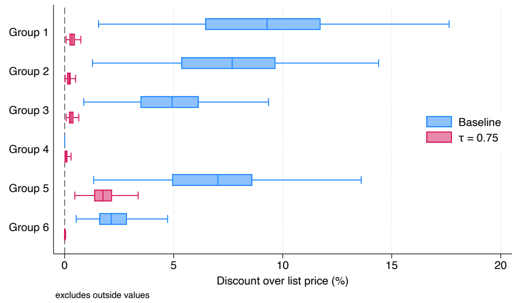
	In-person only	Both channels, restricted online access			
	Baseline	$\tau = 0.90$	$\tau = 0.75$	$\tau = 0.50$	$\tau = 0$
Transaction prices, uniform weights					
Group 1: Age < 40, Inc. $=$ Low	21,825	22,373	22,854	22,828	22,779
Group 2: Age < 40, Inc. $=$ High	22,119	22,903	22,883	22,880	22,867
Group 3: Age $\in [40, 60)$, Inc. $=$ Low	22,611	22,889	22,864	22,861	22,859
Group 4: Age $\in [40, 60)$, Inc. $=$ High	23,583	22,909	22,911	22,989	23,407
Group 5: Age ≥ 60, Inc. $=$ Low	22,222	22,664	22,561	22,489	22,373
Group 6: Age ≥ 60, Inc. $=$ High	23,118	22,916	22,921	22,955	22,985
Online		22,902	22,884	22,891	22,899
Transaction prices, sales-weighted					
Group 1: Age < 40, Inc. $=$ Low	20,585	21,087	21,293	21,315	21,445
Group 2: Age < 40, Inc. $=$ High	22,476	22,935	22,723	22,543	22,782
Group 3: Age $\in [40, 60)$, Inc. $=$ Low	22,555	22,661	22,647	22,656	22,720
Group 4: Age $\in [40, 60)$, Inc. $=$ High	24,736	23,754	23,706	23,926	24,596
Group 5: Age ≥ 60, Inc. $=$ Low	20,727	21,190	21,061	20,985	20,868
Group 6: Age ≥ 60, Inc. $=$ High	23,129	22,849	22,854	22,927	22,995
Online		22,561	22,776	22,845	23,135
Sales, in units					
Group 1: Age < 40, Inc. $=$ Low	101,254	-11,934	-12,056	-6,194	+8,953
Group 2: Age < 40, Inc. $=$ High	71,907	-7,309	-5,637	-2,332	+5,227
Group 3: Age $\in [40, 60)$, Inc. $=$ Low	135,790	-3,690	-520	+5,105	+18,832
Group 4: Age $\in [40, 60)$, Inc. $=$ High	232,803	+21,097	+27,765	+39,558	+65,070
Group 5: Age ≥ 60, Inc. $=$ Low	104,283	-6,580	-3,817	-34	+9,606
Group 6: Age ≥ 60, Inc. $=$ High	314,559	+9,677	+15,841	+25,831	+50,138
Proportion of online sales					
Group 1: Age < 40, Inc. $=$ Low	0	0.400	0.609	0.690	0.722
Group 2: Age < 40, Inc. $=$ High	0	0.590	0.755	0.873	0.910
Group 3: Age $\in [40, 60)$, Inc. $=$ Low	0	0.405	0.488	0.547	0.596
Group 4: Age $\in [40, 60)$, Inc. $=$ High	0	0.644	0.767	0.835	0.860
Group 5: Age ≥ 60, Inc. $=$ Low	0	0.096	0.143	0.180	0.250
Group 6: Age ≥ 60, Inc. $=$ High	0	0.404	0.482	0.529	0.562
Average distance to dealers, in km					
Group 1: Age < 40, Inc. $=$ Low	12.43	+0.21	+1.04	+2.30	+5.94
Group 2: Age < 40, Inc. $=$ High	13.76	+0.30	+0.71	+1.49	+3.45
Group 3: Age $\in [40, 60)$, Inc. $=$ Low	15.97	+0.33	+0.83	+1.83	+4.62
Group 4: Age $\in [40, 60)$, Inc. $=$ High	15.54	+0.33	+0.78	+1.68	+4.21
Group 5: Age ≥ 60, Inc. $=$ Low	16.27	+0.08	+0.35	+0.95	+3.20
Group 6: Age ≥ 60, Inc. $=$ High	13.52	+0.20	+0.54	+1.24	+3.50

Notes: All counterfactual experiments are computed using the 2019 data only. Unless indicated otherwise, in-person sales imply price discrimination and transportation costs and online sales imply a uniform price and reduced transportation costs. Access to the online channel in column (2) to (5) is restricted according to a survey of online purchases, see Table A.1. Transaction prices are in 2018 euros. "Uniform weights" are constructed using the total sales of each product in the baseline scenario, hence are fixed across demographic groups and counterfactual experiments. "Sales weights" use realized sales for each demographic group and counterfactual experiment. For sales and average distances, we report the values at baseline in the first column, and the change from baseline in the other columns.

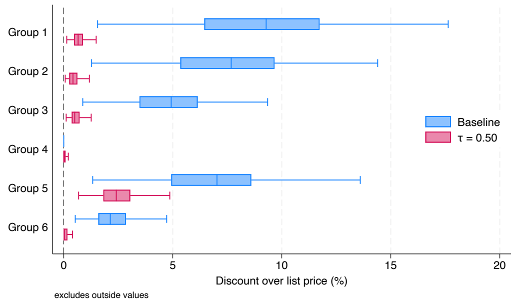
Figure 5: Price dispersion with restricted online access



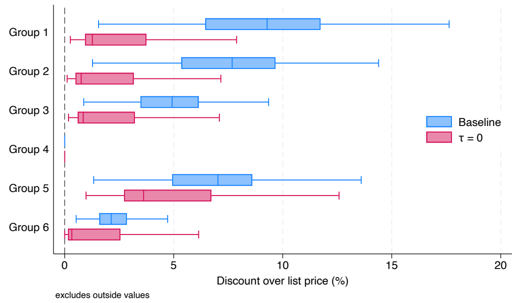
(a) Transportation costs = 90%



(b) Transportation costs = 75%



(c) Transportation costs = 50%



(d) No transportation cost

Notes: These figures represent the price dispersion in the in-person channel when consumers have a restricted access to the online channel, for varying transportation cost savings as per Table 10. Price dispersion is represented as a discount over the list price.

ure 5. An interesting pattern emerges. When the reduction in transportation costs is small, the online channel competes fiercely with in person sales for unrestricted consumers. In this context, firms do not have the leverage to price discriminate, since some unrestricted consumers would simply revert to shopping in person to get a discount. As transportation costs are reduced, purchasing online becomes more and more attractive for unrestricted consumers. This provides an opportunity for firms to reintroduce some level of price discrimination on captive consumers, since the reduction in transportation costs is so large that it offsets discounts from the point of view of unrestricted buyers. In some sense, the higher the reduction in transportation costs, the easier it is for firms to separate captive and non-captive consumers and reintroduce price discrimination.

Table 11: Cross-firm effects on price dispersion

	In person only	Both channels, restricted online access			
	Baseline	$\tau = 0.90$	$\tau = 0.75$	$\tau = 0.50$	$\tau = 0$
Transaction prices – Nissan-Renault group					
Group 1: Age < 40, Inc. $=$ Low	16,547	17,633	17,622	17,631	17,619
Group 2: Age < 40, Inc. $=$ High	16,844	17,661	17,694	17,680	17,713
Group 3: Age $\in [40, 60)$, Inc. $=$ Low	17,398	17,650	17,668	17,658	17,683
Group 4: Age $\in [40, 60)$, Inc. $=$ High	18,389	17,666	17,785	17,739	17,832
Group 5: Age ≥ 60, Inc. $=$ Low	17,038	17,461	17,271	17,312	17,239
Group 6: Age ≥ 60, Inc. $=$ High	17,937	17,679	17,776	17,747	17,802
Online		17,657	17,709	17,688	17,735
Transaction prices – Other manufacturers					
Group 1: Age < 40, Inc. $=$ Low	23,828	23,828	23,828	23,828	23,828
Group 2: Age < 40, Inc. $=$ High	24,120	24,121	24,121	24,121	24,121
Group 3: Age $\in [40, 60)$, Inc. $=$ Low	24,589	24,589	24,589	24,589	24,589
Group 4: Age $\in [40, 60)$, Inc. $=$ High	25,553	25,552	25,552	25,552	25,552
Group 5: Age ≥ 60, Inc. $=$ Low	24,189	24,190	24,190	24,190	24,189
Group 6: Age ≥ 60, Inc. $=$ High	25,084	25,083	25,083	25,083	25,083

Notes: To do. Only Nissan-Renault group offers an online sales channel. Basically the price dispersion effect is mostly to prevent price canibalization with no effet on other firms pricing. All prices are computed using a uniform set of weights, $w_j = \sum_d \phi_d s_{jd} / \sum_{k \in \mathcal{J}_g} \sum_d \phi_d s_{kd}$, where $\mathcal{J}_g$ is the set of products offered by group $g = \{\text{Nissan-Renault}, \text{Other}\}$.

5.3 Cross-firm effects

In the previous sections, we have provided evidence that introducing online sales at a fixed prices reduces price dispersion in the offline channel. We investigate deeper to better understand these dynamics. In particular, we want to know if the reduction in price dispersion is a within-firm effect (e.g., a firm's online channel puts pressure on that firms offline prices) or an across-firm effect (e.g., a firm's online channel puts pressure on other firms offline prices). To achieve this, we consider a counterfactual scenario where one large manufacturer starts selling online and his competitors are restricted to sell in person only. We pick the Nissan-Renault group (which also includes the Dacia and Mitsubishi brands) as our candidate online firm and re-evaluate counterfactual prices for all firms, for varying levels of transportation cost reductions.

The resulting prices are presented in Table 11. We notice in the first panel that introducing the online channel reduces price dispersion in the offline channel as previously. We take this as evidence that firms reduce price dispersion to avoid cannibalizing their own online sales. We do not observe the same effect in other firms, as evidence from the second panel of Table 11. Hence, it is profitable for firms to continue to price discriminate when competitors start selling online. In this case, these firms lose sales to the competitors online channel, but there is nothing they can do in terms of price-setting to offset these losses. For completeness, we also provide the distribution of unobserved discounts by group, see Figure A.1.

Table 12: Effect of online channel on welfare

Counterfactual	Δ Consumer surplus, per capita							Δ Profits
	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6	All	Total
Baseline	411.2	589.7	1,053.5	1,757.6	1,051.7	1,823.7	1,136.1	6,229.8
Restricted access to online channel								
• $\tau = 0.90$	-48.5	-60.7	-22.5	+191.3	-60.0	+84.1	+23.5	-26.8
• $\tau = 0.75$	-45.6	-45.4	+20.9	+262.8	+22.9	+163.1	+70.8	+140.2
• $\tau = 0.50$	-14.8	-14.8	+102.3	+395.3	+185.5	+303.3	+164.2	+404.9
• $\tau = 0$	+66.4	+55.9	+307.8	+708.1	+677.3	+649.2	+403.2	+1,032.3

Notes: All counterfactual experiments are computed using the 2019 data only. All counterfactual have a restricted access to the online channel. Consumer surplus is in 2018 euros. Profits are in million 2018 euros.

5.4 Welfare analysis

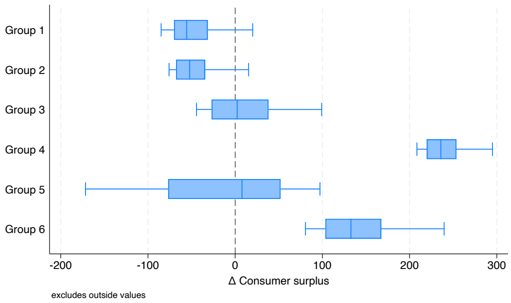
We finally focus our attention on the welfare consequences of the introduction of an online sales channel in the French car market. Our welfare analysis is based on our preferred counterfactual experiment which restricts the access to the online channel for some consumers. We look both at aggregate (group-specific) effects and distributional effects. We begin with aggregate effects, presented in Table 12.

We consider the change in consumer surplus. We uncover some heterogeneity with respect to average consumer gains and losses from the online channel. We can split consumers roughly in three groups: the winners, group 4 and 6 (high-income, aged 40 and above), the losers, group 1 and 2 (all consumers aged below 40), and the rest, formed by group 3 and 5 (low-income, aged 40 and above).

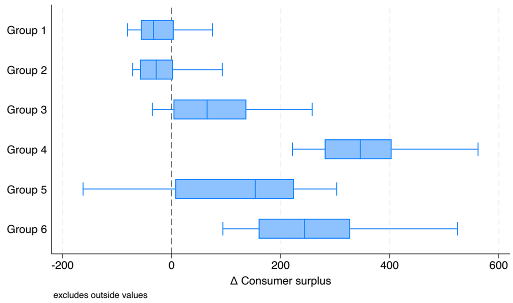
The first group benefits largely from online shopping. They receive both a lower price and reduced transportation costs compared to the baseline. Their surplus can increase by as much as 40% under the most favorable transportation cost reductions. The second group either experience a small loss when the reduction in transportation costs is modest, or a small gain when it is large. These consumers are not very sensitive to driving distance, hence they are mostly affected by the loss of the discount they received in the baseline. The impact of the online channel on their surplus is limited to be between -10% and +10%, depending on the transportation costs reduction.

Our final group is different. They are sensitive to both prices and distance. For the smallest reduction in transportation costs under consideration, their realized gain in consumer surplus is small, around 2%. At that level of transportation costs reduction, the loss of the discount limits their gains in consumer surplus, even though they have a strong distaste for distance. As transportation costs decrease, they gain greatly in terms of surplus. In the most extreme case (i.e., group 5), surplus can increase by as much as 64.4%.

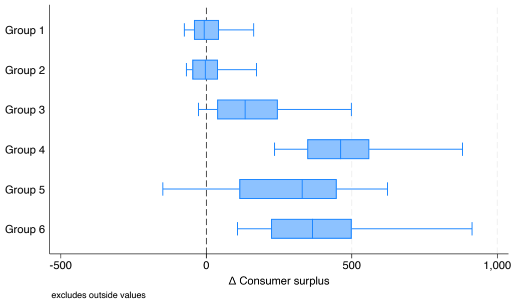
Figure 6: Change in per capita consumer surplus from online store



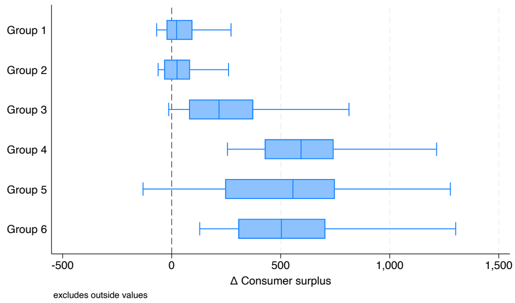
(a) Transportation costs = 90%



(b) Transportation costs = 75%



(c) Transportation costs = 50%



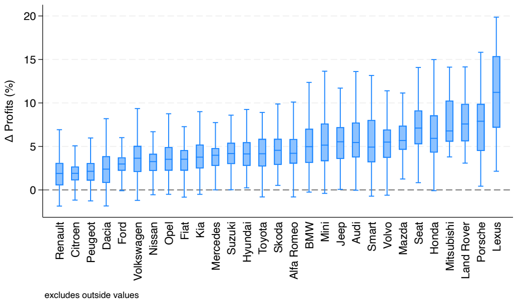
(d) No transportation cost

Notes: These figures represent the change in consumer surplus from introducing the online channel for varying transportation cost savings as per Table 10. Consumer surplus is the average per capita consumer surplus at the municipal level, and it's distribution is weighted by group-specific populations.

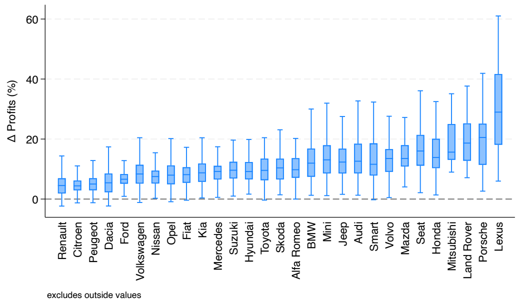
These aggregate figures hide non-negligible heterogeneity within groups. Figure 6 plots the underlying distribution of consumer surplus at the consumer level, by group. In all scenarios considered, a large share of young individuals (groups 1 and 2) experience a decrease in consumer surplus. This is true even when we completely eliminate transportation costs: more than 25% of these individuals are net losers in that case. Groups 4 and 6 (high-income, aged 40 or older) are net winners in all scenarios.

Group 5 (low-income, aged 60 and above) shows the most heterogeneity. This follows from the fact that they both have a very high distaste for traveling and a high price sensitivity. Although most of the consumers in these groups benefit from the introduction of the online sales channel, this is offset by a small number of consumers who experience very large losses. Understanding these distributional effects is important, as these consumers form the most economically vulnerable group from a demographic point of view.

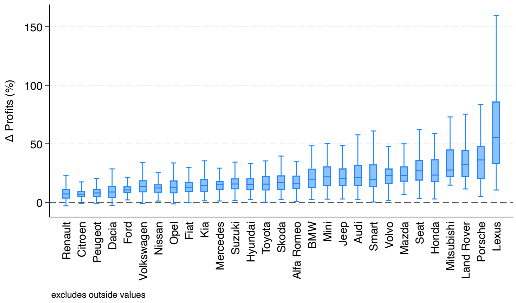
Figure 7: Change in dealer-level profits, online sales reallocated



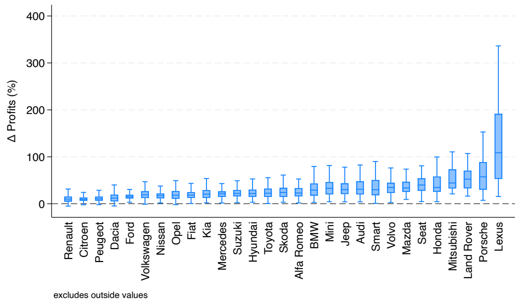
(a) Transportation costs = 90%



(b) Transportation costs = 75%



(c) Transportation costs = 50%



(d) No transportation cost

Notes: These figures represent the change profits at the dealer level from introducing the online channel as per Table 10. Brands are ranked by the total number of dealers, in decreasing order. Profits from the online channel are allocated to each physical store based on consumers' location.

Finally, we consider both industry and dealer-level profits. Industry profits increase for any level of transportation costs reduction. When transportation costs are completely eliminated by the online channel, profits can increase by as much a €1 billion, which corresponds to a 16.7% increase.

Figure 7 presents a breakdown by dealership at the brand-level. To provide the most accurate representation, we reallocate all profits from the online channel to the closest dealers (as if cars where bought in person). We rank firms by their market presence, starting with Renault which operates the largest network and is closest to consumers. We uncover an interesting pattern: firms further away from consumers experience larger increases in profits (in percentage). This aligns with intuition. Once transportation costs are reduced or removed, consumers respond by shifting some of their demands towards products they like more, which sometimes are only sold by dealers located further away.

When we focus solely on profits from the in-person channel, all dealers experience large losses, in the range of 50% or more. This scenario would occur if for example the manufacturer diverted online sales to some new "online only dealer." We believe that this scenario is unlikely, as it would lead to a large push-back from dealers (e.g., some may sue) and potentially end their business altogether. We report on these results in Figure A.2 for completeness.

5.5 Robustness analysis

As a final experiment, we assess the robustness of our welfare analysis to the assumptions that the online channel does not lead to entry or exit of dealers and that it doesn't change the vertical relations between dealers and manufacturers.

One plausible consequence to the introduction of an online sales channel is that it may drive some dealers out of business, which means that we would overstate the associated welfare gains. Since we do not model entry decisions explicitly, we proceed by closing a certain number of stores and re-evaluating welfare in this new environment. We base this experiment on our counterfactual with restricted access to the online dealer and $\tau = 0.50$ (see Table 10, column 4). We focus on the welfare results, presented in Table 13. The detailed counterfactual results are relayed to the Appendix, Table A.6. We consider three experiments: closing the 5% least profitable dealers, the 10% least profitable dealers, and the 25% least profitable dealers.

These experiments reveal that our estimates of consumer surplus are relatively robust to the exit of dealers. We focus on the most extreme scenario where 25% of the dealers exit following the introduction of the online channel. In this case, the loss in consumer surplus range from €6.7 to €39.4 per consumer depending on the demographic group. The average derease over all consumers is around €23 per consumer per year. Profits decrease by €137 million in the worst case scenario. Meanwhile, the average gains in surplus when no dealer exits is around €168 and the total gain in profits is around €405 million. Closing the 25% least profitable dealers thus mitigates around 14% of the realized gains to consumer surplus and 34% of firms realized gains in profits. For a more realistic market reallocation (e.g., 10% exit or less), these effects are of 4% and of 9%, respectively.

Another plausible consequence of the introduction of online sales is that manufacturers could try to eliminate the middleman and sell directly to consumers. We assume that bypassing dealers leads to cost efficiencies, in the sense that the manufacturer will base his pricing decision on the true marginal cost which is lower than the wholesale price

Table 13: Effect of reducing market presence on welfare

Counterfactual	Δ Consumer surplus, per capita							Δ Profits
	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6	All	Total
All stores	396.4	574.9	1,155.8	2,152.8	1,237.2	2,127.1	1,300.4	6,634.7
Reduced market presence								
• 5% fewer stores	-0.5	-0.9	-1.3	-3.7	-2.3	-4.2	-2.2	-13.2
• 10% fewer stores	-1.6	-2.2	-4.5	-9.4	-8.0	-10.8	-6.1	-36.5
• 25% fewer stores	-6.7	-7.4	-18.6	-33.3	-34.2	-39.4	-23.1	-137.4

Notes: All counterfactual experiments are computed using the 2019 data only. All counterfactual have a restricted access to the online channel as described in Section 5.2 and $\tau = 0.5$, as per Table 10, column 3. The first row reports the results from under a counterfactual where all stores are present. Other rows represent the difference with the first row. Consumer surplus is in 2018 euros. Profits are in million 2018 euros.

Table 14: Effect of cost efficiencies on welfare

Counterfactual	Δ Consumer surplus, per capita							Δ Profits
	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6	All	Total
No cost efficiencies	396.4	574.9	1,155.8	2,152.8	1,237.2	2,127.1	1,300.4	6,634.7
Cost efficiencies for online sales								
• 5% lower marg. cost	+63.3	+95.7	+164.4	+268.4	+158.1	+264.9	+171.6	+493.1
• 10% lower marg. cost	+132.4	+200.7	+343.0	+550.2	+334.4	+538.7	+354.0	+1,058.2
Cost efficiencies and delivery cost for online sales								
• 5% lower marg. cost + €400 delivery cost	+30.9	+51.1	+86.4	+149.3	+75.4	+139.5	+90.5	+255.1
• 10% lower marg. cost + €400 delivery cost	+98.1	+153.1	+261.3	+429.0	+246.3	+412.1	+270.3	+802.0

Notes: All counterfactual experiments are computed using the 2019 data only. All counterfactual have a restricted access to the online channel as described in Section 5.2 and $\tau = 0.5$, as per Table 10, column 3. The first row reports the results from under a counterfactual where online sales to not generate cost efficiencies or delivery costs. Other rows represent the difference with the first row. Consumer surplus is in 2018 euros. Profits are in million 2018 euros.

under double marginalization. Since we do not model vertical relations or the wholesale price explicitly, we assume that selling directly to consumers entails a small reduction in marginal cost, in the range of 5 or 10%.[15]

Additionally, we allow for manufacturers to incur an additional cost for delivering the car to consumers' door. To evaluate these costs, we use an online platform specialized in car deliveries, Shiply.com,[16] and ask quotes for various vehicle deliveries for a selection of city pairs in France (shortest distance inquired: 75km, longest distance inquired: 250km). All quotes returned a price between 300 and 500 euros for a single car delivery. We opt for an average shipping cost of 400 euros. The welfare results are presented in Table 14. The detailed counterfactual results are relayed to the Appendix, Table A.5.

Our results are not very robust to changes in the vertical contracts between dealers and manufacturers, leading to large discrepancies in terms of consumer surplus and industry profits. Given that profits from the online channels are redistributed to dealers, our

[15] To fix ideas, in their analysis of vertical relations in the european car market, Brenkers and Verboven (2006) estimate marginal costs to be 7–8% lower under double marginalization that under the traditional assumption that the retail market is perfectly competitive. We use this figure as a basis for determining an appropriate cost reduction.

[16] Source: https://www.shiply.com.

results understate the gains from online sales.

6 Conclusion

In this paper, we have investigated the consequences of the introduction of an online sales channel in the context of the French car market. We focus on the case where firms sell online at the fixed price advertised on the website, but price discriminate in-person by offering personalized discounts to buyers based on their observable characteristics. We propose a structural model of demand with unobserved price discrimination and transportation costs to study firms pricing behavior in this context and the associated welfare effects on consumers.

We show that committing to a uniform price online reduces the extent of price discrimination in-person, as firms do not want to cannibalize their online sales by offering discounts in person. Price transparency is a long standing issue with the car market, and our results imply that committing to sell at a fixed price online may be enough to enforce industry-wide price transparency.

In terms of welfare, we finds that the introduction of an online sales channel benefits some consumers and harm others. These gains and losses depend on two crucial factors: first, on the size of the personalized discounts consumers received before the online channel became available; second, on the size of the reduction in transportation costs that the online channel provides. In the extreme case where buying online eliminates transportation costs altogether, most consumers are net winners. Otherwise, adding the online channel has a redistributive effect on consumer surplus, benefiting consumers with low price sensitivity and high sensitivity to traveling distance (in our context, high-income consumers). Finally, we find that selling online induces a market expansion and increases all car dealers' profits.

We highlight two caveats in our analysis. First, we assume throughout that the introduction of online sales does not generate the entry or exit of dealers. Second, we assume that selling online directly to consumers does not change the vertical contract between the manufacturer and the dealers. We leave these important questions for future work.

References

Ayres, I. and Siegelman, P. (1995). Race and gender discrimination in bargaining for a new car. *The American Economic Review*, pages 304–321.

Berry, S., Levinsohn, J., and Pakes, A. (1995). Automobile prices in market equilibrium. *Econometrica: Journal of the Econometric Society*, pages 841–890.

Berry, S. T. (1994). Estimating discrete-choice models of product differentiation. *The RAND Journal of Economics*, pages 242–262.

Brenkers, R. and Verboven, F. (2006). Liberalizing a distribution system: the european car market. *Journal of the European Economic Association*, 4(1):216–251.

Brown, J. R. and Goolsbee, A. (2002). Does the internet make markets more competitive? evidence from the life insurance industry. *Journal of political economy*, 110(3):481–507.

Brynjolfsson, E., Hu, Y., and Smith, M. D. (2003). Consumer surplus in the digital economy: Estimating the value of increased product variety at online booksellers. *Management science*, 49(11):1580–1596.

Cardell, N. S. (1997). Variance components structures for the extreme-value and logistic distributions with application to models of heterogeneity. *Econometric Theory*, 13(2):185–213.

Chandra, A., Gulati, S., and Sallee, J. M. (2017). Who loses when prices are negotiated? an analysis of the new car market. *The Journal of Industrial Economics*, 65(2):235–274.

Corts, K. S. (1998). Third-degree price discrimination in oligopoly: All-out competition and strategic. *The RAND Journal of Economics*, 29:306–323.

D'Haultfœuille, X., Durrmeyer, I., and Février, P. (2019). Automobile prices in market equilibrium with unobserved price discrimination. *The Review of Economic Studies*, 86(5):1973–1998.

Dubé, J.-P. and Misra, S. (2023). Personalized pricing and consumer welfare. *Journal of Political Economy*, 131(1):131–189.

Duch-Brown, N., Grzybowski, L., Romahn, A., and Verboven, F. (2023). Evaluating the impact of online market integration—evidence from the eu portable pc market. *American Economic Journal: Microeconomics*, 15(4):268–305.

Fan, J., Tang, L., Zhu, W., and Zou, B. (2018). The alibaba effect: Spatial consumption inequality and the welfare gains from e-commerce. *Journal of International Economics*, 114:203–220.

Feenstra, R. C., Inklaar, R., and Timmer, M. P. (2015). The next generation of the penn world table. *American economic review*, 105(10):3150–3182.

Forman, C., Ghose, A., and Goldfarb, A. (2009). Competition between local and electronic markets: How the benefit of buying online depends on where you live. *Management science*, 55(1):47–57.

Goldberg, P. K. (1996). Dealer price discrimination in new car purchases: Evidence from the consumer expenditure survey. *Journal of Political Economy*, 104(3):622–654.

Grieco, P., Murry, C., and Yurukoglu, A. (2023). The evolution of market power in the us automobile industry. *The Quarterly Journal of Economics*.

Harless, D. W. and Hoffer, G. E. (2002). Do women pay more for new vehicles? evidence from transaction price data. *American Economic Review*, 92(1):270–279.

Huang, Y. and Bronnenberg, B. J. (2023). Consumer transportation costs and the value of e-commerce: Evidence from the dutch apparel industry. *Marketing Science*, 42(5):984–1003.

Iaria, A. and Wang, A. (2021). An empirical model of quantity discounts with large choice sets. *Available at SSRN 3946475*.

Miller, N. H. and Osborne, M. (2014). Spatial differentiation and price discrimination in the cement industry: evidence from a structural model. *The RAND Journal of Economics*, 45(2):221–247.

Moraga-González, J. L., Sándor, Z., and Wildenbeest, M. R. (2023). Consumer search and prices in the automobile market. *The Review of Economic Studies*, 90(3):1394–1440.

Morrow, W. R. and Skerlos, S. J. (2011). Fixed-point approaches to computing bertrand-nash equilibrium prices under mixed-logit demand. *Operations research*, 59(2):328–345.

Morton, F. S., Zettelmeyer, F., and Silva-Risso, J. (2001). Internet car retailing. *The Journal of Industrial Economics*, 49(4):501–519.

Murry, C. and Zhou, Y. (2020). Consumer search and automobile dealer colocation. *Management Science*, 66(5):1909–1934.

Nurski, L. and Verboven, F. (2016). Exclusive dealing as a barrier to entry? evidence from automobiles. *The Review of Economic Studies*, 83(3):1156–1188.

Pozzi, A. (2013). The effect of internet distribution on brick-and-mortar sales. *The RAND Journal of Economics*, 44(3):569–583.

Rhodes, A. and Zhou, J. (2024). Personalized pricing and competition. *American Economic Review*, 114(7):2141–2170.

Sagl, S. (2024). Dispersion, discrimination, and the price of your pickup. *Working paper.*

Scott Morton, F., Silva-Risso, J., and Zettelmeyer, F. (2011). What matters in a price negotiation: Evidence from the us auto retailing industry. *Quantitative Marketing and Economics*, 9:365–402.

Shiller, B. R. (2020). Approximating purchase propensities and reservation prices from broad consumer tracking. *International Economic Review*, 61(2):847–870.

Thisse, J.-F. and Vives, X. (1988). On the strategic choice of spatial price policy. *The American Economic Review*, pages 122–137.

Yavorsky, D., Honka, E., and Chen, K. (2021). Consumer search in the us auto industry: The role of dealership visits. *Quantitative Marketing and Economics*, 19:1–52.

A Additional Tables and Figures

Table A.1: Demographic characteristics by group

Description	Mean	Std. dev.	10th pct.	Median	90th pct.	Observations
Group 1: Age < 40, *Income = Low*						
Median income	16,745	2,680	13,052	17,201	19,795	211,397
Average age	26.9	1.0	25.8	26.9	28.0	211,397
Share of female	0.499	0.035	0.468	0.503	0.524	211,397
Average household size	2.22	0.27	1.92	2.16	2.58	211,397
Urban	0.457	0.498	0	0	1	211,397
Shop online (ψ_1)	0.770	0.421	0	1	1	9,541
Share of population (ϕ_1)	0.247					13
Group 2: Age < 40, *Income = High*						
Median income	24,840	4,246	20,832	23,362	31,646	189,754
Average age	27.5	1.1	26.4	27.5	28.7	189,754
Share of female	0.503	0.040	0.462	0.509	0.536	189,754
Average household size	2.35	0.26	1.97	2.38	2.67	189,754
Urban	0.459	0.498	0.000	0.000	1.000	189,754
Shop online (ψ_2)	0.894	0.308	0.000	1.000	1.000	12,131
Share of population (ϕ_2)	0.145					13
Group 3: Age $\in [40, 60)$, *Income = Low*						
Median income	17,891	2,248	15,019	18,388	20,237	195,694
Average age	49.5	0.7	48.9	49.5	50.2	195,694
Share of female	0.510	0.037	0.469	0.517	0.540	195,694
Average household size	2.23	0.27	1.92	2.18	2.59	195,694
Urban	0.367	0.482	0	0	1	195,694
Shop online (ψ_3)	0.505	0.500	0	1	1	14,098
Share of population (ϕ_3)	0.151					13
Group 4: Age $\in [40, 60)$, *Income = High*						
Median income	25,252	5,179	20,858	23,661	31,944	205,510
Average age	49.4	0.7	48.7	49.4	50.1	205,510
Share of female	0.511	0.031	0.474	0.516	0.540	205,510
Average household size	2.33	0.25	1.96	2.36	2.65	205,510
Urban	0.354	0.478	0	0	1	205,510
Shop online (ψ_4)	0.770	0.421	0	1	1	19,614
Share of population (ϕ_4)	0.185					13
Group 5: Age ≥ 60, *Income = Low*						
Median income	18,842	1,568	16,796	19,081	20,600	192,170
Average age	70.5	1.2	69.1	70.5	71.8	192,170
Share of female	0.545	0.050	0.484	0.553	0.596	192,170
Average household size	2.24	0.25	1.94	2.21	2.55	192,170
Urban	0.196	0.397	0	0	1	192,170
Shop online (ψ_5)	0.146	0.353	0	0	1	20,034
Share of population (ϕ_5)	0.093					13
Group 6: Age ≥ 60, *Income = High*						
Median income	24,673	4,711	20,932	23,313	29,837	208,943
Average age	70.1	1.0	68.9	70.2	71.2	208,943
Share of female	0.551	0.040	0.500	0.558	0.592	208,943
Average household size	2.25	0.26	1.92	2.23	2.59	208,943
Urban	0.401	0.490	0	0	1	208,943
Shop online (ψ_6)	0.464	0.499	0	0	1	17,724
Share of population (ϕ_6)	0.179					13

Notes: Statistics concerning the median income, age, household size, the proportion of female, and the level of urbanity are weighted by municipal-level group-specific populations. Statistics concerning the propensity to shop online are weighted by survey weights. For future reference, we denote the probability to shop online by ψ_d (see Section 5). We report a simple year-over-year average of the group-specific population shares, denoted by ϕ_d.

Table A.2: Evidence of price dispersion, by demographic group

	Transaction price		Transaction price - buyback value	
	(1)	(2)	(3)	(4)
Group 1: Age < 40, Income = Low	— Base category (omitted) —			
Group 2: Age < 40, Income = High	1,891.780**	1,672.480*	2,268.324***	2,303.590*
	(744.064)	(910.848)	(710.450)	(1,215.509)
Group 3: Age $[40, 60)$, Income = Low	2,174.785***	1,891.699**	2,766.616***	1,472.772
	(611.006)	(899.649)	(708.275)	(1,307.562)
Group 4: Age $[40, 60)$, Income = High	3,734.329***	3,482.132**	2,961.747***	1,845.544
	(826.167)	(1,418.426)	(853.483)	(1,763.667)
Group 5: Age $\geq g0$, Income = Low	2,412.376***	1,843.251*	1,498.397**	80.061
	(640.156)	(959.263)	(690.785)	(1,148.006)
Group 6: Age ≥ 60, Income = High	2,861.136***	811.124	1,725.900**	-477.457
	(773.938)	(832.586)	(803.734)	(1,130.036)
Female	-306.879	-571.473	-510.496	-516.426
	(346.764)	(487.459)	(372.094)	(621.099)
Value of down payment	0.007	0.006*	0.012	0.011**
	(0.005)	(0.003)	(0.008)	(0.005)
Household: 2 pers.	63.960	-297.050	-133.333	-561.202
	(387.543)	(598.280)	(479.113)	(820.017)
Household: 3 pers.	-151.740	-942.284	-551.323	-1,268.167
	(602.009)	(815.275)	(660.687)	(1,140.575)
Household: 4 pers.	117.470	-1,524.256*	-1,019.702*	-2,687.011**
	(548.932)	(785.119)	(550.327)	(1,132.520)
Household: 5 pers.	-2,166.131***	-1,871.581	-2,437.147***	-4,361.964***
	(794.271)	(1,161.155)	(874.917)	(1,666.250)
Household: 6+ pers.	1,709.236	-1,236.283	3,163.307	-315.518
	(3,021.293)	(1,943.145)	(2,911.596)	(1,961.040)
Urban area: less than 15,000	-1,016.912	2,739.753*	-1,576.286	3,032.271
	(1,326.101)	(1,429.939)	(2,297.180)	(3,751.639)
Urban area: 15,000–24,999	288.247	1,084.762	1,611.884	2,909.341
	(1,581.361)	(1,614.560)	(1,441.358)	(2,866.254)
Urban area: 25,000–34,999	-1,493.065	1,596.119	-1,166.079	1,902.540
	(1,370.082)	(1,903.312)	(1,654.036)	(2,956.974)
Urban area: 35,000–49,999	-1,823.739	-132.945	-1,590.025	1,627.714
	(1,187.328)	(975.182)	(1,281.310)	(1,626.837)
Urban area: 50,000–99,999	-1,339.540	-528.954	-2,156.797***	-1,535.573
	(837.413)	(1,113.736)	(795.857)	(1,116.164)
Urban area: 100,000–199,999	-760.895	316.345	-196.287	325.665
	(803.541)	(876.063)	(699.684)	(998.607)
Urban area: 200,000–499,999	-943.257	215.036	-700.908	84.523
	(713.562)	(835.990)	(617.439)	(1,019.786)
Urban area: 500,000 or more	-971.024	-55.535	-626.253	373.406
	(642.346)	(828.815)	(620.582)	(873.457)
Urban area: Paris greater metro area	-583.891	-67.373	280.185	748.892
	(719.717)	(991.942)	(661.801)	(919.323)
New vehicles only	No	Yes	No	Yes
Fixed effects Car model × engine × new	Yes	Yes	Yes	Yes
Year of purchase × month	Yes	Yes	Yes	Yes
Country of origin of buyer	Yes	Yes	Yes	Yes
Fstat	5.41	1.47	3.96	1.61
Pr(Fstat) > F	< 0.001	0.205	0.002	0.164
Observations	1,283	698	1,283	698
R-squared	0.740	0.801	0.600	0.620

Notes: This Table represents the result of a regression of transaction prices on demographic group dummies and other demographic characteristics of buyers, based on a survey of consumers' expenditure. We have excluded observations where the car was purchased following an insurance claim (i.e., the replacement of a damaged vehicle). Columns (1), (3), and (5) include sales of both new and used cars, purchased at a car dealer. Columns (2) and (4) include only new cars. the buyback value represents the payment that was received by the consumer for trading in his old car. The F-statistic tests for the hypothesis that the coefficients on the group dummies are jointly zero. Standard errors in parenthesis are clustered at the car model × engine × new/used level. Significance: * < 0.10, ** < 0.05, *** < 0.01.

Table A.3: Car dealer's market presence, by brand

Brand	Stores	Market share	Distance to consumers, in km				
		(%)	Mean	Std. dev.	10th pct.	Median	90th pct.
Renault	433	17.8	14.40	14.49	0.19	9.73	34.10
Citroen	400	12.6	13.62	13.05	0.38	10.06	30.89
Peugeot	367	15.9	14.89	14.45	0.28	10.82	34.85
Dacia	331	8.5	18.19	18.73	0.56	12.36	42.82
Ford	258	4.6	16.27	14.10	1.26	12.02	36.02
Volkswagen	212	7.2	20.71	19.45	0.70	14.26	49.50
Nissan	198	2.9	19.80	18.33	1.23	14.19	45.97
Opel	193	3.6	21.98	21.06	1.29	15.00	51.36
Fiat	191	2.7	20.57	19.42	0.65	14.52	48.21
Kia	172	2.3	23.13	22.57	1.87	15.20	56.11
Mercedes	165	2.1	20.98	18.45	2.67	15.21	47.01
Suzuki	161	1.6	23.62	20.49	3.24	17.50	52.79
Hyundai	158	1.7	23.59	21.07	2.56	16.76	55.84
Toyota	151	5.2	24.72	22.00	1.82	18.69	57.45
Skoda	139	1.1	25.40	23.19	3.65	17.57	59.56
Alfa Romeo	129	0.4	29.06	27.22	2.52	20.19	69.39
BMW	104	1.8	35.26	34.44	3.29	23.27	84.62
Mini	101	1.4	35.83	34.41	3.20	24.13	84.62
Jeep	99	0.2	32.38	28.99	3.77	22.80	75.99
Audi	95	2.5	35.46	31.84	4.19	24.05	80.21
Smart	93	0.2	34.92	32.48	3.95	23.23	84.62
Volvo	89	0.5	35.26	31.93	4.60	24.37	80.97
Mazda	88	0.5	35.24	28.84	6.00	27.77	78.80
Seat	87	1.6	35.56	27.23	6.99	28.76	74.42
Honda	69	0.5	38.82	34.19	4.80	27.51	88.87
Mitsubishi	62	0.2	46.31	40.56	7.77	33.25	106.1
Land Rover	50	0.3	53.57	47.64	6.66	37.91	129.6
Porsche	34	0.1	57.68	49.46	8.78	43.32	132.3
Lexus	20	0.2	110.0	97.47	8.61	83.50	249.0
TOTAL	4,649	100					

Notes: Brands are ordered by their market presence, defined by the total number of dealers in 2020. The market share is computed as each brand's sales over total sales. Statistics on the distribution of driving distances are weighted by population.

Table A.4: Effect of reducing market presence

	All stores	Reduced market presence		
		-5%	-10%	-25%
Transaction prices, uniform weights				
Group 1: Age < 40, Inc. $=$ Low	22,828	22,827	22,827	22,824
Group 2: Age < 40, Inc. $=$ High	22,880	22,880	22,880	22,878
Group 3: Age $\in [40, 60)$, Inc. $=$ Low	22,861	22,860	22,860	22,858
Group 4: Age $\in [40, 60)$, Inc. $=$ High	22,989	22,990	22,990	22,994
Group 5: Age ≥ 60, Inc. $=$ Low	22,489	22,488	22,487	22,482
Group 6: Age ≥ 60, Inc. $=$ High	22,955	22,955	22,955	22,954
Online		22,891	22,890	22,889
Transaction prices, sales-weighted				
Group 1: Age < 40, Inc. $=$ Low	21,315	21,309	21,302	21,278
Group 2: Age < 40, Inc. $=$ High	22,543	22,543	22,562	22,585
Group 3: Age $\in [40, 60)$, Inc. $=$ Low	22,656	22,647	22,630	22,586
Group 4: Age $\in [40, 60)$, Inc. $=$ High	23,926	23,922	23,932	23,918
Group 5: Age ≥ 60, Inc. $=$ Low	20,985	20,976	20,953	20,895
Group 6: Age ≥ 60, Inc. $=$ High	22,927	22,917	22,906	22,857
Online		22,839	22,828	22,793
Sales, in units				
Group 1: Age < 40, Inc. $=$ Low	95,060	-129	-418	-1,757
Group 2: Age < 40, Inc. $=$ High	69,575	-111	-272	-920
Group 3: Age $\in [40, 60)$, Inc. $=$ Low	140,895	-198	-660	-2,717
Group 4: Age $\in [40, 60)$, Inc. $=$ High	272,361	-486	-1,225	-4,278
Group 5: Age ≥ 60, Inc. $=$ Low	104,249	-269	-911	-3,898
Group 6: Age ≥ 60, Inc. $=$ High	340,390	-817	-2,108	-7,547
Proportion of online sales				
Group 1: Age < 40, Inc. $=$ Low	0.690	0.690	0.690	0.692
Group 2: Age < 40, Inc. $=$ High	0.873	0.873	0.873	0.874
Group 3: Age $\in [40, 60)$, Inc. $=$ Low	0.547	0.547	0.548	0.550
Group 4: Age $\in [40, 60)$, Inc. $=$ High	0.835	0.836	0.836	0.837
Group 5: Age ≥ 60, Inc. $=$ Low	0.180	0.180	0.181	0.182
Group 6: Age ≥ 60, Inc. $=$ High	0.529	0.529	0.530	0.532
Average distance to dealers, in km				
Group 1: Age < 40, Inc. $=$ Low	14.73	+0.09	+0.27	+1.16
Group 2: Age < 40, Inc. $=$ High	15.24	+0.21	+0.49	+1.62
Group 3: Age $\in [40, 60)$, Inc. $=$ Low	17.80	+0.11	+0.36	+1.50
Group 4: Age $\in [40, 60)$, Inc. $=$ High	17.22	+0.21	+0.52	+1.80
Group 5: Age ≥ 60, Inc. $=$ Low	17.22	+0.06	+0.20	+0.90
Group 6: Age ≥ 60, Inc. $=$ High	14.77	+0.17	+0.42	+1.51

Notes: All counterfactual experiments are computed using the 2019 data only. All counterfactual have a restricted access to the online channel and a 50% reduced transportation cost online. Transaction prices are in 2018 euros. "Uniform weights" are constructed using the total sales of each product in the baseline scenario, hence are fixed across demographic groups and counterfactual experiments. "Sales-weights" use realized sales for each demographic group and counterfactual experiment. For sales and average distances, we report the values at baseline in the first column, and the change from baseline in the other columns. We reduce the market presence of brands by closing 5, 10, or 25% of the less profitable dealers respectively.

Table A.5: Effect of cost efficiencies and delivery costs

	No cost efficiencies	Cost efficiencies, w/o delivery cost		Cost efficiencies, with delivery cost	
		-5%	-10%	-5%	-10%
Transaction prices, uniform weights					
Group 1: Age < 40, Inc. $=$ Low	22,828	22,362	22,029	22,591	22,210
Group 2: Age < 40, Inc. $=$ High	22,880	22,477	22,204	22,674	22,360
Group 3: Age $\in [40, 60)$, Inc. $=$ Low	22,861	22,638	22,597	22,736	22,633
Group 4: Age $\in [40, 60)$, Inc. $=$ High	22,989	23,878	23,646	23,488	23,664
Group 5: Age ≥ 60, Inc. $=$ Low	22,489	22,327	22,245	22,412	22,291
Group 6: Age ≥ 60, Inc. $=$ High	22,955	23,147	23,129	23,058	23,123
Online		22,039	21,278	22,448	21,654
Transaction prices, sales-weighted					
Group 1: Age < 40, Inc. $=$ Low	21,315	21,692	21,363	19,414	20,697
Group 2: Age < 40, Inc. $=$ High	22,543	25,421	23,807	18,750	21,979
Group 3: Age $\in [40, 60)$, Inc. $=$ Low	22,656	22,766	22,678	21,956	22,439
Group 4: Age $\in [40, 60)$, Inc. $=$ High	23,926	26,120	25,160	23,923	24,967
Group 5: Age ≥ 60, Inc. $=$ Low	20,985	20,919	20,816	20,810	20,821
Group 6: Age ≥ 60, Inc. $=$ High	22,927	23,309	23,233	22,738	23,109
Online		22,712	22,939	23,669	23,446
Sales, in units					
Group 1: Age < 40, Inc. $=$ Low	95,060	+13,140	+26,833	+6,422	+19,946
Group 2: Age < 40, Inc. $=$ High	69,575	+10,499	+21,640	+5,620	+16,588
Group 3: Age $\in [40, 60)$, Inc. $=$ Low	140,895	+12,540	+24,331	+6,696	+18,837
Group 4: Age $\in [40, 60)$, Inc. $=$ High	272,361	+19,252	+44,164	+10,656	+33,921
Group 5: Age ≥ 60, Inc. $=$ Low	104,249	+4,487	+8,550	+2,035	+6,481
Group 6: Age ≥ 60, Inc. $=$ High	340,390	+15,494	+34,495	+8,250	+26,086
Proportion of online sales					
Group 1: Age < 40, Inc. $=$ Low	0.690	0.824	0.829	0.761	0.812
Group 2: Age < 40, Inc. $=$ High	0.873	0.951	0.944	0.917	0.937
Group 3: Age $\in [40, 60)$, Inc. $=$ Low	0.547	0.641	0.658	0.598	0.642
Group 4: Age $\in [40, 60)$, Inc. $=$ High	0.835	0.869	0.867	0.858	0.864
Group 5: Age ≥ 60, Inc. $=$ Low	0.180	0.243	0.263	0.214	0.249
Group 6: Age ≥ 60, Inc. $=$ High	0.529	0.579	0.594	0.559	0.583
Average distance to dealers, in km					
Group 1: Age < 40, Inc. $=$ Low	14.73	+0.19	+0.41	+0.15	+0.38
Group 2: Age < 40, Inc. $=$ High	15.24	+0.15	+0.35	+0.13	+0.33
Group 3: Age $\in [40, 60)$, Inc. $=$ Low	17.80	+0.25	+0.59	+0.21	+0.53
Group 4: Age $\in [40, 60)$, Inc. $=$ High	17.22	+0.25	+0.51	+0.20	+0.48
Group 5: Age ≥ 60, Inc. $=$ Low	17.22	+0.18	+0.35	+0.13	+0.30
Group 6: Age ≥ 60, Inc. $=$ High	14.77	+0.21	+0.45	+0.16	+0.41

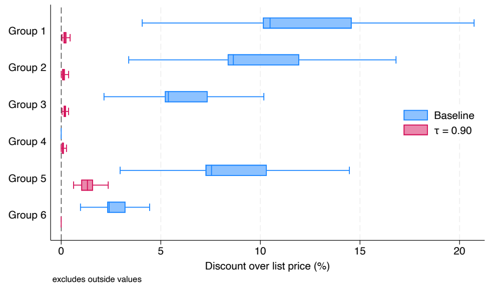
Notes: All counterfactual experiments are computed using the 2019 data only. All counterfactual have a restricted access to the online channel and a 50% reduced transportation cost online. Transaction prices are in 2018 euros. "Uniform weights" are constructed using the total sales of each product in the baseline scenario, hence are fixed across demographic groups and counterfactual experiments. "Sales-weights" use realized sales for each demographic group and counterfactual experiment. For sales and average distances, we report the values at baseline in the first column, and the change from baseline in the other columns. The cost efficiencies are computed as a percentage discount on the marginal cost, applicable to online sales only. Delivery costs are set to €400 when applicable.

Table A.6: Impact of price discrimination for alternative market structures

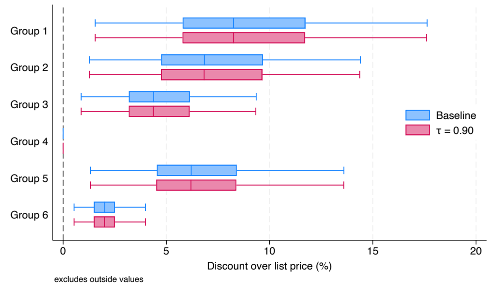
Counterfactual	Δ Consumer surplus, per capita							Δ Profits
	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6	All	Total
Oligopoly (baseline)								
• Price discrimination	411.2	589.7	1,053.5	1,757.6	1,051.7	1,823.7	1,136.1	6,229.8
• Uniform pricing	-65.8	-64.3	-36.5	+168.2	-102.6	+64.6	+4.2	-53.6
Monopoly								
• Price discrimination	386.5	543.5	925.9	1,470.1	918.8	1,512.0	974.1	6,320.4
• Uniform pricing	-101.4	-106.0	-63.7	+205.0	-120.6	+117.5	+2.8	-61.5
Competition								
• Price discrimination	416.1	598.6	1,082.3	1,818.6	1,099.3	1,916.4	1,178.3	6,181.1
• Uniform pricing	-56.6	-53.5	-27.2	+165.8	-103.8	+44.3	+4.2	-49.5

Notes: All counterfactual experiments are computed using the 2019 data only. Oligopoly is the observed ownership structure in the data. Monopoly implies that one firm sells all product, or equivalently the full collusive equilibria. Competition implies J firms, each selling one product. Consumer surplus is in 2018 euros per capita. Profits are in million 2018 euros.

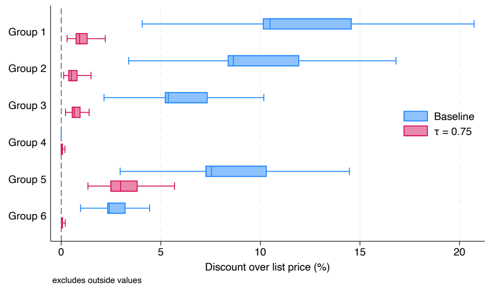
Figure A.1: Price dispersion, single firm online



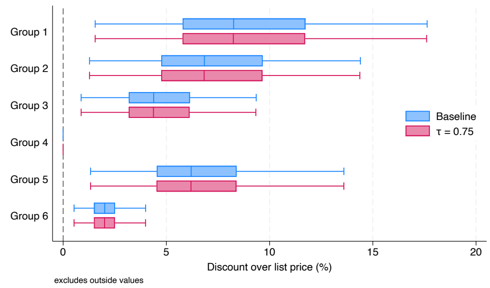
(a) Nissan-Renault group ($\tau = 0.90$)



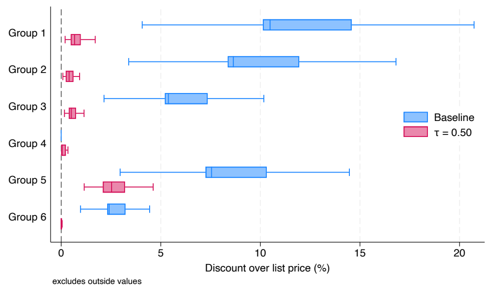
(b) Other manufacturers ($\tau = 0.90$)



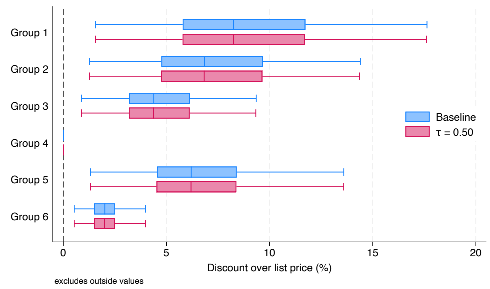
(c) Nissan-Renault group ($\tau = 0.75$)



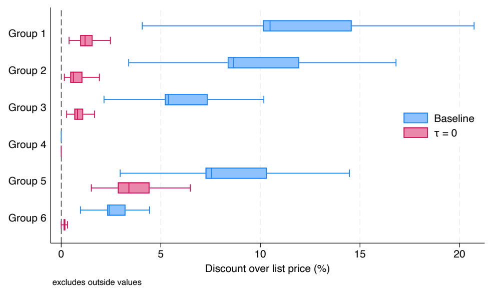
(d) Other manufacturers ($\tau = 0.75$)



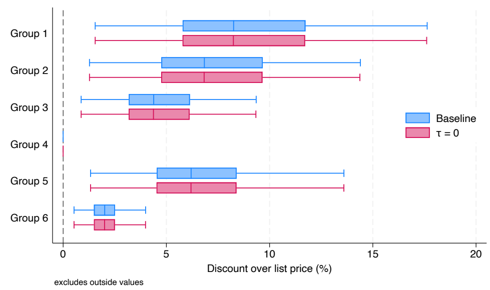
(e) Nissan-Renault group ($\tau = 0.50$)



(f) Other manufacturers ($\tau = 0.50$)



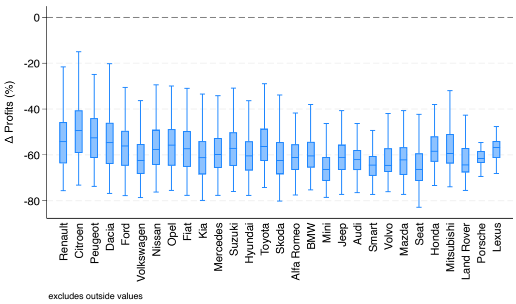
(g) Nissan-Renault group ($\tau = 0$)



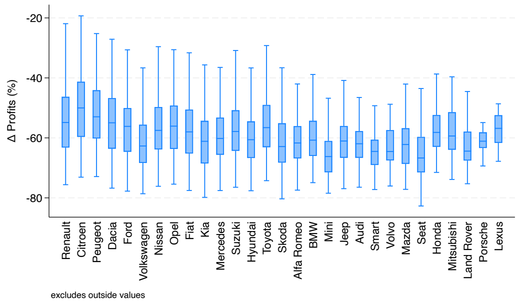
(h) Other manufacturers ($\tau = 0$)

Notes: These figures represent the distribution of in-person discounts for a counterfactual where only the Nissan-Renault group offers online sales as per Table 11. All other manufacturers are restricted to selling in person only.

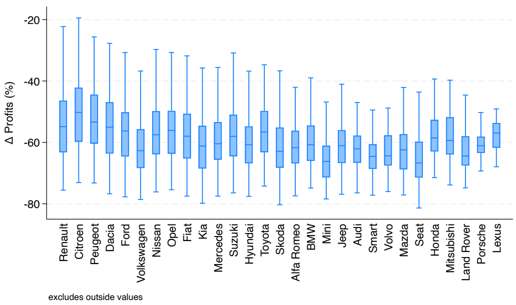
Figure A.2: Change in dealer-level profits, in-person stores only



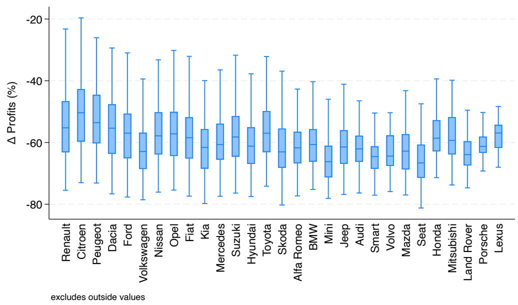
(a) Transportation costs = 90%



(b) Transportation costs = 75%



(c) Transportation costs = 50%



(d) No transportation cost

Notes: These figures represent the change profits at the dealer level from introducing the online channel as per Table 10. Brands are ranked by the total number of dealers, in decreasing order. Profits are computing from sales occurring in person only.

B Computational details

In this section, we provide additional computational details related to the data, the estimation routine, and the computation of counterfactuals.

B.1 Additional details on the data

Construction of demographic groups. We provide more details on the construction of the demographic groups. We collect data from two sources, a survey of populations by municipalities and age groups, available every 5 years, and a survey of incomes by municipalities and age groups available on a yearly basis. Both datasets are available from the Institut National de la Statistique et des Études Économiques (INSEE).[17] We select the following categorizations for age: young (39 and below), middle aged (between 40 and 59 inclusively), and old (60 and above). Within age category, we split municipalities into two evenly-sized groups, high and low income, according to the median income reported in the income files. Since income is reported in increments smaller than our age categories, we use a population-weighted average of the median incomes within age classes and municipalities to assign an income group. In some cases where populations are very small, income is not reported separately by age, so we use the municipalities median income in that case. We drop a small number of municipalities which are too small to report income at all (along with the associated car sales). We obtain 6 demographic groups, described in Table B.1 below.

Table B.1: Demographic groups definition

Group	Definition	
Group 1	Age 39 and below	Income in bottom half of age-specific distribution
Group 2	Age 39 and below	Income in top half of age-specific income distribution
Group 3	Age between 40 and 59	Income in bottom half of age-specific distribution
Group 4	Age between 40 and 59	Income in top half of age-specific distribution
Group 5	Age 60 and above	Income in bottom half of age-specific distribution
Group 6	Age 60 and above	Income in top half of age-specific distribution

The population files include both the population by municipality and age and the number of household. We use the number of household to define market sizes, and we compute the average household size using the ratio of populations to the number of households. The average age can be approximated with the population data, taking the midpoint of age intervals (5-year increments) and using population-weights. These files also offer a breakdown of populations by gender which allows us to compute the share of female

[17] Source: https://www.insee.fr/

by group. We merge these data with a survey of population densities, also available at INSEE. Population density is available as an municipal level categorical variable indicating whether a given municipality is urban, suburban, or rural. Finally, we rely on a survey of attitudes towards online shopping to determine the propensity to shop online, by demographic group (see Section 5.2 for details). A summary of the demographic characteristics of buyers by demographic group is available in Table A.1.

Construction of the car data. Our car data comes from AAA data, which collects data on all car registrations in France. We obtain all new car registrations between 2009 and 2021. The data is aggregated by car model (i.e., products), age groups (in 5-year increments), and municipalities. We merge the data with consumers demographics to recover demographic groups based on the age and municipality of residence of buyers. We form two main datasets. The first one is aggregated at the level of the brand-model-engine-year-demographic group—this is our aggregated dataset used in the estimation of demand. The second one is aggregated at the level of the brand-model-engine-year-municipality-demographic group—this is our disaggregated dataset used to compute micro moments.

We keep the 29 most prominent brands, and keep products with a net price (adjusted for the French feebate program) below 100,000 euros. The car data includes the list prices, and some common car characteristics such as the horsepower. Horsepower and fuel consumption are not available for electric vehicles in the data. We impute the missing horsepower using an alternative data source, and we set the fuel consumption of electric vehicle to their fuel equivalent electricity consumption. We compute fuel costs using various fuel prices interacted with fuel consumption, depending on the engine type (e.g., diesel prices for diesel engines). Finally, we obtain each vehicle's marketing segment (i.e., compact, SUV) and the country of origin of each model (e.g., the location of the plant that produces each model) from Jato Dynamics. A breakdown of sales by model and demographic group and a summary of the main car characteristics is presented in Table 2 for completeness.

B.2 Additional details on the estimation

Notation. We provide the list of various notations used throughout the paper, see Table B.2 below.

Table B.2: Model notation

Notation	Description
i	Individuals
j	Products
d	Demographic groups
m	Municipalities
t	Markets (years)
f	Firms
P	In-person channel
O	Online channel
M_{dmt}	Number/set of individuals in municipality m, demographic group d, and market t
M_{dt}	Number/set of individuals in demographic group d and market t
M_t	Number/set of individuals in market t
$\mathcal{M}$	Set of all municipalities
J_t	Number/set of products available in market t
$\mathcal{J}_{ft}$	Set of products offered by firm f in market t
D	Number/set of demographic groups
T	Number/set of markets (years)
α_d	Price sensitivity of group d
γ_d	Distance sensitivity of group d
β_d	Preference parameters of group d
λ_1, λ_2	Cost function parameters
θ	Set of all parameters $(\alpha_1, ..., \alpha_D, \gamma_1, ..., \gamma_D, \beta_1, ..., \beta_D, \lambda_1, \lambda_2)$
ψ_{dt}	Share of consumers from demographic group d in market t that have access to the online channel
τ	Transportation cost reduction for online sales
σ	Nesting parameters, online vs in-person channel
p^P_{jdt}	In-person price of product j for group d
p^O_{jt}	Online (uniform) price of product j
$\overline{p}_{jt}$	List price of product j
c_{jt}	Marginal cost of product j
s^P_{jdt}	In-person market share of product j for group d
s^O_{jdt}	Online market share of product j for group d
ϕ_{dt}	Share of consumers in market t that belong to demographic group d (M_{dt}/M_t)
w_{dmt}	Share of consumers in demographic group d and market t that live in municipality m (M_{dmt}/M_{dt})

Specification. We provide additional details on the chosen specification. Table B.3 presents a description of the variables that are included in the demand and supply estimation. Micro moments are constructed using distances as per Section 2.3. Unless indicated otherwise, car characteristics are used both in the demand-side and the supply-side estimations. Cost shifters are used in the supply side only. We also report on the demand- and supply-side instruments used in the estimation, which are constructed

Table B.3: Specification details

Variable	Description
Car characteristics	
Price	Price, net of French feebate program, in 10,000 2018 euros (Demand only)
Distance	Driving distance, in 10 km (Demand only)
	Note: Driving distance is measured from the centroid of the municipality of residence of consumers to the centroid of the zipcode of car dealers.
Horsepower	Horsepower, in 100 kW
Weight	Curb weight, in 1,000 kg
Fuel cost	Cost for driving 100km, in 2018 euros (Demand only)
Fuel consumption	Fuel consumption, in L / 100 km (Supply only)
Diesel	=1 if Diesel
Electric	=1 if Electric
Plug-in hybrid	=1 if Plug-in hybiid
Hybrid	=1 if Hybrid
Station wagon	=1 if Station Wagon
Convertible	=1 if Convertible
Trend	Time trend (Supply only)
Cost shifters	
Input price index	Composite price index based on steel price (56%), polypropylene price (8%), iron price (8%), and aluminum price (10%), interacted with vehicle weight (Supply only)
Real exchange rate	Penn World Table 10.0, `pl_con`, see Grieco et al. (2023) (Supply only)
	Note: Both cost shifters are lagged one period to reflect planning horizons.
Instruments	
Demand-side	(1) Sum of characteristics of competitors using horsepower, weight, fuel cost
	(2) Number of competitors' products
	(3) Number of competitors' products with same engine type
	(4) Number of competitors' products with same body trim
	Note: Demand-side instruments are the same for all demographic groups.
Supply-side	(1) Sum of characteristics of competitors using horsepower, weight, fuel consumption, input price index, real exchange rate
	(2) Number of competitors' products
	(3) Number of competitors' products with same engine type
	(4) Number of competitors' products with same body trim

following Berry et al. (1995).

Estimation. The estimation is done on our data which includes only in-person sales. We add the online channel in counterfactual experiments only. We exclude Tesla from the estimation as we do not observe their locations and cannot compute the distance to dealers in that case. They represent a very small share of total sales. The model is estimated by generalized method of moments, following Berry et al. (1995) and subsequent best practices. We stack the moment conditions $g(\theta) = (g_1(\theta)', g_2(\theta)', g_3(\gamma)')'$ and construct the following estimator

$$\widehat{\theta} = \arg\min_{\theta} \quad g(\theta)'\mathbf{W}g(\theta) \tag{21}$$

where $\mathbf{W} = \text{diag}(\mathbf{W}_{11}, ..., \mathbf{W}_{1D}, \mathbf{W}_2, \mathbf{W}_3)$ is a block-diagonal symmetric weighting matrix. The linear parameters $(\beta_1', ..., \beta_D', \lambda_1', \lambda_2')$ are concentrated out and the estimation is focused around identifying price sensitivities $(\alpha_1, ..., \alpha_D)$ and the distance sensitivities $(\gamma_1, ..., \gamma_D)$. We use a tight convergence criterion of 1e–12 for the inner loop and a sample of 3,000 municipalities to compute the integral in the market share equation. We draw municipalities by systematic sampling separately for each demographic group and each market. Finally, we solve for the discriminatory prices at each iteration of the estimation.

We use the same set of instruments for all demographic groups in the demand side, denoted by Z_1, and a different set of instruments for the supply side, denoted Z_2. Details about these instruments are available in Table B.3 above. The weighting matrices are set to $W_{1d} = (Z_1' \cdot Z_1)^{-1}$ for all $d = 1, ..., D$ and $W_2 = (Z_2' \cdot Z_2)^{-1}$.

Micro moments are computed at the demographic group by market level, meaning that we have 78 additional moments (6 groups $\times$ 13 markets) to identify 6 distance coefficients. The weighting matrix for the micro moments is set to the identity matrix, $W_3 = I$, scaled by a factor of 1e–4.

Market size and market shares. The various market sizes (M_{dmt}, M_{dt}, and M_t) are computed as the number of households (by demographic group, municipality, and market, as appropriate) divided by four. In the data, we observe a few products with a market share of zero in some demographic groups and markets. Assuming that these products are unavailable at the national level only to certain demographic groups did not seem appropriate in this case. To circumvent this issue, we follow D'Haultfœuille et al. (2019) and compute observed market shares as

$$s_{jdt} = \frac{q_{jdt} + 0.5}{M_{dt}},$$

where q_{jdt} is the total quantity sold of product j for demographic group d and market t and M_{dt} is the market size. We perform some robustness checks using a standard IV logit model to verify that this doesn't affect the estimation—removing the products with zero shares and estimating the model computing market shares as usual. We find that the estimated coefficients are statistically unaffected by the change.

B.3 Solving for prices

Market shares. We precise how we compute counterfactuals prices. We want to approach our utility specification with a nested logit structure. Recall that the utility of consumer i from purchasing car model j from channel $\ell \in \{P, O\}$ is given by

$$U^{\ell}_{ijdm} = \delta_{jd} + \mu^{\ell}_{jdm} + \zeta_{ijdm} + (1-\sigma)\epsilon^{\ell}_{ijdm},$$

where $\zeta_{ijdm} + (1-\sigma)\epsilon^{\ell}_{ijdm}$ distributed as extreme value type I.

Consider the case where consumer i belongs to demographic group d and lives in municipality m. The subscript t is omitted. The market shares of product j for each distribution channel are given by

$$s^{O}_{jdm} = s_{O|jdm} \cdot s_{jdm},$$

$$s^{P}_{jdm} = s_{P|jdm} \cdot s_{jdm},$$

where

$$s_{O|jdm} = \frac{\exp\left(\frac{\delta_{jd}+\mu^{O}_{jdm}}{1-\sigma}\right)}{\sum_{\ell\in\{P,O\}} \exp\left(\frac{\delta_{jd}+\mu^{\ell}_{jdm}}{1-\sigma}\right)},$$

$$s_{P|jdm} = \frac{\exp\left(\frac{\delta_{jd}+\mu^{P}_{jdm}}{1-\sigma}\right)}{\sum_{\ell\in\{P,O\}} \exp\left(\frac{\delta_{jd}+\mu^{\ell}_{jdm}}{1-\sigma}\right)},$$

$$s_{jdm} = \frac{\exp\left(IV_{jdm}\right)}{1+\sum_k \exp\left(IV_{kdm}\right)},$$

and

$$IV_{jdm} = (1-\sigma)\ln\left[\sum_{\ell\in\{P,O\}} \exp\left(\frac{\delta_{jd}+\mu^{\ell}_{jdm}}{1-\sigma}\right)\right].$$

Note that IV_{jjdm} is the nested logit "inclusive value". Aggregating over individual market shares, we get

$$s^{\ell}_{jd} = \sum_{m\in\mathcal{M}} w_{dm} \cdot s^{\ell}_{jdm}(\text{dist}_{1m}, ..., \text{dist}_{Jm}),$$

for $\ell = \{P, O\}$, where $\mathcal{M}$ is the set of all municipalities, and w_{dm} are municipality group-specific population weights.

Derivatives. We now show explicitly what are the various derivatives that enters the firms' first-order conditions. Notice that prices at the regular and online dealerships affect the market shares at both types of retailers.

$$\frac{\partial s^P_{jd}}{\partial p^P_{jd}} = \sum_{m\in\mathcal{M}} \frac{\alpha_d}{1-\sigma} \cdot s^P_{jdm}\left[1 - \sigma s_{P|jdm} - (1-\sigma)s^P_{jdm}\right] \cdot w_{dm},$$

$$\frac{\partial s^P_{jd}}{\partial p^P_{kd}} = \sum_{m\in\mathcal{M}} -\frac{\alpha_d}{1-\sigma} \cdot s^P_{jdm}s^P_{kdm} \cdot w_{dm},$$

$$\frac{\partial s^P_{jd}}{\partial p^O_{j}} = \sum_{m\in M} -\frac{\alpha_d}{1-\sigma} \cdot s^P_{jdm}\left[\sigma s_{O|jdm} + (1-\sigma)s^O_{jdm}\right] \cdot w_{dm},$$

$$\frac{\partial s^P_{jd}}{\partial p^O_{k}} = \sum_{m\in\mathcal{M}} -\frac{\alpha_d}{1-\sigma} \cdot s^P_{jdm}s^O_{kdm} \cdot w_{dm},$$

$$\frac{\partial s^O_{jd}}{\partial p^P_{jd}} = \sum_{m\in\mathcal{M}} -\frac{\alpha_d}{1-\sigma} \cdot s^O_{jdm}\left[\sigma s_{P|jdm} + (1-\sigma)s^P_{jdm}\right] \cdot w_{dm},$$

$$\frac{\partial s^O_{jd}}{\partial p^P_{kd}} = \sum_{m\in\mathcal{M}} -\frac{\alpha_d}{1-\sigma} \cdot s^O_{jdm}s^P_{kdm} \cdot w_{dm},$$

$$\frac{\partial s^O_{jd}}{\partial p^O_{j}} = \sum_{m\in\mathcal{M}} \frac{\alpha_d}{1-\sigma} \cdot s^O_{jdm}\left[1 - \sigma s_{O|jdm} - (1-\sigma)s^O_{jdm}\right] \cdot w_{dm},$$

$$\frac{\partial s^O_{jd}}{\partial p^O_{k}} = \sum_{m\in\mathcal{M}} -\frac{\alpha_d}{1-\sigma} \cdot s^O_{jdm}s^O_{kdm} \cdot w_{dm}.$$

Solving Counterfactuals. We adapt the methodology in Morrow and Skerlos (2011) to our setup. We begin by stacking the price vectors, the marginal cost vectors, and the market share vectors to solve for the equilibrium in one step, that is

$$\mathbf{p} = \begin{bmatrix} p^P_1 \\ \cdots \\ p^P_D \\ p^O \end{bmatrix}, \qquad \mathbf{c} = \begin{bmatrix} c \\ \cdots \\ c \\ c \end{bmatrix}, \qquad \mathbf{s} = \begin{bmatrix} \phi_1 s^P_1 \\ \cdots \\ \phi_D s^P_D \\ \sum_d \phi_d s^O_d \end{bmatrix}. \tag{22}$$

Notice that the market share of each group is multiplied by the share of consumers in that demographic group, such that multiplying the full vector of market shares by total population yields total sales by demographic group. It is also important to note that equilibrium prices cannot be solved separately for the various demographic groups as is the case when online dealers are not present, since the uniform online price affect the personalized prices and vice versa through the derivatives above.

Let $\mathcal{H}$ be the ownership matrix and $\mathcal{D}_d^{\ell,\kappa}$ be the $J \times J$ matrix with element (j,k) equal to $\partial s_{kd}^{\ell} / \partial p_{jd}^{\kappa}$ for $(\ell,\kappa) \in \{P,O\}^2$. As previously, let us also define $\widetilde{\mathcal{D}}_d^{\ell,\kappa} = \mathcal{H} \odot \mathcal{D}_d^{\ell,\kappa}$. Then, we can compute the matrix of demand derivatives for the stacked vectors defined in equation (22) as

$$\widetilde{\mathcal{D}}(\mathbf{p}) = \begin{bmatrix} \phi_1 \widetilde{\mathcal{D}}_1^{P,P} & 0 & 0 & 0 & 0 & 0 & \phi_1 \widetilde{\mathcal{D}}_1^{O,P} \\ 0 & \phi_2 \widetilde{\mathcal{D}}_2^{P,P} & 0 & 0 & 0 & 0 & \phi_2 \widetilde{\mathcal{D}}_2^{O,P} \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots \\ 0 & 0 & 0 & \dots & 0 & 0 & \dots \\ 0 & 0 & 0 & 0 & \dots & 0 & \dots \\ 0 & 0 & 0 & 0 & 0 & \phi_D \widetilde{\mathcal{D}}_D^{P,P} & \phi_D \widetilde{\mathcal{D}}_D^{O,P} \\ \phi_1 \widetilde{\mathcal{D}}_1^{P,O} & \phi_2 \widetilde{\mathcal{D}}_2^{P,O} & \dots & \dots & \dots & \phi_D \widetilde{\mathcal{D}}_D^{P,O} & \sum_d \phi_d \widetilde{\mathcal{D}}_d^{O,O} \end{bmatrix}. \tag{23}$$

Solving for counterfactuals is then a straightforward fixed-point iteration, based on Morrow and Skerlos (2011), on the stacked system of first-order conditions, that is,

$$\mathbf{p} = \mathbf{c} + \Lambda(\mathbf{p})^{-1} \cdot \widetilde{\Gamma}(\mathbf{p}) \cdot (\mathbf{p} - \mathbf{c}) - \Lambda(\mathbf{p})^{-1} \cdot \mathbf{s}$$

where $\widetilde{\mathcal{D}}(\mathbf{p}) = \Lambda(\mathbf{p}) - \widetilde{\Gamma}(\mathbf{p})$ is the same decomposition as in Morrow and Skerlos (2011).

Extrapolation. A computational issue that arises when approximating market shares and their derivatives during the price optimization routine is that $s_{jd}^{\ell} \to \frac{\infty}{\infty}$ for $\sigma \approx 1$ using conventional software packages (i.e., both the numerator and the denominator "blow up" past the threshold for infinity which is around 1e700). This prevents the evaluation of counterfactual price vectors in the limit as $\sigma \to 1$.

To circumvent this issue, we rely on a linear extrapolation. We evaluate counterfactual prices for two very high values of σ, say $\sigma_1 = 0.95$ and $\sigma_2 = 0.96$, then approximate counterfactual prices as

$$p_{jd}^{\ell}(\sigma \approx 1) = \lim_{\sigma \to 1} p_{jd}^{\ell}(\sigma) \approx \frac{1 - \sigma_2}{\sigma_2 - \sigma_1} \cdot \left(p_{jd}^{\ell}(\sigma_2) - p_{jd}^{\ell}(\sigma_1) \right). \tag{24}$$

We do the same for market shares. We compute

$$s_{jd}^{\ell}(\sigma \approx 1) = \lim_{\sigma \to 1} s_{jd}^{\ell}(\sigma) \approx \frac{1-\sigma_2}{\sigma_2 - \sigma_1} \cdot \left(s_{jd}^{\ell}(\sigma_2, p(\sigma_2)) - s_{jd}^{\ell}(\sigma_1, p(\sigma_1)) \right), \tag{25}$$

where $p(\sigma) = \{p_1^P, ..., p_D^P, p^O\}$ includes both the group-specific prices and the uniform online price of all products. We perform several robustness checks to verify in particular that the market shares obtained from the extrapolation match the market shares computed using the max{ , } formulation in (16) and (17). We find the the approximation yields market shares and aggregate sales that are very close to each other (see Table B.4).

Table B.4: Robustness check on market share calculations

	Total sales						
	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6	All
In person							
• Nested logit	29,599	8,910	64,004	44,923	85,762	160,573	393,771
• Maximum	33,723	9,021	66,751	46,251	85,807	156,928	398,481
• Difference	-4,124	-111	-2,747	-1,327	-45	3,645	-4,709
Online							
• Nested logit	65,835	61,039	77,266	227,811	18,861	180,191	631,003
• Maximum	62,069	61,271	74,796	227,113	18,877	184,302	628,428
• Difference	3,766	-233	2,470	699	-16	-4,110	2,576
Both channels							
• Nested logit	95,434	69,949	141,269	272,735	104,623	340,764	1,024,774
• Maximum	95,792	70,292	141,547	273,364	104,684	341,229	1,026,908
• Difference	-358	-343	-277	-629	-61	-465	-2,133

Notes: To do. This is with restricted buyers and a transportation cost of 50%.

C Price discrimination in the online channel

For completeness, we consider a counterfactual experiment in which firms can price discriminate both online and offline. While this does not align with firms intentions about online sales, we acknowledge that they could in principle price discriminate online, as data on consumers is readily available online. For example, consumers demographics could be inferred from browsing histories by a third-party data broker and resold to car manufacturers. We focus on the pricing behavior of firms in this case. The results are presented in Table B.5.

Table B.5: Online and offline price discrimination

	Baseline	Restricted online access			
		$\tau = 0.90$	$\tau = 0.75$	$\tau = 0.50$	$\tau = 0$
Transaction prices – In-person channel					
Group 1: Age < 40, Inc. $=$ Low	21,825	21,826	21,827	21,830	21,839
Group 2: Age < 40, Inc. $=$ High	22,119	22,120	22,121	22,125	22,134
Group 3: Age $\in [40, 60)$, Inc. $=$ Low	22,611	22,612	22,614	22,619	22,633
Group 4: Age $\in [40, 60)$, Inc. $=$ High	23,583	23,585	23,590	23,600	23,623
Group 5: Age ≥ 60, Inc. $=$ Low	22,222	22,223	22,224	22,228	22,243
Group 6: Age ≥ 60, Inc. $=$ High	23,118	23,120	23,123	23,130	23,150
Transaction prices – Online channel					
Group 1: Age < 40, Inc. $=$ Low		21,825	21,827	21,830	21,840
Group 2: Age < 40, Inc. $=$ High		22,120	22,121	22,125	22,134
Group 3: Age $\in [40, 60)$, Inc. $=$ Low		22,611	22,614	22,620	22,637
Group 4: Age $\in [40, 60)$, Inc. $=$ High		23,585	23,590	23,601	23,628
Group 5: Age ≥ 60, Inc. $=$ Low		22,221	22,225	22,235	22,269
Group 6: Age ≥ 60, Inc. $=$ High		23,119	23,123	23,133	23,157

Notes: This table represents the results from a counterfactual experiment in which all manufacturers can price discriminate online and offline. Column (1) represents the baseline scenario without online sales. Column (2) to (4) represent counterfactuals where some consumers are restricted to shop in-person, for varying levels of transportation cost reductions. The transaction prices are weighted by a uniform set of weights constructed from the total sales of each product under baseline.

We notice that firms set almost the same price in both sales channels when it is possible for them to price discriminate online. Prices are increasing slightly as we reduce transportation costs, a sign that firms increase prices slightly to capture part of the reduction in transportation costs as profits. Also, online prices are slightly higher compared to offline prices when transportation costs are removed completely. In that case, firms can price both captive and non-captive consumers separately as before, but the price difference is only a few dollars on average. One last observation is that the extent of price discrimination remains the same when firms are able to price discriminate in both channels; price dispersion is of the same magnitude in the counterfactuals as under baseline in this case, even though the average prices are slightly higher.