Automobile Prices in Market Equilibrium with Unobserved Price Discrimination*

Xavier D'Haultfœuille† Isis Durrmeyer‡ Philippe Février§

September 13, 2018

Abstract

In markets where sellers are able to price discriminate, individuals pay different prices that may be unobserved by the econometrician. This paper considers the structural estimation of a demand and supply model à la Berry et al. (1995, henceforth BLP) with such price discrimination and limited information on prices taking the form of, e.g., observing list prices from catalogues or average prices. Within this framework, identification is achieved not only with usual moment conditions on the demand side, but also through supply-side restrictions. The model can be estimated by GMM using a nested fixed point algorithm that extends BLP's algorithm to our setting. We apply our methodology to estimate the demand and supply in the French new automobile market. Our results suggest that discounting arising from price discrimination is important. The average discount is estimated to be 9.6%, with large variation depending on buyers' characteristics and cars' specifications. Our results are consistent with other evidence on transaction prices in France.

Keywords: demand and supply, unobserved prices, price discrimination, automobiles.

JEL codes: C51, D12, D22.

*We would like to thank the editor Stéphane Bonhomme, five anonymous referees, Steven Berry, Juan Esteban Carranza, Rob Clark, Pierre Dubois, Liran Einav, Philippe Gagnepain, Gautam Gowrisankaran, Alessandro Iaria, John Morrow, Kathleen Nosal, Mario Pagliero, Ariel Pakes, John Rust, Philipp Schmidt-Dengler, Michelle Sovinsky, Sri Srikandan, Hidenori Takahashi, Yuya Takahashi, Frank Verboven, Naoki Wakamori and the participants of various seminars and conferences for their valuable comments. Finally, we thank Pierre-Louis Debar and Julien Mollet from the CCFA for sharing their data with us.

†CREST. E-mail: xavier.dhaultfoeuille@ensae.fr

‡Toulouse School of Economics, Université Toulouse Capitole. E-mail: isis.durrmeyer@tse-fr.eu

§CREST. E-mail: philippe.fevrier@ensae.fr

1 Introduction

The standard aggregate-level estimation of demand and supply models of differentiated products relies on the observation of market shares and products characteristics, in particular prices (see Berry, 1994). Yet, prices for an identical product can differ significantly across transactions. Temporary promotions, coupons, local prices and negotiation between buyers and sellers are examples of factors that lead to such price dispersion. Precise data on transaction prices may be hard to obtain. One typically observes posted prices from catalogues, or average prices over the different transactions. In the two cases, the common practice is simply to ignore price variation across transactions, and consider the observed prices to be the prices paid by all consumers.

One may argue that the instrumental variables approach developed by Berry et al. (1995, henceforth BLP) to control for price endogeneity also solves this measurement error problem on prices. However, to the extent that prices are not randomly assigned to consumers, the difference between observed prices and transaction prices are generally correlated with the instruments. Hence, ignoring this issue generally results in inconsistent estimators of the structural parameters and biases in policy simulation exercises.

This paper proposes a method to estimate structural demand and supply models with unobserved transaction prices. Our rationale for the existence of price dispersion is that firms price discriminate between heterogeneous consumers in order to extract more surplus than they would with a uniform pricing strategy, as long as other firms also price discriminate. We suppose that sellers use observable characteristics of the buyers to price discriminate and set optimal prices. Of course, it is unlikely that sellers observe perfectly the consumers' preferences, and we allow for individual heterogeneity in consumers' preferences that is unobserved by both the sellers and the econometrician. This model applies to several of the cases mentioned above: temporary discounts if the population of consumers differs along the time of purchase, local prices if individuals in different geographical areas have different willingness to pay (e.g., poor versus rich neighborhoods), coupons when they are used by the most price sensitive consumers. Price negotiation may also be seen as a special case of price discrimination, provided that there is no informational friction and sellers make take-it or leave-it offers to consumers. This set-up seems in particular appropriate to model business to business markets with small buyers and large sellers.

We therefore extend the random coefficient discrete choice model of demand popularized by BLP to allow for unobserved price discrimination. To identify this demand model, we rely on usual moment conditions on the demand side, but also on supply-side restrictions. The latter are key here, insofar as prices are only partially observed. First, we consider third-degree price discrimination, as opposed to (unobserved) second-degree price discrimination. Third-degree price discrimination is realistic whenever sellers can price discriminate consumers on the exact same product. Second-degree price discrimination arises, on the other hand, when sellers let consumers self-select through menus of contracts corresponding to different quality or quantity levels and prices. In this latter case, the transaction price paid by a given consumer depends on his specific unobserved preferences, something our approach cannot handle.

In this setup we rely on three identifying assumptions. First, we assume that the econometrician observes the market shares of products for each group of consumers used for price discrimination. This requires in particular to observe the characteristics of consumers used by sellers to price discriminate. Second, the marginal cost of a product is supposed to be identical for all the buyers. This amounts to neglecting differences in the cost of selling to different consumers in the total cost of a product. This assumption is credible when the major part of the marginal cost is production, not sale, or when the cost of selling has no reason to differ from one consumer to another. The third condition for identification is that there is a known relationship between observed and transaction prices. This assumption is satisfied if the econometrician observes average prices, or at least one transaction price for each product. It also holds if list prices are observed, as these prices correspond to the maximal transaction prices whenever some consumers do pay these prices.

Under these three assumptions, we show that we can estimate the model with the general method of moments (henceforth, GMM) in a similar spirit as BLP, but with a modified algorithm to account for unobserved prices. In addition to obtaining the average utilities through the numerical inversion of the market share equations, we recover transaction prices by using the first-order conditions of profit maximization. We consider a fixed-point algorithm to compute both average utilities and transaction prices. We show that this algorithm converges under, basically, some conditions on the amount of heterogeneity among consumers. We establish for that purpose that the corresponding function is a contraction, similarly to BLP's celebrated contraction mapping.

To assess the credibility of our model, we develop a formal test of price discrimination versus uniform pricing, where all consumers pay the same prices. Finally, we consider a few extensions to our initial set-up. First, our method extends, under some conditions, to the case where only aggregate market shares are observed. Second, we show that the assumption of identical marginal costs can be relaxed, as long as we observe the underlying cost shifters, and that they vary across products. We discuss additional extensions in our supplement.

We apply our method to estimate the primitives of the French new car market. Up to now, the demand for automobiles has always been estimated with posted or average prices, whenever transaction prices were unobserved. There is however much evidence of price dispersion and price discrimination in this market (see below for references). We estimate our model using an exhaustive dataset recording all the registrations of new cars bought by households in France between 2003 and 2008. Apart from detailed car attributes, including list prices, we observe the age of the buyers and their expected income given their age class and municipality of residence. As these characteristics are easily observed by sellers and presumably strong determinants of car purchases, we suppose that they are used by sellers to price discriminate.

Our results suggest that price discrimination is significant in France. First, our statistical test clearly favors price discrimination over the uniform pricing model. Second, the average discount is estimated to be 9.6% of the posted price. The distribution of estimated discounts spreads mostly between 0 and 25% depending on the car model purchased and demographic characteristics. As expected, age and income are negatively correlated to the value of discount. We also show that price discrimination may be important in ex ante policy analysis. Finally, we compare our estimated transaction prices with a sample of observed transaction prices from the French consumer expenditure survey. These data suggest that our method provides reasonable estimates of transaction prices.

Related literature. Our paper is related to three strands of the literature. First, on a methodological side, it is close to empirical papers estimating demand and supply with imperfect information on prices. Miller and Osborne (2014) develops a method to estimate the demand for cement when only average prices and total quantities are observed, allowing for price discrimination (see also Thomadsen, 2005, for a related approach with unobserved quantities). While they compute optimal prices using equilibrium conditions, as we do here, their model and estimation strategy are very different from ours. First, they do not account for differences in preferences across groups, which is usually the rationale for price discrimination. Second, they rule out any systematic unobserved preference term, implying in particular that prices are exogeneous. These restrictions may hold in homogeneous good markets such as cement but are less realistic for differentiated product markets, in which preference heterogeneity and unobserved components such as quality play a crucial role.

Our paper is also related to Dubois and Lasio (2018), which estimates marginal costs when observed prices are regulated, so that marginal costs can no longer be identified from these prices. They use the first-order conditions of the firms in countries that do not regulate prices to identify these costs, under a similar assumption as us on the marginal costs. Contrary to us, however, they do not use the first-order conditions of the firms to identify the demand model, as they observe the prices paid by consumers.

Our paper also builds on the recent literature that considers hybrid models of bargaining in which sellers post a sticker price and offer the possibility to bargain for discounts. This pricing strategy might be profitable for sellers when consumers have heterogeneous bargaining costs or are imperfectly informed on their ability to bargain (see Gill and Thanassoulis, 2009, 2016). Jindal and Newberry (2016) develop a structural model of demand in which all the buyers are able to negotiate but have heterogeneous bargaining cost. The authors estimate both the bargaining power and the distribution of bargaining costs. However, their framework is very different from ours since they omit competition and they observe the transaction prices at the individual level. Structural models in which prices are set by a bargaining process have also been recently developed and estimated, but for the case of business to business markets where there are few identifiable actors (see, e.g. Crawford and Yurukoglu, 2012; Grennan, 2013; Gowrisankaran et al., 2015). In such models, bargaining modifies the supply side but contrary to our case, the demand can be estimated in a standard way.

Finally, our application quantifies the importance of price discrimination on the automobile market. Several papers investigate this issue using either detailed data on transaction prices or dealers margins, see, e.g., Ayres and Siegelman (1995), Goldberg (1996), Harless and Hoffer (2002), Scott Morton et al. (2003), Langer (2016) and Chandra et al. (2017). Noteworthy, our method is still useful to estimate demand with data on prices paid by each consumer, because the prices of products that are not chosen by the consumers remain unobserved. Huang (2016) is closest to our paper. He develops a structural model of demand for cars when some dealers make second-degree price discrimination by proposing the list price immediately or a discounted price later. He estimates both demand parameters and the discounts using market shares and list prices. Rather than relying on the supply-side conditions, as here, identification is achieved by leveraging the existence of non-negotiating car dealers.

Structure of the paper. The second section presents the theoretical model and identifying assumptions. Section 3 develops our estimation method and consider some extension. The application on the French new car market is developed in Section 4, while Section 5 concludes. We present our estimation algorithm and all the proofs in the appendix. Finally, we present additional extensions, Monte Carlo simulations and further details on the application in the supplement (D'Haultfœuille et al., 2018).

2 Theoretical model and identifying assumptions

2.1 The unobserved price discrimination model

We first present our theoretical model. The approach is identical to the BLP model, except that the demand arises from a finite number of heterogeneous groups of consumers. Firms are supposed to observe the group of each consumer and the corresponding distribution of preferences (price sensitivities and preferences for car attributes). They price discriminate among these groups, in order to take advantage of the difference in preferences from one group to another.

Specifically, consumers are supposed to be segmented into n_D groups of consumers, and we denote by D_i the group of consumer i. As in the standard BLP model, we allow consumers to be heterogeneous within a group, but assume sellers are not able to discriminate based on this unobserved heterogeneity. Each consumer chooses either to purchase one of the J products or not to buy any, which corresponds to the outside option denoted by 0. As usual, each product is assimilated to the bundle of its characteristics. Consumers maximize their utility, and the utility of choosing j is assumed to be a linear function of product characteristics:

$$U_{ij}^d = X_j'\beta_i^d + \alpha_i^d p_j^d + \xi_j^d + \varepsilon_{ij}^d,$$

where X_j corresponds to the vector of observed characteristics and ξ_j^d represents the valuation of unobserved characteristics.[1] p_j^d is the price set by the seller for the category d and is not observed by the econometrician. Consumers with characteristics d are supposed to face the same transaction price p_j^d. The case where prices differ across unobserved groups is considered in Section 3.4.2 below. We also assume that ξ_j^d is common to all individuals from group d. This was shown by Berry and Haile (2014) to be necessary for identifying demand models nonparametrically from aggregated data. As typical in the literature, the idiosyncratic error terms ε_{ij}^d are extreme-value distributed.

We make the usual parametric assumption about the intra-group heterogeneity. Specfiically, the individual parameters can be decomposed linearly into a mean, an individual deviation from the mean and a deviation related to individual characteristics:

$$\begin{cases} \beta_i^d = \beta_0^d + \pi_0^{X,d} E_i + \Sigma_0^{X,d} \zeta_i^X \\ \alpha_i^d = \alpha_0^d + \pi_0^{p,d} E_i + \Sigma_0^{p,d} \zeta_i^p, \end{cases}$$

where E_i denotes demographic characteristics that are unobserved by the firm for each purchaser, but whose distribution is common knowledge. $\zeta_i = (\zeta_i^X, \zeta_i^p)$ is a random vector with a specified distribution, such as the standard multivariate normal distribution.

The utility function satisfies $U_{ij}^d = \delta_j^d(p_j^d) + \mu_j^d(E_i, \zeta_i, p_j^d) + \varepsilon_{ij}^d$, with the mean utility $\delta_j^d(p_j^d)$ and the individual deviation $\mu_j^d(E_i, \zeta_i, p_j^d)$ satisfying

$$\begin{aligned} \delta_j^d(p_j^d) &= X_j' \beta_0^d + \alpha_0^d p_j^d + \xi_j^d, \\ \mu_j^d(E_i, \zeta_i, p_j^d) &= X_j' \left(\pi_0^{X,d} E_i + \Sigma_0^{X,d} \zeta_i^X \right) + p_j^d \left(\pi_0^{p,d} E_i + \Sigma_0^{p,d} \zeta_i^p \right). \end{aligned} \tag{1}$$

We let the dependence of δ_j^d and μ_j^d in p_j^d explicit for reasons that will become clear below. Because of the logistic assumption on the ε_{ij}^d, the aggregate market share $s_j^d(p^d)$ of good j for group d satisfies, when prices are set to $p^d = (p_1^d, ..., p_J^d)$,

$$s_j^d(p^d) = \int s_j^d(e, u, p^d) dP_{E,\zeta}^d(e, u), \tag{2}$$

where $P_{E,\zeta}^d$ is the distribution of (E, ζ) for group d and

$$s_j^d(e, u, p^d) = \frac{\exp\left(\delta_j^d(p_j^d) + \mu_j^d(e, u, p_j^d)\right)}{\sum_{k=0}^J \exp\left(\delta_k^d(p_k^d) + \mu_k^d(e, u, p_k^d)\right)}.$$

Now, we consider a Bertrand competition model where firms are able to price discriminate by setting different prices to each of the n_D groups of consumers. Letting $\mathcal{J}$ denote the set of

[1] Following the literature, we focus for simplicity on a single market. The different groups cannot be seen as different markets because the $(\xi_j^d)_{d=1,\dots,n_D}$ are allowed to be correlated. In case of multiple markets, product characteristics, including prices and costs, vary with markets. Also, provided that the distribution of random coefficients is constant across markets, we could include product and market fixed effects in the X_j.

products sold by a firm, the profit of this firm when the vector of all prices for group d is p^d satisfies

$$\Pi = M \sum_{d=1}^{n_D} P(D=d) \sum_{j\in\mathcal{J}} s_j^d(p^d) \times \left(p_j^d - c_j^d\right),$$

where $P(D=d)$ is the fraction of the group d in the population, $s_j^d(p^d)$ is the market share of product j for group d when prices are equal to p^d and M is the total number of potential consumers. c_j^d is the marginal cost of the product j for group d.

The first-order condition stemming from the profit maximization for group d yields

$$p_j^d = c_j^d + \left[(\Omega^d)^{-1} s^d\right]_j, \tag{3}$$

where $[.]_j$ indicates that we consider the j-th line of the vector only. Ω^d is the matrix of typical (j,k) term equal to $-\partial s_k^d/\partial p_j$ when j and k belong to the same firm, 0 otherwise. The firms set prices optimally by making the traditional arbitrage between increasing prices and lowering sales. When a monopoly seller is able to price discriminate, it is less constrained than with a uniform pricing strategy since this arbitrage is made for each group separately. If a group is particularly price sensitive, the monopoly seller offers a low price and is still able to extract a large surplus from the less price sensitive group by setting a higher price for this group. In a competitive setting, this effect is mitigated by the fact that, for a given group of consumers, the competition among sellers is reinforced.

Unlike standard demand estimation, our estimation method below relies both on the demand and supply models. This is necessary because we only partially observe prices. Note that an assumption on conduct is necessary anyway to run counterfactual simulations, which is most often the main purpose in empirical industrial organization. We focus here on the most standard model of competition between firms with differentiated products, but our methodology can be adapted directly to different competitive settings and collusive models. On the other hand, it cannot be applied to models of second degree price discrimination where sellers do not observe consumers' characteristics but can propose different levels of quality that, like prices, are unobserved by the econometrician.[2] In such cases, the sellers offer menus of contracts to induce consumers to self-select, implying a dependence between prices and consumers' unobserved preferences. We rule out such a dependence here.

2.2 Identifying assumptions

To identify the model, we crucially rely on the following three conditions.

Assumption 1. *(Observability of market shares) For all d and j, we observe s_j^d.*

Assumption 2. *(Constant marginal costs across consumers) For all d and j, $c_j^d = c_j$.*

[2] This applies as well to bargaining, where quality corresponds to the time needed for the consumer to obtain a given discount.

Assumption 3. *(Partial observation of prices) For all $j \in \{1, ..., J\}$, we observe $\widetilde{p}_j$ such that*

$$\widetilde{p}_j = f_j(p_j^1, ..., p_j^{n_D}),$$

where f_j is known and satisfies $f_j(0, ..., 0) = 0$ and for all $(c, u^1, ..., u^{n_D}) \in \mathbb{R}^{n_D+1}$,

$$f_j(c + u^1, ..., c + u^{n_D}) = c + f_j(u^1, ..., u^{n_D}). \tag{4}$$

The first requirement to apply our methodology is to observe market shares of products for all groups of consumers. This basically means that the econometrician and the sellers have the same information about consumers. We show that we can relax this assumption, at the cost of additional restrictions on the heterogeneity across consumers (see Section 3.4.2).

Assumption 2 amounts to neglecting differences in the costs of selling to different consumers in the total cost of a product. This assumption is likely to be satisfied in many settings. In the automobile market, production costs do not vary with consumers' type, and distribution costs are also likely to be the same for all consumers. However in some markets, this assumption might be more problematic. This is the case for insurance providers that offer different prices to consumers based on their observable characteristics (e.g., age, gender, driving experience), because those characteristics imply different risk classes and different costs for insurers. Note that our method can still be applied if the econometrician observes the risk classes. We show in Section 3.4.1 that we can introduce cost differences when they are due to observables and vary across products.

Let us turn to Assumption 3. First, note that $f_j(0, ..., 0) = 0$ is a mere normalization. If it does not hold, one can simply replace $\widetilde{p}_j$ by $\widetilde{p}_j - f_j(0, ..., 0)$, hand the corresponding modified function f_j then satisfies this requirement. Condition (4), on the other hand, is not a normalization. It is nonetheless satisfied in several settings. First, suppose that we only observe the maximum of all transaction prices, so that

$$\widetilde{p}_j = \max\left(p_j^1, ..., p_j^{n_D}\right). \tag{5}$$

Then f_j satisfies Equation (4). Such a case arises when firms post their highest discriminatory price $\widetilde{p}_j$ and then offer some discounts according to observable characteristics of buyers in order to reach optimal discriminatory prices. This assumption is in line with empirical evidence on the automobile market, showing in particular that some consumers pay the posted prices (for France and the UK, see, e.g., the reports of L'Observatoire Cetelem in 2013 and the UK Competition Commission in 2000). Furniture, kitchens, mobile phone contracts are other examples for which there is either documented or anecdotal evidence that some consumers receive some discounts over the posted prices. More generally, Shelegia and Sherman (2015) provide evidence, through a field experiment in Austria, that discounting is common in varied retail shops. Moreover, even if observed, these discounts may not correspond directly to price reductions, but rather to non-pecuniary benefits that are difficult to value in monetary terms (e.g. more flexibility, extended warranty, free shipping, coupon for a future purchase). In such cases, our method below is well-suited to identify the monetary equivalent of these advantages for consumers. We refer to, e.g., Grubb and Osborne (2015, pp.240-241) for an example of such advantages on mobile phone contracts for college students.

While Equation (5) imposes that the maximal transaction price corresponds to the list price, the group that pays this maximal price (the pivot group hereafter) is neither supposed to be known ex ante, nor constant across different products. Finally, in the case where the minimal rebate r_j is not zero, Assumption 3 still holds provided that the econometrician observes r_j. The price $\widetilde{p}_j$ then simply corresponds to the list price minus r_j. On the other hand, r_j cannot be identified with the method we propose below.

Another case where Assumption 3 holds is when we observe, through survey data for instance, the price paid by at least one consumer group for each product. This is typically the case with survey data where, for a given consumer, the price paid for the chosen product is observed, while the prices of the other available products are not. With such data, it may be possible to reconstruct all the prices when the numbers of products and consumer groups are small, but this is typically not the case when the number of groups and products are large. If we consider for instance panel data on grocery items for which there is spatial price discrimination, the number of geographical areas (n_D) is potentially very large. It is then unlikely to observe the prices of all products within each location. A similar problem is likely to arise if there are many products available, as, again, in the automobile market (see, e.g., Langer, 2016; Allcott and Wozny, 2014; Murry, 2017, for papers relying on such data in this market). If we do not observe all the transaction prices $(p_j^d)_{d=1,\dots,n_D}$ corresponding to product j, but at least one price $p_j^{d_j}$ for each j, then we can use $\widetilde{p}_j = p_j^{d_j}$ and apply our methodology since Assumption 3 holds.

A third case for which Assumption 3 holds is when we only observe the average price paid over all consumers for each product. This is the case if we only have access to sales revenue and units sold for each product within the market. Such data are typically available from marketing companies or company tax declarations. Then we observe $\widetilde{p}_j = \sum_{d=1}^{n_D} s_j^d p_j^d$, which, again, satisfies Assumption 3.

As a side remark, since it can be costly to gather transaction prices for all the consumer groups and products, our method below constitutes a way to reduce the collection of such transaction data, while still being able to use them for demand estimation. This can be particularly relevant for quantitative analysis in antitrust cases such as merger analysis and evaluation of damages due to anti-competitive practices.

3 Inference

3.1 Bias from ignoring price discrimination

First, let us recall the standard case where the true prices are observed. Let

$$\theta_0^d = (\beta_0^d, \alpha_0^d, \pi_0^{X,d}, \Sigma_0^{X,d}, \pi_0^{p,d}, \Sigma_0^{p,d})$$

denote the true vector of parameters for group d. The standard approach for identification and estimation of θ_0^d, initiated by BLP, is to use the exogeneity of Z_j, which includes the characteristics X_j and other instruments (typically, function of characteristics of other products or cost shifters) to derive moment conditions involving θ_0^d. The exogeneity condition takes the form $E\left[Z_j \xi_j^d\right] = 0$. The idea is then to use the link between ξ_j^d and the true parameters θ_0^d through Equation (2). Specifically, we know from Berry (1994) that for any given vector θ^d, Equation (2), where θ_0^d is replaced by θ^d, defines a bijection between market shares and mean utilities of products δ_j^d. Hence, we can define $\delta_j^d(s^d, p^d; \theta^d)$, where $s^d = (s_1^d, ..., s_J^d)$ denotes the vector of observed market shares. Once $\delta_j^d(s^d, p^d; \theta^d)$ is obtained, the vector $\xi_j^d(p^d; \theta^d)$ of unobserved characteristics corresponding to θ^d and rationalizing the market shares follows easily since

$$\xi_j^d(p^d; \theta^d) = \delta_j^d(s^d, p^d; \theta^d) - X_j \beta^d - \alpha^d p_j^d.$$

The moment conditions used to identify and estimate θ_0^d are then

$$E\left[Z_j \xi_j^d(p^d; \theta_0^d)\right] = 0. \tag{6}$$

When the observed prices are different from the true prices paid by consumers, the former approach is not valid in general. To see this, consider the simple logit model, where $\pi_0^{X,d}, \Sigma_0^{X,d}, \pi_0^{X,d}$ and $\Sigma_0^{X,d}$ are known to be zero. In this case $\delta_j^d(s^d, p^d; \theta^d)$ takes the simple form

$$\delta_j^d(s^d, p^d; \theta^d) = \ln s_j^d - \ln s_0^d$$

and does not depend on p^d. In this context, using posted prices $\widetilde{p}$ instead of the true prices amounts to relying on

$$\xi_j^d(\widetilde{p}; \theta^d) = \ln s_j^d - \ln s_0^d - X_j \beta^d - \alpha^d \widetilde{p}_j,$$

instead of relying on $\xi_j^d(p^d; \theta^d)$. A problem arises because $\widetilde{p}_j - p_j^d$ is not a classical measurement error. The true price depends on the characteristics of the good and of the cost shifters. If, for instance, a group of consumers values particularly the horsepower of automobiles, powerful cars will be priced higher for this group, and $\widetilde{p}_j - p_j^d$ will be negatively correlated with horsepower. Because horsepower is one of the instruments, we have $E[Z_j(\widetilde{p}_j - p_j^d)] \neq 0$, and $E\left[Z_j \xi_j^d(\widetilde{p}; \theta_0^d)\right]$ is no longer equal to zero. In the general random coefficient model, this problem is still present but in addition to it, $\delta_j^d(s^d, p^d; \theta^d)$ generally depends on p^d. Thus, Z_j is also correlated with $\delta_j^d(s^d, \widetilde{p}; \theta^d) - \delta_j^d(s^d, p^d; \theta^d)$.

To illustrate this issue, we estimated the usual BLP model on simulated data satisfying our assumptions. The detailed results, presented in Section 2.4 of our supplementary material, show that the biases on key parameters are not only large but also unpredictable. The errors on average mark-ups are up to 78% and vary a lot, depending on the group of consumers and the function f_j we consider. A similar conclusion holds on the preference parameters, with average errors up to 70% of the true values. Errors on average price elasticities are lower, but can still reach around 8%.

3.2 Consistent GMM estimation

Instead of simply replacing p^d by $\widetilde{p}$, we use the supply model together with Assumptions 1-3 to obtain consistent estimators. The idea is first to compute, for a given value of the vector of parameters $\theta = (\theta^1, ..., \theta^{n_D})$, the transaction prices $p_j^d(\theta)$ that rationalize the market shares and the supply-side model. Precisely, Equation (3) and Assumptions 2-3 imply that

$$\widetilde{p}_j = c_j + f_j \left(\left[(\Omega^1)^{-1} s^1 \right]_j, ..., \left[(\Omega^{n_D})^{-1} s^{n_D} \right]_j \right). \tag{7}$$

Then, the discriminatory prices satisfy

$$p_j^d = \widetilde{p}_j - f_j \left(\left[(\Omega^1)^{-1} s^1 \right]_j, ..., \left[(\Omega^{n_D})^{-1} s^{n_D} \right]_j \right) + \left[(\Omega^d)^{-1} s^d \right]_j, \tag{8}$$

which shows that for a given vector of parameters θ, the discriminatory prices are identified up to the matrices Ω^d. Now, taking the derivative of the market share function (Equation (2)) with respect to the price p_j^d, we obtain

$$\frac{\partial s_j^d}{\partial p_j^d}(p^d) = \int \left(\alpha_0^d + \pi_0^{p,d} e + \Sigma_0^{p,d} u \right) s_j^d(e, u, p^d)(1 - s_j^d(e, u, p^d)) dP_{E,\zeta}^d(e, u) \tag{9}$$

We obtain a similar expression for $\partial s_j^d / \partial p_l^d(p^d)$. These expressions show that Ω^d only depends on the parameters θ_0^d, the vector of prices p^d and $\delta^d = (\delta_1^d, ..., \delta_J^d)$, through $s_j^d(e, u, p^d)$. We emphasize this dependence by writing $\Omega^d(\theta_0^d, p^d, \delta^d)$.

Besides, the observed vector of market shares $s = (s_1^1, ..., s_J^1, ..., s_1^{n_D}, ..., s_J^{n_D})$ satisfies

$$s_j^d = \int \frac{\exp\left(\delta_j^d + \mu_j^d(e, u, p_j^d)\right)}{\sum_{k=0}^J \exp\left(\delta_k^d + \mu_k^d(e, u, p_k^d)\right)} dP_{E,\zeta}^d(e, u). \tag{10}$$

By Berry (1994), for any vector p^d of transaction prices, there exists a unique δ_θ^d such that Equation (10) holds. We denote by $\delta_\theta^d(p^d)$ this solution. Let

$$\begin{aligned} g_{\theta,j}^d(p) = \widetilde{p}_j - f_j & \left(\left[\Omega^1(\theta^1, p^1, \delta_\theta^1(p^1))^{-1} s^1 \right]_j, ..., \left[\Omega^{n_D}(\theta^{n_D}, p^{n_D}, \delta_\theta^{n_D}(p^{n_D}))^{-1} s^{n_D} \right]_j \right) \\ & + \left[\Omega^d(\theta^d, p^d, \delta_\theta^d(p^d))^{-1} s^d \right]_j \end{aligned}$$

and $g_\theta(p) = (g^1_{\theta,1}(p), ..., g^1_{\theta,J}(p), ..., g^{n_D}_{\theta,1}(p), ..., g^{n_D}_{\theta,J}(p))$. Then Equation (8) implies that the true vector of transaction prices $p(\theta) = (p^1_1(\theta), ..., p^1_J(\theta), ..., p^{n_D}_1(\theta), ..., p^{n_D}_J(\theta))$ satisfies

$$g_\theta(p(\theta)) = p(\theta).$$

This suggests that we can obtain $p(\theta)$ as the fixed point of g_θ. However, it is unclear at this stage whether $p(\theta)$ is the sole fixed point of g_θ. Even if this is the case, finding this fixed point may be difficult in practice, given the potentially large dimension of the problem. In our application, for instance, $p \in \mathbb{R}^{571\times 6}$. To solve both problems, we prove that if the heterogeneity on price sensitivity is not too large, g_θ is a contraction.[3] This result is similar to that of BLP, who exhibit a contraction mapping for Equation (10) in δ when transaction prices are known.

Theorem 1. *Suppose that f_j is 1-Lipschitz for all j. Then, for any $\overline{p} > 0$ there exists $\overline{\Sigma}^p > 0$ and $\overline{\pi}^p > 0$ such that for all $\Sigma^p = (\Sigma^{p,1}, ..., \Sigma^{p,n_D}) \in [0, \overline{\Sigma}]^{n_D}$ and $\pi^p = (\pi^{p,1}, ..., \pi^{p,n_D}) \in [-\overline{\pi}^p, \overline{\pi}^p]^{n_D}$, g_θ is a contraction on $[0, \overline{p}]^{Jn_D}$.*

This theorem ensures not only that there is a unique fixed point to g_θ, but also that the sequence $(p_n)_{n\in\mathbb{N}}$ defined by a given p_0 and $p_{n+1} = g_\theta(p_n)$, for all $n \in \mathbb{N}$, always converges to $p(\theta)$, irrespective of p_0. The result relies on two conditions. The Lipschitz condition on f_j, first, holds in the first three examples we mentioned above, namely $f_j(p^1, ..., p^{n_D}) = \max(p^1, ..., p^{n_D})$, $f_j(p^1, ..., p^{n_D}) = \sum_{d=1}^{n_D} s^d_j p^d / \sum_{d=1}^{n_D} s^d_j$ and $f_j(p^1, ..., p^{n_D}) = p^{d_j}$. The second condition is that $p^d_j \leq \overline{p}$ for all (j, d). Note however that we place no restriction on $\overline{p}$.

Though the proof of Theorem 1 is technical (see Appendix B.1), the intuition behind it is simple. Without heterogeneity on price sensitivity ($\Sigma^{p,d} = \pi^{p,d} = 0$ for all d), the function g_θ is constant, since both δ_θ and $\Omega^d(\theta, \cdot, \delta^d)$ are constant. Then g_θ is obviously a contraction with a Lipschitz coefficient of 0. The contraction result then holds because we can prove that this coefficient moves continuously with all the $(\Sigma^{p,d}, \pi^{p,d})$. Another interesting consequence of the observation that g_θ is constant when $\Sigma^{p,d} = \pi^{p,d} = 0$ is that the convergence of the aforementioned sequence $(p_n)_{n\in\mathbb{N}}$ is immediate in this case. We can then expect quick convergence with moderate values of $\Sigma^{p,d}$ and $\pi^{p,d}$.

We can apply the GMM to identify and estimate $\theta_0 = (\theta^1_0, ..., \theta^{n_D}_0)$. Let $\delta^d_j(s, \theta)$ and $p^d_j(s, \theta)$ denote the mean utility and price of product j for group d when market shares and the vector of parameters are respectively equal to s and θ. Let also

$$G^d_J(\theta) = \frac{1}{J} \sum_{j=1}^{J} Z_j \left(\delta^d_j(s, \theta) - X_j \beta^d - \alpha^d p^d_j(s, \theta) \right)$$

denote the empirical counterpart of the moment conditions corresponding to Equation (6).[4]

[3] A contraction is a K-Lipschitz function with $K < 1$, where we recall that a function g is K-Lipschitz if for all p, p', $\|g(p) - g(p')\| \leq K\|p - p'\|$. The norm we consider here is the supremum norm.

[4] In case of multiple markets, we consider averages not only on products but also on markets.

Let $G_J(\theta) = (G_J^1(\theta)', ..., G_J^{n_D}(\theta)')'$ and define

$$Q_J(\theta) = G_J(\theta)' W_J G_J(\theta),$$

where W_J is a positive definite matrix. Our GMM estimator of θ_0 finally satisfies $\widehat{\theta} = \arg\min_\theta Q_J(\theta)$.

Compared to the estimation of the standard BLP model, the computation of $\widehat{\theta}$ requires to optimize over a larger space. In the standard BLP model where s_j^d is observed for each group d and true prices are observed or supposed to be equal to posted prices, optimization can be performed for each group separately. Further, we only need to optimize over $(\pi^{X,d}, \Sigma^{X,d}, \pi^{p,d}, \Sigma^{p,d})$, because we can easily concentrate the objective function with respect to (α^d, β^d), by running two-stage least squares of δ_j^d on (p_j^d, X_j) instrumented by Z_j. In our case, we cannot estimate θ^d separately from $\theta^{d'}$, for $d' \neq d$, because $\theta^{d'}$ matters for determining $p_j^d(s, \theta)$, as Equation (8) shows. Also, while we can concentrate the objective function with respect to β^d, we cannot do this with α^d, as it appears in Equation (8). Second, for each θ, we need to solve not only Equation (10), but also simultaneously Equation (8), in order to obtain both the mean utilities and the transaction prices. Therefore, estimating the model is computationally more costly than estimating the standard random coefficient model.

That said, it is possible to lower the computational burden. We describe in details our algorithm based on Theorem 1 in Appendix A. To investigate both its performance and reliability, we also display the results of Monte Carlo simulations in Section 2 of the supplement. We show therein that the fixed point algorithm suggested above works well in practice, and we confirm the idea of quick convergence with moderate values of unobserved heterogeneity on prices. We also show that the optimization problem corresponding to the GMM remains feasible in a reasonable amount of time, and that the estimator is accurate in practice.

3.3 Test of the model

The assumption that firms practice price discriminate may be debatable for some markets, especially when direct evidence based on transaction prices is not available. We now develop a formal test of the model of price discrimination against the model of uniform pricing, where prices are supposed to satisfy $p_j^1 = ... = p_j^{n_D} = \widetilde{p}_j$. The idea is to consider a demand model nesting both. Specifically, let us define

$$r_j^d = f_j\left(\left[(\Omega^1)^{-1} s^1\right]_j, ..., \left[(\Omega^{n_D})^{-1} s^{n_D}\right]_j\right) - \left[(\Omega^d)^{-1} s^d\right]_j.$$

Under the price discrimination model, $r_j^d = \widetilde{p}_j - p_j^d$. We then consider the following demand model:[5]

$$U_{ij}^d = X_j \beta_i^d + \left[\widetilde{p}_j - r_j^d \kappa\right] \alpha_i^d + \xi_j^d + \varepsilon_{ij}^d. \tag{11}$$

This model nests both models since under the discrimination model, $\kappa = 1$, while under the uniform pricing model, $\kappa = 0$. Note that we exclude here the very special case where $r_j^d = 0$ for all (j, d). This happens if firms could price discriminate, but the groups of consumers are all identical, resulting in equal transaction prices. In such a case, the two models are actually identical and our test cannot be implemented since κ is not identified.

We therefore consider the test of H_0: $\kappa = 0$ versus H_1: $\kappa = 1$. We treat H_0 and H_1 symmetrically, so that the errors of first and second types should be (approximately) identical. To construct such a test, we first estimate r_j^d using our unobserved price discrimination model. Then we estimate κ by the standard BLP demand model corresponding to utilities defined by Equation (11), replacing r_j^d by its estimates. In a third step, we compute consistent estimators $\widehat{\sigma}_k^2$ of the asymptotic variance of $\widehat{\kappa}$ under H_k ($k = 0, 1$). $\widehat{\sigma}_0^2$ is simple to obtain because one can show that the estimation of the discount in the first stage does not have any effect on the standard error of $\widehat{\kappa}$ under H_0. It does have an effect, however, under H_1. In such a case, $\widehat{\kappa}$ may be seen as a two-step GMM estimator, and we can then apply the corresponding standard formula (see, e.g., Newey and McFadden, 1994). Finally, we compute the test statistic T defined by

$$T = J \left[\left(\frac{\widehat{\kappa}}{\widehat{\sigma}_0} \right)^2 - \left(\frac{\widehat{\kappa} - 1}{\widehat{\sigma}_1} \right)^2 \right].$$

T would simply be the likelihood ratio test of H_0 versus H_1 if $\widehat{\kappa} \sim \mathcal{N}(\kappa, \sigma_\kappa^2/J)$. We consider tests where we accept H_0 if $T < s$, H_1 otherwise. Instead of finding the threshold s^* such that the errors of first and second types are (asymptotically) identical, it is simpler to compare the p-values p_0 and p_1 under both hypotheses. The following proposition formalizes this idea.[6]

Proposition 1. *Suppose that $\sqrt{J}\widehat{\kappa}/\widehat{\sigma}_0 \sim \mathcal{N}(0, 1)$ under H_0 and $\sqrt{J}(\widehat{\kappa} - 1)/\widehat{\sigma}_1 \sim \mathcal{N}(0, 1)$ under the alternative, with $\widehat{\sigma}_0 \neq \widehat{\sigma}_1$. Then the test where we accept H_0 if $p_0 > p_1$ and H_1*

[5] Equation (11) corresponds to a more general demand model than that considered in Section 2.1. It could be interpreted as a model where discounts are given randomly to consumers with a uniform probability κ. In our setting where discounts are unobserved, however, we cannot separately identify κ from the discounts themselves, and therefore we cannot estimate a demand model of this kind.

[6] The test of Rivers and Vuong (2002) has often been used in the literature to discriminate between alternative supply-side models, given a consistent demand estimation, by comparing the corresponding R^2 of the cost equations (see, e.g., Jaumandreu and Moral, 2006; Bonnet and Dubois, 2010; Ferrari and Verboven, 2012). The advantage of this latter test is that it can be applied even if both models are misspecified. In our context, the test statistic would consist in taking the standardized difference between the GMM objective functions of the two models. The main problem in applying such a test is to obtain a consistent estimator of the standard error of this difference, i.e. check Assumption 8 in Rivers and Vuong (2002). When both models are wrong, the estimated error terms of the two models depend in general on all the $(s_j^d)_{j=1,\dots,J}$ and are therefore not independent from each other, even asymptotically. Moreover, the dependence between these error terms has an unknown form. Thus, neither the standard GMM formula based on independence, nor the standard bootstrap, would deliver consistent estimators of the standard error of the aforementioned difference.

otherwise is symmetric in both hypotheses. Moreover,

$$p_0 = \mathbb{1}\{\widehat{\sigma}_1 > \widehat{\sigma}_0\} + sgn(\widehat{\sigma}_1 - \widehat{\sigma}_0)\left(\Phi(r_1) - \Phi(r_2)\right),$$
$$p_1 = \mathbb{1}\{\widehat{\sigma}_1 < \widehat{\sigma}_0\} + sgn(\widehat{\sigma}_1 - \widehat{\sigma}_0)\left(\Phi(r_4) - \Phi(r_3)\right),$$

where $r_1 \leq r_2$ *(resp.* $r_3 \leq r_4$*) are the two roots of* $x \mapsto (\widehat{\sigma}_1^2 - \widehat{\sigma}_0^2)x^2 + 2\sqrt{J}\widehat{\sigma}_0 x - (J + \widehat{\sigma}_1^2 T)$ *(resp.* $x \mapsto (\widehat{\sigma}_1^2 - \widehat{\sigma}_0^2)x^2 + 2\sqrt{J}\widehat{\sigma}_1 x + (J - \widehat{\sigma}_0^2 T)$*). If there is no such root,* $p_0 = 1 - p_1 = \mathbb{1}\{\widehat{\sigma}_1 > \widehat{\sigma}_0\}$*.*

The advantage of our test is that it confronts one model against the other. We could simply perform a test of $\kappa = 0$ versus $\kappa \neq 0$, or $\kappa = 1$ versus $\kappa \neq 1$. These would rather correspond to specification tests of the uniform pricing model and price discrimination model, respectively. Of course, a more direct approach for specification testing is through standard overidentification GMM tests.

3.4 Extensions

3.4.1 Allowing for cost differences

Our methodology relies on Assumption 2, which supposes constant marginal costs across groups of consumers. This assumption might not be valid in some settings. When the demand segmentation is based on geographic variables, the costs can vary across groups because of varying prices in production factors. In some cases, the transportation cost from the factory to the retailers is significant and vary within the territory. We can then relax the constant marginal cost assumption by considering a model in which the cost of product j for group d satisfies:

$$c_j^d = c_j + W_j^{d\prime}\lambda,$$

where W_j^d are cost shifters that vary across products. In this case, Equation (8) becomes:

$$\widetilde{p}_j = c_j + f_j\left(\left[(\Omega^1)^{-1}s^1\right]_j + W_j^{1\prime}\lambda, ..., \left[(\Omega^{n_D})^{-1}s^{n_D}\right]_j + W_j^{n_D\prime}\lambda\right).$$

Then the optimal price for group d is:

$$p_j^d = \widetilde{p}_j - f_j\left(\left[(\Omega^1)^{-1}s^1\right]_j + W_j^{1\prime}\lambda, ..., \left[(\Omega^{n_D})^{-1}s^{n_D}\right]_j + W_j^{n_D\prime}\lambda\right) + \left[(\Omega^d)^{-1}s^d\right]_j + W_j^{d\prime}\lambda.$$

Our method then applies as previously, with λ one of the component of θ. When W_j^d does not vary with j, on the other hand, we may not be able to separately identify λ from the intercept of the utility function. Intuitively, we can rationalize any price gap (constant across products) between groups by either a marginal cost difference or a corresponding difference in the intercepts of the utility function.

3.4.2 Inference with unobserved groups

We sketch here how the previous ideas can be adapted to cases where only the aggregated market shares s_j, rather than the $(s_j^d)_{d=1,...,n_D}$, are observed. More details on this case are provided in Section 1.1 of the supplement. This setting is in particular relevant for combining aggregate data on sales with survey data where both consumers' characteristics and transaction prices are observed. Another important example is when only data on total revenue and total quantities are available. Supermarket scanner data, for instance typically report weekly revenues and units sold for all grocery items. The corresponding average prices then hide temporary promotions.

To handle such cases, we assume that consumers are homogeneous inside each group. Consumers inside a group only differ in their product-specific terms, which, as usual, are supposed to be i.i.d. and extreme-value distributed. We also suppose that the unobserved preference terms ξ_j^d are the same across groups of consumers. A similar assumption is also key for identification in Berry et al. (1995), as discussed in Berry and Haile (2014).

Assumption 4. *$\Sigma^p = 0$, $\pi^p = 0$ and $\xi_j^1 = ... = \xi_j^{n_D}$ for all $j \in \{1, ..., J\}$.*

The demand model features a discrete unobserved heterogeneity, with n_D points of support on the random coefficients. Such a model has also been used by Berry and Jia (2010) and Kalouptsidi (2012). Under Assumption 4, the market share of product j for consumer group d given the vector of prices $p = (p_1^1, ..., p_J^1, ..., p_1^{n_D}, ..., p_J^{n_D})$, the vector of unobserved preferences $\xi = (\xi_1, ..., \xi_J)$ and $\theta = (\alpha^1, \beta^1, ..., \alpha^{n_D}, \beta^{n_D})$ satisfies

$$s_j^d(p, \xi, \theta) = \frac{\exp(X_j'\beta^d + \alpha^d p_j^d + \xi_j)}{\sum_{k=0}^J \exp(X_k'\beta^d + \alpha^d p_k^d + \xi_k)}.$$

As a result, the aggregate market share of product j is:

$$s_j(p, \xi, \theta) = \sum_{d=1}^{n_D} \phi^d s_j^d(p, \xi, \theta), \tag{12}$$

where $\phi^d = \Pr(D = d)$ can be assumed to be known or added to the vector of parameters θ.

For a given θ and vector of transaction prices p, the system of nonlinear equations in $(\xi_1, .., \xi_J)$ given by (12) can be seen as a particular case of the system studied by Berry (1994), with the $(\xi_1, .., \xi_J)$ playing the role of the $(\delta_1, ..., \delta_J)$ in his setting. By his result, the market share function is invertible and there is a unique solution to this system. Hence, we can define the vector $\xi(p, s, \theta)$ of the $(\xi_1, ..., \xi_J)$ corresponding to transaction prices p, the vector of observed market shares $s = (s_1, ..., s_J)$ and θ. Note that the condition $\xi_j^1 = ... = \xi_j^{n_D}$ is key here to invert the market share equations and obtain $\xi(p, s, \theta)$.

We do not observe transaction prices, however, so we cannot compute directly $\xi(p, s, \theta)$ to form the moment conditions. Instead and as above, we solve both for ξ and p, using not only the market share equations but also the first-order conditions on prices. Because there is no unobserved individual heterogeneity inside groups, these first-order conditions are simply:

$$p_j^d = c_j - \frac{1}{\alpha^d(1-\sum_{k\in\mathcal{J}_j} s_k^d(p,\xi,\theta))}, \tag{13}$$

where $\mathcal{J}_j$ denotes the set of products sold by the same firm as the one selling j. These first-order conditions imply:

$$\begin{aligned} p_j^d =& \widetilde{p}_j + f_j\left(\frac{1}{\alpha^1(1-\sum_{k\in\mathcal{J}_j} s_k^1(p,\xi,\theta))},...,\frac{1}{\alpha^{n_D}(1-\sum_{k\in\mathcal{J}_j} s_k^{n_D}(p,\xi,\theta))}\right) \\ &- \frac{1}{\alpha^d(1-\sum_{k\in\mathcal{J}_j} s_k^d(p,\xi.\theta))}. \end{aligned} \tag{14}$$

Replacing ξ by $\xi(p,s,\theta)$ in this equation, we obtain that the vector of prices p is the fixed point of the function $M_{s,\theta} = (M_{s,\theta,1}^1,...,M_{s,\theta,J}^1,...,M_{s,\theta,1}^{n_D},...,M_{s,\theta,J}^{n_D})$ defined by

$$\begin{aligned} M_{s,\theta,j}^d(p) =& \widetilde{p}_j - \frac{1}{\alpha^d(1-\sum_{k\in\mathcal{J}_j} s_k^d(p,\xi(p,s,\theta),\theta))} + f_j\left(\frac{1}{\alpha^1(1-\sum_{k\in\mathcal{J}_j} s_k^1(p,\xi(p,s,\theta),\theta))},...,\right. \\ &\left.\frac{1}{\alpha^{n_D}(1-\sum_{k\in\mathcal{J}_j} s_k^{n_D}(p,\xi(p,s,\theta),\theta))}\right). \end{aligned}$$

For any $\xi \in \mathbb{R}^J$, let $p(\xi,\theta)$ denote the vector of equilibrium prices, provided that there is a unique solution to the system defined by Equations (12) and (13). Then let

$$s(\xi,\theta) = (s_1(p(\xi,\theta),\xi,\theta),...,s_J(p(\xi,\theta),\xi,\theta))$$

denote the market shares corresponding to ξ and $p(\xi,\theta)$. In a similar way as Theorem 1 above, we show in Section 1.1 of the supplement that for at least some θ, the sequence $(p_n)_{n\in\mathbb{N}}$ defined by $p_0 = \widetilde{p}_j$ and $p_{n+1} = M_{s(\xi,\theta),\theta}(p_n)$ converges towards $p(\xi,\theta)$. Once $p(\xi,\theta)$ is obtained, we can compute the corresponding ξ by the standard BLP inversion, and then compute the GMM objective function in the same way as in Section 3.2. We provide details on the estimation algorithm and the results of Monte Carlo simulations in Section 2.6 of the supplement. There, we assume n_D and the proportions $(\phi^d)_{d=1,...,n_D}$ to be known. This is the case when the econometrician knows the groups that are used for price discrimination and their proportion in the population, but does not observe their specific demand (e.g. male/female in our application, if there is price discrimination with respect to gender).

3.4.3 Other extensions

We consider other extensions in our supplement. First, we show how to adapt our methodology to other supply models and how to include additional moment conditions from the supply side. Second, we consider cases where discrimination is based on an unobserved characteristic but a proxy of this unobserved characteristic is available. Then we provide details on how one can leverage individual-level data in addition to aggregate-level market shares. Finally, we discuss how to account for a nonlinear effect of price on utility.

4 Application to the French new car market

4.1 Data and methodology

We apply our methodology to estimate demand and supply together with unobserved discounts in the new automobile industry in France. Automobile sellers are well-known to price discriminate, negotiate or offer discounts over the sticker price to close the deal. As in our theoretical model, we only observe here posted prices that come from manufacturers catalogues. Apart from such posted prices, we use data on all the registrations of new cars in France between 2003 and 2008. We observe the corresponding car attributes as well as the municipality of residence and the age of the car owner.

We define groups of buyers by interacting three age classes and two income classes. We choose the commonly used thresholds of 40 and 60 for the age classes, and 27,000 euros per year for the income classes. The age is presumably a strong determinant of purchase, and it seems plausible that the seller can observe the age class of the buyer even if they do not know each other before the transaction. Income is also likely to affect preferences for different car attributes and price sensitivity. Income is, however, likely to be unobserved by the seller but instead inferred. We compute a predictor of buyer's income, namely the median household income in his age class and in his municipality. We therefore assume that both age and this income predictor are used by the automobile sellers to price discriminate.[7]

Hereafter, we estimate the random coefficient model with uniform pricing and with unobserved price discrimination. In the latter, we assume that the observed price, i.e. the posted price, is the maximal price among all prices paid by the different groups of consumers. In all specifications, we control for the main characteristics of the cars such as horsepower, weight, the cost of driving 100 kilometers, dummies for station wagon body-style and three doors in the utility function. Finally, we include year and brand dummies, constraining the coefficients of the latter to be identical for all demographic groups. We allow for unobserved heterogeneity of preferences inside groups of consumers in terms of price sensitivity, constraining again the distribution of this coefficient to be identical for all demographic groups in order to obtain more accurate results.[8]

In addition to exogenous characteristics, we include the following instruments for car j. The first is the weight of j multiplied by a composite price index that aims at approximating the

[7] We refer to Section 3.1 in our supplement for more details on the data and evidence of heterogeneity in purchase patterns across consumer groups.

[8] We estimated a specification with heterogeneity on the fuel cost and for the utility of buying a new car. These coefficients turned out to be imprecisely estimated, so we preferred to drop them.

average input price. Specifically, we use a weighted average of steel, aluminium and plastic prices taken in January 2015. The weights we use are equal to 0.77, 0.11 and 0.12, respectively, reflecting the relative importance of each of these inputs in car manufacturing. The other instruments are close to those suggested by BLP. We include the sum of the continuous exogenous characteristics (namely weight, horsepower and fuel cost) of the other brands' products. We also consider the sums of these characteristics over the other brands' products in the same segment as j, supposing that they are closer substitutes to j.[9] Finally, we include the sums of the continuous characteristics of the other products belonging to the same segment as j and produced by the brand producing j.[10]

4.2 Parameter estimates and comparison with the standard model

The results for the models with uniform pricing and unobserved price discrimination are presented in Table 1. The two models produce different price sensitivities. They are always smaller for the price discrimination model, except for the group of old with high income for which we obtain a higher price sensitivity under the uniform pricing model. This group is the least price sensitive group and turns out to be always pivot in the model with unobserved price discrimination. More generally, price sensitivity decreases with both age and income.

The parameters of the intercept are negative, reflecting the fact that the major part of consumers choose the outside option, namely not to buy a car or buy one on the second-hand market. The heterogeneity of this parameter across demographic groups does not follow a clear pattern. As expected, consumers display a preference for horsepower, but the groups differ in how much they value it. Young consumers have a high valuation for the engine power while the eldest care less about this attribute. As expected, all groups of consumers dislike large fuel expenses. The parameters of sensitivity to the fuel cost are consistent with the parameters of sensitivity to the car price. The old purchasers with high income appear to be also less sensitive to the cost of driving while the most sensitive consumers are also the young and middle-age groups with a low income. As weight is a proxy for the size and the space of the car, it is positively valued by all the consumers. Three doors and station wagon vehicles are negatively valued, reflecting that most of the consumers buy sedan or hatchback cars with five doors (four doors plus the trunk).

[9] See Table 14 for the details of the segmentation we use.

[10] Armstrong (2016) has shown that such instruments could be weak when the number of products is large. Note however that identification is secured here by the inclusion of the cost shifter. Nonetheless, we checked that the instruments are relevant for prices. We used for that purpose the F-statistic of the joint nullity of the coefficients of the instruments in the linear regression of prices on the characteristics and these instruments. We obtained $F \simeq 24.1$, which is far above the threshold of 10 suggested by Staiger and Stock (1997) and usually used to detect weak instruments. This is therefore reassuring on the identification of the model and the validity of inference here.

	Uniform		Price discrimination	
	Parameter	Std-err	Parameter	Std-err
Price sensitivity				
Age $<$ 40, I = L	-4.57**	0.149	-4.83**	0.12
Age $<$ 40, I = H	-4.33**	0.145	-4.52**	0.119
Age $\in$ [40,59], I = L	-3.98**	0.137	-4.32**	0.118
Age $\in$ [40,59], I = H	-3.73**	0.133	-3.96**	0.116
Age $\geq$ 60, I = L	-3.85**	0.149	-4.21**	0.133
Age $\geq$ 60, I = H	-3.61**	0.15	-3.05**	0.134
Std. dev. (σ^p)	1.15**	0.058	0.98**	0.086
Intercept				
Age $<$ 40, I = L	-5.4**	0.211	-6.24**	0.208
Age $<$ 40, I = H	-6.25**	0.209	-6.92**	0.207
Age $\in$ [40,59], I = L	-6.12**	0.212	-6.85**	0.208
Age $\in$ [40,59], I = H	-6.36**	0.21	-6.9**	0.207
Age $\geq$ 60, I = L	-5.73**	0.228	-6.48**	0.226
Age $\geq$ 60, I = H	-6.25**	0.236	-6.31**	0.282
Horsepower				
Age $<$ 40, I = L	3.88**	0.214	2.68**	0.179
Age $<$ 40, I = H	3.13**	0.191	2.09**	0.166
Age $\in$ [40,59], I = L	2.27**	0.195	1.87**	0.161
Age $\in$ [40,59], I = H	1.65**	0.17	1.37**	0.162
Age $\geq$ 60, I = L	1.11**	0.202	1.12**	0.178
Age $\geq$ 60, I = H	0.64**	0.225	0.28	0.251
Fuel cost				
Age $<$ 40, I = L	-6.08**	0.179	-5.5**	0.181
Age $<$ 40, I = H	-5.1**	0.171	-4.63**	0.176
Age $\in$ [40,59], I = L	-5.24**	0.173	-4.97**	0.17
Age $\in$ [40,59], I = H	-4.18**	0.164	-4.03**	0.17
Age $\geq$ 60, I = L	-3.51**	0.171	-3.46**	0.17
Age $\geq$ 60, I = H	-2.73**	0.174	-2.59**	0.177
Weight				
Age $<$ 40, I = L	5.67**	0.218	6.63**	0.221
Age $<$ 40, I = H	5.83**	0.213	6.61**	0.22
Age $\in$ [40,59], I = L	5.7**	0.218	6.69**	0.22
Age $\in$ [40,59], I = H	5.55**	0.211	6.28**	0.215
Age $\geq$ 60, I = L	4.55**	0.23	5.59**	0.243
Age $\geq$ 60, I = H	4.64**	0.24	4.13**	0.289
Three doors				
Age $<$ 40, I = L	0.09	0.199	0.16	0.209
Age $<$ 40, I = H	-0.05	0.197	0.02	0.208
Age $\in$ [40,59], I = L	-0.05	0.196	-0.04	0.206
Age $\in$ [40,59], I = H	-0.2	0.199	-0.18	0.209
Age $\geq$ 60, I = L	-0.52**	0.194	-0.53**	0.205
Age $\geq$ 60, I = H	-0.59**	0.194	-0.51*	0.203
Station wagon				
Age $<$ 40, I = L	-0.74**	0.131	-0.75**	0.144
Age $<$ 40, I = H	-0.61**	0.13	-0.61**	0.143
Age $\in$ [40,59], I = L	-0.64**	0.13	-0.66**	0.142
Age $\in$ [40,59], I = H	-0.71**	0.132	-0.72**	0.143
Age $\geq$ 60, I = L	-0.73**	0.128	-0.76**	0.14
Age $\geq$ 60, I = H	-0.72**	0.128	-0.65**	0.132
Value of objective function	2,343		1,739	

Notes: Significance levels: †: 10% *: 5% **: 1%. Standard errors are computed using the standard GMM formula. "Horsepower" is the fiscal horsepower, "Fuel cost" is in euros/10 kilometres and "Weight" is in tons. Year and brand fixed effects and the within-group heterogeneity parameter of price sensitivity are constrained to be identical across groups of consumers.

Table 1: Parameter estimates for the standard uniform BLP model and our model with unobserved price discrimination.

If qualitatively similar, the results we obtain with the two models exhibit some quantitative differences, as we further illustrate below. It is therefore important to discriminate between the two models. Recall that in the test developed in Section 3.3, $\kappa = 0$ corresponds to uniform pricing, while $\kappa = 1$ corresponds to the price discrimination model. We estimate $\hat{\kappa} = 1.08$ and obtain $p_0 = 0$ and $p_1 = 0.52$.[11] Hence, this test clearly points towards the discrimination model over the uniform pricing model.

To understand what the differences of the estimates between the two models imply, we compare the corresponding price elasticities and mark-up rates. As Table 2 shows, price elasticities are, in absolute terms, lower for the model with uniform pricing for all groups, except for the pivot group, namely the group assumed to pay the posted prices. Hence the overestimation of prices in the uniform pricing model is more than compensated by the underestimation of price sensitivity parameters.

In the discriminatory pricing model, we find average price elasticities varying from -3.9 to -6.4. Such elasticities are in line with those obtained by BLP (between -3.5 and -6.5) but below those of Langer (2016) who finds, using transaction prices, a range between -6.4 to -17.8. Our price elasticities imply an average mark-up of 20.6% under the price discrimination model and 21.6% under the uniform pricing model, with, as we could expect, sizable heterogeneity across groups in the price discrimination model. As in the simulations (see Table 4 in the supplement), the uniform pricing model underestimates the mark-up firms obtain on the pivot group but overestimates the mark-ups of the other groups. The average mark-up for the group of young, low-income consumers is around 17.6%, contrasting with the 28.5% the firms obtain for the old and high-income group. Similarly, the costs are always overestimated in the uniform pricing model, with an average difference of 9.5%. The relative cost differences even exceed 18% for 2.9% of the products. We refer to Section 3.3 of the supplementary material for more details.

Group of consumers	Price elasticity		Average mark-up		Average surplus	
	Disc.	Unif.	Disc.	Unif.	Disc.	Unif.
Age < 40, I $=$ L	-6.4	-6.15	17.5	21.4	13,220	16,552
Age < 40, I $=$ H	-6.18	-5.89	18.2	21.1	14,473	18,261
Age $\in$ [40,59], I $=$ L	-5.99	-5.28	18.9	21.2	15,465	21,018
Age $\in$ [40,59], I $=$ H	-5.53	-4.92	20.4	21.3	18,480	25,411
Age ≤ 60, I $=$ L	-5.52	-4.75	20.7	22.2	15,574	20,774
Age ≤ 60, I $=$ H	-3.94	-4.5	28.5	22	32,442	26,590
Average	-5.61	-5.2	20.6	21.6	17,916	21,651

Reading notes: Mark-ups are in percentage.

Table 2: Comparison of average price elasticities, mark-ups and consumer surplus under the uniform pricing and unobserved price discrimination models.

We investigate further the differences between the two models by looking at the results of two counterfactual exercises. First we measure the welfare effects of a purchase subsidy for young households that are below 40. We consider three policy designs: (i) a uniform subsidy of €1,000, (ii) a subsidy of €1,000 for cars that are more fuel efficient than the average (148 cars

[11] The standard errors $\widehat{\sigma}_0$ and $\widehat{\sigma}_1$ are equal to 0.07 and 0.92, respectively. $\widehat{\sigma}_0$ is much smaller than $\widehat{\sigma}_1$ because under H_0, the variance of the discounts has no effect on the asymptotic variance of $\widehat{\kappa}$, and it turns out that the variance of the discounts has an important effect on $\widehat{\sigma}_1$.

out of 571) and (iii) a feebate system that provides a rebate of €1,000 for cars that are more fuel efficient than the average and a tax of €1,000 for the other cars. Results are displayed in Table 3. All the scenarios imply welfare effects qualitatively similar but quantitatively very different. Under the uniform pricing we always obtain lower effects on profits, consumers and policy cost. The differences are the most striking for the variation in consumer surplus: 30%, 44% and 92% for respectively scenario (i), (ii) and (iii).

Secondly, we measure the welfare implications of a potential merger between manufacturers. We investigate the unilateral effects of a merger between Peugeot group (PSA) and the European branch of General Motor (Opel and Vauxhall).[12] The results, displayed in the last two columns of Table 3, reveal that the two alternative models imply welfare effects that differ by 1.5 million euros. We estimate the total welfare loss to be 18% lower under the uniform pricing model than under the model with price discrimination. The effects on profits differ by 57%, the uniform pricing model implying larger impacts on the profitability of the industry.

	Subsidy for young buyers						Merger	
	scenario (i)		scenario (ii)		scenario (iii)		PSA/GM	
	Disc	Unif	Disc	Unif	Disc	Unif	Disc	Unif
ΔConsumer surplus	959.59	674.32	588.23	327.49	284.92	21.56	-9.56	-8.56
ΔProfits	161.05	151.64	107.48	102.87	69.89	67.88	0.99	1.56
# firms better off	20	19	15	15	9	8	19	19
Policy cost	269.09	255.88	17.02	16.2	13.25	12.18	0	0
ΔWelfare	851.55	570.08	678.69	414.16	341.56	77.26	-8.57	-7
ΔProfits for PSA/GM							0.34	0.53

Reading notes: All monetary values are in million euros. The differences in consumer surplus and profits are summed over all consumers and firms, respectively. The average surplus CS^d for group d is computed in euros, using the standard formula $\text{CS}^d = \int \log\left(1 + \sum_{j=1}^{J} \exp(\delta_j^d + \mu_j(u, p_j^d))\right) dF_\varsigma(u)$. We thus take into account the substitution from the outside good. The welfare is computed as the sum of manufacturers' profits and all consumers' surplus, minus the cost of the policy.

Table 3: Welfare analysis of the hypothetical purchase subsidy for young households and a hypothetical merger between PSA and Opel.

4.3 Analysis of the discounts

Table 4 presents the average discount for each demographic group estimated using the model with unobserved price discrimination. We compute average discounts weighted by actual sales in each group but also using the same weighting scheme for all groups of consumers, namely, the overall product market shares ("basket-weighted" method). This allows us to eliminate the potential group-specific demand composition effect. The results with both weighting methods are nevertheless very similar. As expected, the pattern on average discounts across groups is similar to that on price elasticity. The estimated pivot group is identical for all the products

[12]This merger analysis is inspired by acquisition of Opel by PSA in the beginning of March 2017. Note that we do not pretend that our results are credible to evaluate this merger as our results use market conditions from 2007 and are no longer relevant.

and corresponds to the group with the lowest price elasticity. These are the 13.2% of the population over 60 years of age with income over 27,000 euros.

On average, the sales-weighted discount is 9.6%, with a large heterogeneity across consumers. Around 25% of transactions occurred with a discount greater than or equal to 12.2%. Clearly, income and age are both important determinants of the discount obtained. On average, young, low-income purchasers pay 13.4% less than the posted price, while young, high-income buyers get an average discount of 12.0%. These percentages represent a gross gain of around 2,500 euros. Middle-aged consumers get smaller discounts (11.4% for the low-income group and 9.6% for the high-income group). Finally, whereas old, low-income individuals receive an average discount of 10.4%, the old, high-income buyers receive no discount since they constitute the pivot group for all the products.

	Average discount (in % of posted price)		Average gross discount (in euros)	
Group of consumers	Sales-weighted	Basket-weighted	Sales-weighted	Basket-weighted
Age $<$ 40, I = L	13.3	13.53	2,594	2,813
Age $<$ 40, I = H	12.01	12.27	2,523	2,568
Age $\in$ [40,59], I = L	11.36	11.33	2,482	2,385
Age $\in$ [40,59], I = H	9.56	9.53	2,156	2,032
Age $\geq$ 60, I = L	10.37	10.28	2,084	2,174
Age $\geq$ 60, I = H	0	0	0	0
Average	9.64	9.68	2,023	2,038

Reading notes: the "basket-weighted" discounts are obtained by using the same artificial basket of cars for all groups.

Table 4: Average discounts by group of consumers

To put these discounts into perspective, we provide a rough assessment of the importance of third- versus second-degree price discrimination. We define the latter as the variations in prices of the different versions of a given car model, fuel type, body style and year. These different versions of a car model are associated to different characteristics such as cylinder capacity or horsepower for instance. We first compare the sales-weighted variance of list prices without second-degree price discrimination, thus considering only baseline models, with the variance of list prices that includes second-degree price discrimination. We find that second-degree price discrimination increases the variance of log list prices by 2.7%. We then turn to third-degree price discrimination. Using our estimated prices, we observe a further increase in the variance of 2.3%.[13] We can make a similar assessment on relative price ranges, defined for a given car model name, fuel type and body style as the ratio between the maximal and minimal prices minus one. While the average relative price range with second-degree price discrimination only is equal to 38%, this average relative price range reaches 56% when introducing third-degree price discrimination. Hence, at the end of the day, third-degree price discrimination

[13] Because we do not define products at the finest possible level in our model in order to measure markets shares with enough precision, we do not have a specific estimate of discounts for each version of the different car models. To compute such an estimate, we assume that the discounts represent the same percentage of the list prices for all the different versions of a car model.

appears to be a determinant of price dispersion nearly as important as second-degree price discrimination.

Further results are displayed in Section 3.3 of the supplement. In particular we analyze the heterogeneity of discounts across car models (Section 3.3.2). We also investigate the importance of price discrimination on firms and consumers in Section 3.3.3. Counterfactual simulations suggest that compared to an equilibrium with uniform pricing, price discrimination matters. Most firms and consumers gain from discrimination, but the high-income group of consumers older than 60 is significantly worse off.

4.4 Plausibility of the results

Note first that a direct comparison with data on discounts for each product and consumer group is not possible since, to the best of our knowledge, such data are not available. Nevertheless, we confirm indirectly the plausibility of our results using data from the 2006 French consumer expenditure survey of Insee (BdF survey hereafter, for "Budget des Familles"). Similar data was used by Goldberg (1996) in her study on price discrimination against women and minorities. Each household in this survey must indicate whether they bought a new car in the three years before the survey (2004, 2005 or 2006). If so, they indicate how much they paid for it, the brand and model's names (e.g., Volkswagen Golf) and the type of fuel. We also observe the age of the head of the household, the administrative region and the type of urban area of their residence.[14]

Let $\widetilde{X}$ denote the characteristics of the car and its owner that are available both in the BdF survey and in the CCFA. $\widetilde{X}$ includes the brand name and model, the type of fuel, the owner's age, region and urban area. Because the BdF survey is representative, the average prices we estimate with our demand and supply model and those observed in the survey (conditional on $\widetilde{X}$) should match, if the model is correct. Formally, under the hypothesis that the price discrimination model holds, we have

$$E(p^{BdF}|\widetilde{X}) = E(p^{CCFA}|\widetilde{X}). \tag{15}$$

To test the equality described by (15), we first compute $p(\widetilde{X}) \equiv E(p^{CCFA}|\widetilde{X})$ using the CCFA data and our estimates of the transaction prices. Then (15) may be seen as a standard nonparametric specification test $E(p^{BdF}|p(\widetilde{X})) = p(\widetilde{X})$. We rely on Yatchew's differencing test (see Yatchew, 1998, Section 4.2.1), which has the advantage of not relying on any tuning parameter. Note that we ignore that $p(.)$ itself is estimated here. This means that the test is overrejecting, and thus plays against our model. Given that the CCFA database is much larger than the BdF sample, this is most likely not a first-order concern here.

[14]There were 22 administrative regions at this period and there are five types of urban area defined by the size of the population: less than 5,000 inhabitants, between 5,000 and 20,000, between 20,000 and 100,000, more than 100,000 without Paris and Paris agglomeration.

Equality (15) also has two simple implications. First, prices should be equal on average, i.e. $E(p^{BdF}) = E(p^{CCFA})$. Second, we should have $E(p^{BdF} - p(\widetilde{X})|\widetilde{X}) = 0$. We can test the latter by considering the linear regression of $p^{CCFA} - p(\widetilde{X})$ on $\widetilde{X}$, or components of $\widetilde{X}$, and testing whether all the coefficients are equal to zero.

The results of these different tests are displayed in Table 5. We obtain for the non-parametric test a $t-$statistic of 0.52 when using our estimated transaction prices, meaning that we accept the null hypothesis that (15) holds at all standard levels. Conversely, the $t-$statistic is equal to 3.50 when considering list prices instead of our discounted prices. We also observe that the average of our estimated transaction prices is very close to the average price obtained in BdF, while the average list price is clearly higher. Finally, the parametric test using owner characteristics also indicates that our estimated transaction prices are not significantly different from BdF prices, whereas the test of equality with list prices is rejected. When controlling for both car and owner characteristics, the tests of equality are both rejected at standard levels. But again, the test statistic of equality between list prices and BdF prices is much larger than that involving discounted prices. Thus, at the end of the day, these tests suggest that our estimated transaction prices are indeed reasonable approximations of the true transaction prices.

Finally, we provide additional robustness checks in Section 3.4 of the supplement. In particular, we estimate our model and the uniform pricing one with a nested logit structure instead of random coefficients. We obtain similar results. We also perform sensitivity analyses of our results with respect to temporary promotions, price discrimination based on unobserved characteristics, spatial price dispersion, old cars as trade-ins, etc. Our results suggest that these issues are not of first-order concern here.

	$p^{CCFA} =$ discounted prices	$p^{CCFA} =$ list prices
Nonparametric test (t-stat)	0.52	3.50^{**}
$E(p^{BdF} - p^{CCFA})$ (in euros)	112	$-1{,}899^{**}$
Linear reg. on owner characteristics (F-test)	1.46	8.82^{**}
Linear reg. on owner & car characteristics (F-test)	2.85^{**}	8.48^{**}

Reading notes: owner characteristics include the age, the type of urban area of residence and the year of purchase. Car characteristics include the brand, the list price and type of fuel. Significance levels: †: 10%, *: 5%, **: 1%. Data source: Budget des familles survey - 2005-2006, INSEE, Centre Maurice Halbwachs (CMH).

Table 5: Comparison with BdF data.

5 Conclusion

This paper investigates the pervasive issue of partial observation of prices in structural models of demand and supply in markets with differentiated products. We propose an approach that incorporates unobserved price discrimination by firms based on observable individual characteristics. We use this model to estimate demand and supply on the French new car market where price discrimination may occur through discounts. Our results suggest significant discounting by manufacturers, in line with other evidence on this market.

While we have considered several extensions of our baseline model, we have maintained the assumption that consumer groups are fixed ex ante. Yet, in several cases, relevant characteristics of the consumers are unobserved by the sellers, who then offer menus of contracts to price discriminate consumers. Adapting our methodology to such second-degree price discrimination does not seem obvious, and is left for future research.

References

Allcott, H. and Wozny, N. (2014), 'Gasoline prices, fuel economy, and the energy paradox', *Review of Economics and Statistics* **96**(5), 779–795.

Armstrong, T. B. (2016), 'Large market asymptotics for differentiated product demand estimators with economic models of supply', *Econometrica* **84**(5), 1961–1980.

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

Berry, S. (1994), 'Estimating discrete-choice models of product differentiation', *The RAND Journal of Economics* **25**, 242–262.

Berry, S. and Haile, P. (2014), 'Identification in Differentiated Products Markets Using Market Level Data', *Econometrica* **82**(5), 1749–1797.

Berry, S. and Jia, P. (2010), 'Tracing the woes: An empirical analysis of the airline industry', *American Economic Journal: Microeconomics* **2**(3), 1–43.

Berry, S., Levinsohn, J. and Pakes, A. (1995), 'Automobile prices in market equilibrium', *Econometrica* **63**, 841–890.

Bonnet, C. and Dubois, P. (2010), 'Inference on vertical contracts between manufacturers and retailers allowing for nonlinear pricing and resale price maintenance', *The RAND Journal of Economics* **41**(1), 139–164.

Carter, M. (2001), *Foundations of Mathematical Economics*, MIT Press.

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.

Crawford, G. S. and Yurukoglu, A. (2012), 'The welfare effects of bundling in multichannel television markets', *American Economic Review* **102**(2), 643–85.

D'Haultfœuille, X., Durrmeyer, I. and Février, P. (2018), Supplement to automobile prices in market equilibrium with unobserved price discrimination.

Dubé, J.-P., Fox, J. T. and Su, C.-L. (2012), 'Improving the numerical performance of static and dynamic aggregate discrete choice random coefficients demand estimation', *Econometrica* **80**(5), 2231–2267.

Dubois, P. and Lasio, L. (2018), 'Identifying industry margins with unobserved price constraints: structural estimation on pharmaceuticals', *American Economic Review* **forthcoming**.

Ferrari, S. and Verboven, F. (2012), 'Vertical control of a distribution network : an empirical analysis of magazines', *The RAND Journal of Economics* **43**(1), 26–50.

Gill, D. and Thanassoulis, J. (2009), 'The impact of bargaining on markets with price takers: Too many bargainers spoil the broth', *European Economic Review* **53**(6), 658–674.

Gill, D. and Thanassoulis, J. (2016), 'Competition in posted prices with stochastic discounts', *The Economic Journal* **126**(594), 1528–1570.

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

Gowrisankaran, G., Nevo, A. and Town, R. (2015), 'Mergers when prices are negotiated: Evidence from the hospital industry', *The American Economic Review* **105**(1), 172–203.

Grennan, M. (2013), 'Price discrimination and bargaining: Empirical evidence from medical devices', *The American Economic Review* **103**(1), 145–177.

Grubb, M. D. and Osborne, M. (2015), 'Cellular service demand: Biased beliefs, learning, and bill shock', *American Economic Review* **105**(1), 234–71.

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

Huang, G. (2016), When to haggle, when to hold firm? lessons from the used car retail market. Working paper.

Jaumandreu, J. and Moral, M. J. (2006), Identifying behaviour in a multiproduct oligopoly: Incumbents reaction to tariffs dismantling. Working paper.

Jindal, P. and Newberry, P. (2016), 'To bargain or not to bargain: The role of fixed costs in price negotiations', *Journal of Marketing Research* .

Kalouptsidi, M. (2012), 'From market shares to consumer types: Duality in differentiated product demand estimation', *Journal of Applied Econometrics* **27**(2), 333–342.

Langer, A. (2016), (dis)incentives for demographic price discrimination in the new vehicle market. Working paper.

Lee, J. and Seo, K. (2016), 'Revisiting the nested fixed-point algorithm in blp random coefficients demand estimation', *Economics Letters* **149**, 67–70.

L'Observatoire Cetelem (2013), L'automobile en europe: 5 leviers pour rebondir, Technical report, http://observatoirecetelem.com/wp-content/uploads/2013/08/observatoire-cetelem-automobile-2013.pdf.

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.

Murry, C. (2017), Advertising in vertical relationships: An equilibrium model of the automobile industry. Working paper.

Newey, W. K. and McFadden, D. (1994), Large sample estimation and hypothesis testing, *in* 'Handbook of econometrics, vol.4', Elsevier, pp. 2111–2245.

Rivers, D. and Vuong, Q. (2002), 'Model selection tests for nonlinear dynamic models', *The Econometrics Journal* **5**, 1–39.

Scott Morton, F., Zettelmeyer, F. and Silva-Risso, J. (2003), 'Consumer information and discrimination: Does the internet affect the pricing of new cars to women and minorities?', *Quantitative Marketing and Economics* **1**(1), 65–92.

Shelegia, S. and Sherman, J. (2015), When the price you see is not the price you get: A field study. Working Paper.

Staiger, D. and Stock, J. H. (1997), 'Instrumental Variables Regression with Weak Instruments', *Econometrica* **65**(3), 557–586.

Thomadsen, R. (2005), 'The effect of ownership structure on prices in geographically differentiated industries', *RAND Journal of Economics* pp. 908–929.

UK Competition Commission (2000), New cars: A report on the supply of new motor cars within the uk, Technical report, http://webarchive.nationalarchives.gov.uk/20111202193730tf_/http://www.competition-commission.org.uk/rep_pub/reports/2000/439cars.htm.

Yatchew, A. (1998), 'Nonparametric regression techniques in economics', *Journal of Economic Literature* **36**(2), 669–721.

A Computation of the GMM estimator

We first provide additional details on how to compute our GMM estimator in practice. As in BLP, we rely on a nested fixed point (NFP) algorithm to solve the system of non-linear equations given by (8) and (10) for each value of θ,. Specifically, the algorithm involves the following steps:

1. Start from initial values for p^d, for each group d. We can use the observed prices $\widetilde{p}$ or previous transaction prices obtained for another θ.

2. Given the current vector of transaction prices p_n^d, compute $\delta_n^d = \delta(s^d, p_n^d; \theta^d)$ by inverting Equation (10).

3. Given δ_n^d and p_n^d, compute the corresponding matrix Ω^d and update the transaction prices, using Equation (8).

4. Iterate 2 and 3 until convergence of prices.

The construction of the moment conditions therefore involves two nested inner loops. The *price-loop* searches over the vector of prices for all the consumer groups. Inside the *price-loop*, the *delta-loop* searches over the mean utilities δ^d. For each value of transaction prices, we have to invert the market share equation to solve for the mean utility vectors δ^d. We use for that purpose the contraction mapping proposed by Lee and Seo (2016) which relies on Newton's method and converges more rapidly than BLPs contraction mapping.

If the computational cost of our algorithm is greater than for the BLP estimator, it is possible to parallelize the market share inversion as well as the computation of the mark-up terms ($(\Omega^d)^{-1}s^d$), as they are independent across markets and demographic groups. We also save time by updating the initial values of the mean utilities after each iteration of the inner *price-loop* and by updating initial values of prices across iterations of the outer loop that involves the parameters θ (for more details on computational aspects, see Section 2.2 in the supplement).

We do not rely on the minimization program with equilibrium constraints (MPEC) approach suggested by Dubé et al. (2012) here. Simulations suggest that in our set-up, this approach is much slower than our NFP algorithm. This result is consistent with the findings of Dubé et al. (2012), who report that with markets of size 500 or more, MPEC was less efficient than NFP for the standard BLP model. Here, we can easily reach problems of size 500, since the optimization problem has a size larger than $2Jn_D$.

Finally, we can reduce the computational cost by considering restricted versions of the model. In particular, the computation of the GMM estimator is much simpler if we assume no heterogeneity on price sensitivity within a group of consumers ($\Sigma_0^p = 0$ and $\pi_0^{p,d} = 0$), so that $\alpha_i^d = \alpha_0^d$. This assumption may be reasonable in particular if we have a fine segmentation of consumers. In this case, we still have to optimize over $\theta = (\theta^1, ..., \theta^{n_D})$. On the other hand, solving the system defined by Equations (8)-(10) is easy. Equation (10) reduces to the standard inversion of market shares, whereas Equation (8) provides an explicit expression for transaction prices, since Ω^d does not depend on p^d. Thus, the computational cost is significantly reduced compared to the general model.

Another restricted version of our model is when utility parameters do not vary with d. Then $\theta = \theta^1$ is of much lower dimension, making again the optimization much easier. This assumption is realistic if consumers preferences vary with individual characteristics E, but sellers only observe a proxy of E trough the discrete variable D. Then individual preferences are independent of D conditional on E. Note that price discrimination with respect to D is still relevant for sellers because the distribution of E conditional on $D = d$ varies with d, i.e. the distribution of preferences differs in the n_D groups.

Finally, another alternative is to rely on the logit or nested logit models. In the simple logit model, we have seen above that the matrix Ω^d only depends on $(\alpha^1, ..., \alpha^{n_D})$. In the nested logit, it also depends on the parameters $(\sigma^1, ..., \sigma^{n_D})$ that drive substitutions within nests. But at the end, we also obtain a quite simple nonlinear optimization over $(\alpha^1, \sigma^1, ..., \alpha^{n_D}, \sigma^{n_D})$ only. We refer to Section 2.5 of the supplement for a detailed discussion on the computational and statistical performances of the GMM estimator with the logit and nested logit models.

B Proofs

B.1 Proof of Theorem 1

In the sequel, the components of θ other than $\Sigma^p = (\Sigma^{p,1}, ..., \Sigma^{p,n_D})$ and $\pi^p = (\pi^{p,1}, ..., \pi^{p,n_D})$ are held fixed. Hence, we identify θ (resp. θ^d) with (Σ^p, π^p) (resp. $(\Sigma^{p,d}, \pi^{p,d})$). We also let $K_r = [0, \overline{p}]^r$ for any $r \in \mathbb{N}$ and introduce the function

$$q_{\theta,j}^{d,d'}(p) = \left[\Omega^{d'}(\theta^{d'}, p^{d'}, \delta^{d'}_{\theta^{d'}}(p^{d'}))^{-1} s^{d'}\right]_j - \left[\Omega^{d}(\theta^{d}, p^{d}, \delta^{d}_{\theta^{d}}(p^{d}))^{-1} s^{d}\right]_j .$$

The proof is divided in three steps. First, we show that g_θ is a contraction if $q_{\theta,j}^{d,d'}(.)$ is a contraction, for all (j, d, d'). Second, we show that $(\theta, p) \mapsto q_{\theta,j}^{d,d'}(p)$ is continuously differentiable (C^1). Third, we prove that $q_{\theta,j}^{d,d'}(.)$ is a contraction for all $\theta = (\Sigma^p, \pi^p)$ in a neighborhood of 0.

1. g_θ is a contraction if $q_{\theta,j}^{d,d'}$ is a contraction, for all (j, d, d').

First, because we consider the supremum norm here, g_θ is a contraction if for all j, d, and θ in a neighborhood of 0, $g_{\theta,j}^d$ is a contraction. By Assumption 3, we have

$$g_{\theta,j}^d(p) - g_{\theta,j}^d(p') = f_j(q_{\theta,j}^{d,1}(p), ..., q_{\theta,j}^{d,n_D}(p') - f_j(q_{\theta,j}^{d,1}(p'), ..., q_{\theta,j}^{d,n_D}(p)).$$

By assumption, f_j is 1-Lipschitz for the supremum norm. As a result,

$$\|g_{\theta,j}^d(p') - g_{\theta,j}^d(p)\| \leq \max_{d'} \left| q_{\theta,j}^{d,d'}(p') - q_{\theta,j}^{d,d'}(p) \right| .$$

The first step follows.

2. $(\theta^d, p) \mapsto q_{\theta,j}^{d,d'}(p)$ **is** C^1**.**

First, we show that $(\theta^d, p^d) \mapsto \delta^d_{\theta^d}(p^d)$ is C^1. Let $\mu^d_{\theta,j}(e,u,p^d_j)$ be defined as in Equation (1), except that we let the dependence on θ explicit. Then let

$$s_{\theta^d}(e,u,p^d,\delta^d) = (s_{\theta^d,1}(e,u,p^d,\delta^d), ..., s_{\theta^d,J}(e,u,p^d,\delta^d)),$$

with

$$s_{\theta^d,j}(e,u,p^d,\delta^d) = \frac{\exp\left(\delta^d_j + \mu^d_j(e,u,p^d_j)\right)}{\sum_{k=0}^J \exp\left(\delta^d_k + \mu^d_k(e,u,p^d_k)\right)}.$$

Finally, let $Q_{\theta^d}(p^d,\delta^d) = (Q_{\theta^d,1}(p^d,\delta^d), ..., Q_{\theta^d,J}(p^d,\delta^d))$, with

$$Q_{\theta^d,j}(p^d,\delta^d) = \int s_{\theta^d}(e,u,p^d,\delta^d) dP^d_{E,\zeta}(e,u) - s^d_j.$$

Then $\delta^d_{\theta^d}(p^d)$ is defined by $Q_{\theta^d}(p^d, \delta^d_{\theta^d}(p^d)) = 0$. By the dominated convergence theorem, $(\theta^d, p^d, \delta^d) \mapsto Q_{\theta^d}(p^d,\delta^d)$ is C^1. Moreover,

$$\frac{\partial Q_{\theta^d,j}}{\partial \delta^d_k} = \int s_{\theta^d,j}(e,u,p^d,\delta^d\delta^d)\left(\mathbb{1}\{j=k\} - s_{\theta^d,k}(e,u,p^d,\delta^d\delta^d)\right) dP^d_{E,\zeta}(e,u).$$

Thus,

$$\begin{aligned}\sum_{k\neq j}\left|\frac{\partial Q_{\theta^d,j}}{\partial \delta^d_k}\right| &\leq \int s_{\theta,j}(\delta^d,p^d,u,e)\left(1 - s_{\theta,0}(\delta^d,p^d,u,e) - s_{\theta,j}(\delta^d,p^d,u,e)\right) dP^d_{E,\zeta}(e,u) \\ &< \frac{\partial Q_{\theta^d,j}}{\partial \delta^d_j}.\end{aligned}$$

In other words, the jacobian matrix of $\delta^d \mapsto Q_{\theta^d}(p^d,\delta^d)$ is diagonally dominant, and thus invertible. Hence, the conditions of the implicit function theorem hold, and $(\theta^d, p^d) \mapsto \delta^d_{\theta^d}(p^d)$ is C^1.

Second, for any products (i,j) produced by the same firm, the (i,j)-th term of the matrix $\Omega^d(\theta^d,p^d,\delta^d)$ satisfies

$$\begin{aligned}&\Omega^d_{i,j}(\theta^d,p^d,\delta^d) \\ &= \int \left(\alpha^d + \pi^{p,d}e + \Sigma^{p,d}u^p\right) s^d_{\theta^d,i}(e,u,\delta^d,p^d)(\mathbb{1}\{i=j\} - s^d_{\theta^d,j}(e,u,\delta^d,p^d))dP^d_{E,\zeta}(e,u).\end{aligned}$$

Then, by the dominated convergence theorem, $(\theta^d,p^d,\delta^d) \mapsto \Omega^d_{i,j}(\theta^d,p^d,\delta^d)$ is C^1 on $\mathbb{R}\times K_J \times \mathbb{R}^J$. This is also the case if i and j are not produced by the same firm, since in this case $\Omega^d_{i,j}(\theta^d,p^d,\delta^d)$ is simply equal to 0.

Third, the inverse mapping for matrices, $A \mapsto A^{-1}$, is C^1 on any subset of $\mathcal{A}^+ = \{A : \det(A) >$

$0\}$ or $\mathcal{A}^- = \{A : \det(A) < 0\}$. Let us show that

$$\left\{\Omega^d(\theta^d, p^d, \delta^d), (\theta^d, p^d) \in \mathbb{R} \times K^J\right\} \subset \mathcal{A}^+ \text{ or } \left\{\Omega^d(\theta^d, p^d, \delta^d), (\theta^d, p^d) \in \mathbb{R} \times K^J\right\} \subset \mathcal{A}^-. \tag{16}$$

Suppose this is not the case. Then, by the intermediate value theorem, $\det(\Omega^d(\theta^d, p^d, \delta^d)) = 0$ for some (θ^d, p^d). But a same reasoning as above shows that $\Omega^d(\theta^d, p^d, \delta^d)$ is diagonally dominant. Thus, it is invertible, a contradiction. Hence, (16) holds.

Finally, by the chain rule, $(\theta, p) \mapsto q^{d,d'}_{\theta,j}(p)$ is C^1.

3. For all (j, d, d') and $(\theta^d, \theta^{d'})$ in a neighborhood of 0, $q^{d,d'}_{\theta,j}$ is a contraction.

By the maximum theorem (see e.g. Carter, 2001, Theorem 2.3) and Step 2, the function

$$R^{d,d'}_j : (\theta^d, \theta^{d'}) \mapsto \max_{p \in K_J} \sum_{k=1}^{Jn_D} \left|\partial q^{d,d'}_{\theta,j} / \partial p_k(p)\right|$$

is continuous. Let $\theta_0 = (\theta^1_0, ..., \theta^{n_D}_0)$ be such that $\theta^d_0 = \theta^{d'}_0 = 0$. $q^{d,d'}_{\theta_0,j}$ is a constant function, since neither $\delta_{\theta_0}(p)$ nor $\Omega^d(\theta^d_0, p^d, \delta^d)$ depend on p^d. Hence, $R^{d,d'}_j(\theta^d_0, \theta^{d'}_0) = 0$. By continuity of $R^{d,d'}_j$, there exists $\overline{\Sigma}^d$ such that for all $\theta^d \in [0, \overline{\Sigma}^d] \times \{0\}$, $R^{d,d'}_j(\theta^d, \theta^{d'}_0) < 1$. $\pi^{p,d} \mapsto \max_{\Sigma^{p,d} \in [0, \overline{\Sigma}^d]} R^{d,d'}_j(\Sigma^{p,d}, \pi^{p,d})$ is also continuous and smaller than 1 at $\pi^{p,d} = 0$. Then there exists $\overline{\pi}^d$ such that for all $(\Sigma^{p,d}, \pi^{p,d}) \in [0, \overline{\Sigma}^d] \times [-\overline{\pi}^d, \overline{\pi}^d]$, $R^{d,d'}_j(\Sigma^{p,d}, \pi^{p,d}, \theta^{d'}_0) < 1$.

Repeating this argument for $\theta^{d'}$ instead of θ^d, we see finally that there exists $\overline{\Sigma}^{d,d'} > 0$ and $\overline{\pi}^{d,d'}$ such that for all $(\theta^d, \theta^{d'}) \in ([0, \overline{\Sigma}^{d,d'}] \times [-\overline{\pi}^{d,d'}, \overline{\pi}^{d,d'}])^2$, $R^{d,d'}_j(\theta^d, \theta^{d'}) < 1$. In turn, this implies that $q^{d,d'}_{\theta,j}$ is a contraction for all $\Sigma^p \in \left[0, \overline{\Sigma}^{d,d'}\right]^{n_D}$ and all $\pi^p \in \left[-\overline{\pi}^{d,d'}, \overline{\pi}^{d,d'}\right]^{n_D}$. The result follows.

B.2 Proof of Proposition 1

A symmetric test is to accept H_0 if $T < s^*$ and accept H_1 otherwise, with $P_{H_0}(T > s^*) = P_{H_1}(T \leq s^*)$. The first probability is strictly decreasing in s^* while the second is increasing in s^*. Hence, $t < s^*$ if and only if $P_{H_0}(T > t) > P_{H_1}(T \leq t)$. In other words, $T < s^*$ if and only if $p_0 > p_1$, and the test where we accept H_0 if $p_0 > p_1$ and accept H_1 otherwise is symmetric in both hypotheses. Now, to compute p_0 and p_1, note that under H_0,

$$P_{H_0}(T > t) = P(aZ^2 + bZ + c > 0),$$

where $Z \sim \mathcal{N}(0, 1)$ and with $a = \widehat{\sigma}^2_1 - \widehat{\sigma}^2_0$, $b = 2\sqrt{J}\widehat{\sigma}_0$ and $c = -J - \widehat{\sigma}^2_1 t$. The result follows using standard arguments on quadratic inequalities. The same reasoning applies for p_1.