Automobile Prices in Market Equilibrium with Unobserved Price Discrimination[*]

Xavier D'Haultfœuille[†]   Isis Durrmeyer[‡]   Philippe Février[§]

October 15, 2017

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) 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 by using supply-side conditions, provided that the marginal costs of producing and selling the goods do not depend on the characteristics of the buyers. 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 transaction prices, price discrimination, automobiles.

JEL codes: C51, D12, D22.

[†]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 third-degree price discrimination that is unobserved by the econometrician. We rely for that purpose not only on the demand model but also on the supply model, and two identifying assumptions. First, 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 second 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 two 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 these 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. Simulations confirm that this method works well, both computationally and statistically. Finally, we develop a formal test of our model versus the uniform pricing model, for which there is no unobserved price discrimination between consumers.

In our baseline model, 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 the consumers used by sellers to price discriminate. Nevertheless, we show that our method extends, under some conditions, to the case where only aggregate market shares are observed. We also 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. Next, we discuss how information on transaction prices for a subsample of products can be included in the estimation through the construction of additional micro-moments, as in Berry et al. (2004). Finally, while we focus in our baseline model on the standard utility specification and supply-side model, we show that our methodology extends to other functional forms and supply models.

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 on this precise issue). 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 show that price discrimination has a significant effect on consumers' surplus and firms' profits, and that accounting for it 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, and that third-degree price discrimination accounts for one third of the unexplained price dispersion in transaction prices. Other survey data and anecdotal evidence are also consistent with our findings.

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. 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 exogenous. 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 (2014), 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 (2014) 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 observed transaction prices or directly 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). Even with data on prices paid by each consumer our method is still useful to estimate demand because for a given consumer the prices of products that are not chosen 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 explains how to estimate the model with unobserved transaction prices. In Section 4, we present several extensions of the model and in particular the extension to the case of unobserved consumer groups. Section 5 describes our estimation algorithm and presents the results of Monte Carlo simulations. The application on the French new car market is developed in the sixth part of the paper. We conclude in Section 7. The appendix gathers all proofs. In the supplementary material, we consider additional simulation results and provide more details on the application.

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, as well as their corresponding preferences, such as the distributions of price sensitivities and preferences for car attributes. They price discriminate among these groups, in order to take advantage of the heterogeneity in preferences.

Specifically, heterogeneous 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. This is crucial, but not more restrictive than the assumption 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 can be expressed as a mean utility and an individual deviation from this mean:

$$U_{ij}^d = \delta_j^d(p_j^d) + \mu_j^d(E_i, \zeta_i, p_j^d) + \varepsilon_{ij}^d,$$

with the mean utility and the individual deviation satisfying

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

---

[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.

We let the dependence 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 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. Prices are optimally set by the firms 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.

2.2 Identifying assumptions

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

Assumption 1. *(Constant marginal costs across consumers) For all d and j, $c_j^d = c_j$.*

Assumption 2. *(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 $(c, u^1, ..., u^{n_D}) \in \mathbb{R}^{n_D+1}$,

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

Assumption 1 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, such as the automobile market, where most costs stem from producing, not selling the goods. 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 4.4 that we can introduce cost differences when they are due to observables and vary across products.

Let us turn to Assumption 2. 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 $\widetilde{p}_j - f_j(0, ..., 0)$. 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 (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. In the case where the minimal rebate r_j is not zero, Assumption 2 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 2 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, only the price paid for the chosen product is observed and not for the entire choice set faced by the consumer. If this is not an issue when the numbers of products and consumer groups are small, this is more complicated 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 virtually impossible to observe the transaction prices for 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 2 holds.

A third case for which Assumption 2 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 2.

As a side, 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

In this section, we assume that the econometrician observes the market shares s_j^d corresponding to each consumer group while the prices p_j^d are only partially observed. This is the case when some characteristics of the purchasers and the product characteristics are observed, but not the individual transaction prices. In our application for instance, we use data coming from the registration of new cars in France. While many characteristics of the cars and some individual characteristics (age and zip code of the municipality) are recorded in this dataset, neither

transaction prices nor list prices are included. One then needs to merge these data with prices from another source. Here we use prices from catalogues, arguing that they correspond to maximal prices. In other settings, it may be easier to obtain data from surveys (see e.g. Langer, 2016), in which case we would observe transaction prices p_j^d for every product and at least one group of consumers.

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

$$E\left[Z_j \xi_j^d\right] = 0. \tag{6}$$

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

$$E\left[Z_j \xi_j^d(p^d; \theta_0^d)\right] = 0. \tag{7}$$

Now, when the observed prices are different from the true prices paid by consumers, for example when posted prices are used instead of transaction prices, 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 the DGP considered in our simulations (see Section 5.2 below). The detailed results, presented in Section 1.1 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 1-2 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 1-2 imply that

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

Then, the discriminatory prices satisfy

$$p_j^d = \widetilde{p}_j - f_j\left(\left[(\Omega^1)^{-1} s^1\right]_j, ..., \left[(\Omega^{n_D})^{-1} s^{n_D}\right]_j\right) + \left[(\Omega^d)^{-1} s^d\right]_j, \tag{9}$$

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

$$\frac{\partial s_j^d}{\partial p_j^d}(p^d) = \int \left(\alpha_0^d + \pi_0^{p,d} e + \Sigma_0^{p,d} u\right) s_j^d(e, u, p^d)(1 - s_j^d(e, u, p^d)) dP_{E,\zeta}^d(e, u) \tag{10}$$

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

$$s_j^d = \int \frac{\exp\left(\delta_j^d + \mu_j^d(e, u, p_j^d)\right)}{\sum_{k=0}^J \exp\left(\delta_k^d + \mu_k^d(e, u, p_k^d)\right)} dP_{E,\zeta}^d(e, u). \tag{11}$$

By Berry (1994), for any vector p^d of transaction prices, there exists a unique δ_θ^d such that Equation (11) holds. We denote by $\delta_\theta^d(p^d)$ this solution. Let

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

and $g_\theta(p) = (g_{\theta,1}^1(p), ..., g_{\theta,J}^1(p), ..., g_{\theta,1}^{n_D}(p), ..., g_{\theta,J}^{n_D}(p))$. Then Equation (9) implies that the true vector of transaction prices $p(\theta) = (p_1^1(\theta), ..., p_J^1(\theta), ..., p_1^{n_D}(\theta), ..., p_J^{n_D}(\theta))$ 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.[2] This result is similar to that of BLP, who exhibit a contraction mapping for Equation (11) 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 $[0, \overline{p}]^{J n_D}$.*

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 $f_j(p^1, ..., p^{n_D}) = \max(p^1, ..., p^{n_D})$, $f_j(p^1, ..., p^{n_D}) = \sum_{d=1}^{n_D} s_j^d p^d / \sum_{d=1}^{n_D} s_j^d$ and $f_j(p^1, ..., p^{n_D}) = p^{d_j}$. The second condition is that $p_j^d \le \overline{p}$ for all (j, d). Note however that we place no restriction on $\overline{p}$.

Though the proof of Theorem 1 is technical (see Appendix A.1), the intuition behind 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

---

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

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 $\pi^{p,d}$. This is also what we find in the simulations in Section 5.4 below.

We can apply the GMM to identify and estimate $\theta_0 = (\theta_0^1, ..., \theta_0^{n_D})$. Let $\delta_j^d(s,\theta)$ and $p_j^d(s,\theta)$ 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

$$G_J^d(\theta) = \frac{1}{J} \sum_{j=1}^J Z_j \left(\delta_j^d(s,\theta) - X_j \beta^d - \alpha^d p_j^d(s,\theta) \right)$$

denote the empirical counterpart of the moment conditions corresponding to Equation (7).[3] 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 is then

$$\widehat{\theta} = \arg \min_\theta Q_J(\theta). \tag{12}$$

Compared to the estimation of the standard BLP model, estimating our model in practice raises two challenges. First, we have to optimize over a larger space than in the BLP setting. In the standard BLP model where we observe the market share of product j for each group d and true prices are observed or supposed to be equal to posted prices, we could optimize only on θ^d (abstracting from supply-side conditions), for each group separately. We even 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 (9) 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 (9). Second, for each θ, we need to solve not only Equation (11), but also simultaneously Equation (9), 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. We describe in details our algorithm based on Theorem 1 in Section 5 below. We also show that this optimization problem remains feasible in a reasonable amount of time.

We can also reduce the computational cost by considering restricted versions of the model. In

---

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

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 (9)-(11) is easy. Equation (11) reduces to the standard inversion of market shares, while Equation (9) 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 1.2 of the supplementary material for a detailed discussion on the computational and statistical performances of the GMM estimator with the logit and nested logit models.

3.3   Supply-side conditions

As in the standard BLP model, it is possible to include moments corresponding to the supply side by imposing some additional structure on marginal costs. Let X_j^s be the vector of cost shifters for product j. X_j^s may be different from X_j but typically share some common components. We may suppose for instance that the marginal costs are log-linear, $\ln(c_j) = X_j^s \gamma + \omega_j$, where ω_j denotes the unobserved cost shock. This shock is supposed to satisfy $E[Z_j^s \omega_j] = 0$, where Z_j^s denotes a vector of instruments for the supply side. As for the demand, we construct the moment conditions by first recovering the marginal cost $c_j(s, \theta)$ associated to s and a given vector of parameters θ. Specifically, by Equation (8),

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

We then obtain $\omega_j(s, \theta, \gamma)$ by

$$\omega_j(s, \theta, \gamma) = \ln\left(c_j(s, \theta) \right) - X_j^s \gamma.$$

The supply-side moment conditions are

$$G_J^s(\theta, \gamma) = \frac{1}{J} \sum_{j=1}^{J} Z_j^s \left[\ln\left(c_j(s, \theta)\right) - X_j^s \gamma \right].$$

We can proceed as previously, simply replacing $G_J(\theta)$ by

$$G_J(\theta, \gamma) = (G_J^1(\theta)', ..., G_J^{n_D}(\theta)', G_J^s(\theta, \gamma))'.$$

Whether or not we add moments from the supply side, it is important to note that our GMM estimator relies fundamentally on the supply side, contrary to the demand estimator of BLP. First-order conditions of the firms' program are key to recover the transaction prices $p(\theta)$, and in turn the residuals $\xi_j^d(p(\theta), \theta)$. Because the unobserved transaction prices depend on the nature of the competition between firms, it is impossible to estimate the demand without making an assumption on the supply side. We do not see this as a strong limitation, however, because the supply side is usually modeled in addition to the demand, since it is crucial to perform counterfactual analysis.

Related to this, we have considered here the standard set-up where firms and retailers are integrated and prices are fixed through a Bertrand competition. Our methodology generalizes straightforwardly to different supply-side models and competitive settings. In particular, it applies directly to models with collusion or vertical relations. In the latter case, our method constitutes the first step of the analysis, where the margins of the retailers are recovered. The second step corresponds to the modeling of vertical relations, and can incorporate any kind of vertical arrangement (bargaining, non-linear pricing...). Third-degree price discrimination on the downstream market only matters for the value of margins and the profits of the retailers.

3.4 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:

$$U_{ij}^d = X_j \beta_i^d + \left[\widetilde{p}_j - r_j^d \kappa \right] \alpha_i^d + \xi_j^d + \varepsilon_{ij}^d. \tag{13}$$

This model nests both models since under the discrimination model, $\kappa = 1$, while under the uniform pricing model, $\kappa = 0$. We therefore consider the test of H_0: $\kappa = 0$ versus H_1: $\kappa = 1$. Note that we wish to 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 (13), 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 $T < s$ and accept 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.

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 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. $x \mapsto (\widehat{\sigma}_1^2 - \widehat{\sigma}_0^2)x^2 + 2\sqrt{J}\widehat{\sigma}_1 x + (J - \widehat{\sigma}_0^2 T)$).

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 that Assumption 8 in Rivers and Vuong

(2002) is satisfied. Specifically, when both models are wrong, the estimated error terms of the two models depend in general on all the $(s_j^d)_{j=1,...,J}$ 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.

4 Extensions

4.1 Inference with unobserved groups and no proxy variables

We now extend the baseline model further by supposing that market shares $(s_j^d)_{d=1,...,n_D}$ are unobserved. We propose a methodology relying only on aggregate level market shares of products, as in the standard BLP model. As before, we observe only partially the prices paid by the different groups of consumers, and assume that Assumptions 1 and 2 hold.

This setting is relevant for combining aggregate data on sales with survey data where both consumers' characteristics and transaction prices are observed. In many cases, and in particular when J is large, the sample size is not large enough to construct accurate estimates of market shares. It is then preferable to rely on aggregate level data to estimate the demand. On the other hand, the survey data allow one to observe some transaction prices, so that Assumption 2 holds in this setting, with $\widetilde{p}_j = p_j^{d_j}$.

A canonical example is when the econometrician has data from consumer surveys (e.g. Kantar Worldpanel for data on grocery items from supermarkets) in markets with spatial price discrimination. The sample is never large enough to observe transaction prices for all the products and the geographical areas and to estimate precisely the market shares $(s_j^d)_{d=1,...,n_D}$ for all j. The common practice consists in aggregating demand at the national level and using the average price by item/supermarket brand. Instead, we suggest to rely on the same data but to apply our methodology to account for unobserved spatial price discrimination. We also refer to Miller and Osborne (2014), who estimate a different model from ours in the case of spatial price discrimination in the U.S. cement industry.

This methodology also constitutes an alternative strategy to model the airline industry, which shares common features with the previous example. Specifically, price dispersion for a given flight is important, and the prices of the alternative flights available at the moment of the purchase are not observable.[4] One may be reluctant to define ex ante consumer groups in this example, contrary to the one above where groups naturally correspond to geographical areas, but we can avoid this issue by simply using maximal prices or average prices for $\widetilde{p}_j$, instead of $p_j^{d_j}$.

---

[4]To deal with these issues Berry and Jia (2010) consider all tickets with different prices to be different products, while Ciliberto and Williams (2014) aggregate tickets at the route level and use the average prices.

Another example where our methodology below applies is when only data on total revenue and total quantities are available. To consider a very popular example in empirical industrial organization, supermarket scanner data typically report weekly revenues and units sold for all grocery items. The corresponding average prices then hide temporary promotions. Einav et al. (2010) document these discrepancies by comparing such prices obtained from retailers to those from a panel of consumers. In this set-up, the groups of consumers are defined by the day of the week when they make their purchase. The weekly sales and revenues for product j allow one to construct s_j and the sales-weighted average price $\widetilde{p}_j$ over the week. Because the demand is heterogeneous across the different days of the week, supermarkets may have an incentive to use temporary sales in order to price discriminate. Warner and Barsky (1995) analyze daily prices for a broad subset of consumer goods and find significantly lower prices during the weekend.

Another class of examples where we could use total revenues and quantities is the entertainment industry. Different prices are generally set for the same movie, concert or show, depending on some specified characteristics of the purchaser (age, professional activity, family size...). Considering for instance the demand for movies, available data typically consist in weekly revenues and the number of seats sold for a given movie (see, e.g. Einav, 2007; de Roos and McKenzie, 2014).

In all these cases where only the aggregate market shares s_j are observed, we can extend the approach developed in Section 3.2 under additional assumptions. As before, the demand is segmented into a finite number of groups of consumers. Preferences are heterogeneous across groups, but we now assume they 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 are the same across groups of consumers. This assumption was previously unnecessary because the $(s_j^d)_{d=1,...,n_D}$ were observed. It is also key for identification in Berry et al. (1995), as discussed in Berry and Haile (2014).

Assumption 3. $\Sigma^p = 0$, $\pi^p = 0$ 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 3, 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:

$$s_j(p, \xi, \theta) = \sum_{d=1}^{n_D} \phi^d s_j^d(p, \xi, \theta), \tag{14}$$

where $\phi^d = \Pr(D = d)$. ϕ^d can be assumed to be known or added to the vector of parameters θ. Now, for a given θ and vector of transaction prices p, the system of nonlinear equations in $(\xi_1, .., \xi_J)$ given by (14) 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 $\xi(p, s, \theta)$.

Now, we do not observe transaction prices, so we cannot compute directly $\xi(p, s, \theta)$ to form the moment conditions. As in the previous section, we then 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:

$$p_j^d = c_j - \frac{1}{\alpha^d (1 - \sum_{k \in \mathcal{J}_j} s_k^d(p, \xi, \theta))}, \tag{15}$$

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

$$p_j^d = \widetilde{p}_j + f_j \left(\frac{1}{\alpha^1 (1 - \sum_{k \in \mathcal{J}_j} s_k^1(p, \xi, \theta))}, ..., \frac{1}{\alpha^{n_D} (1 - \sum_{k \in \mathcal{J}_j} s_k^{n_D}(p, \xi, \theta))} \right)$$
$$- \frac{1}{\alpha^d (1 - \sum_{k \in \mathcal{J}_j} s_k^d(p, \xi.\theta))}. \tag{16}$$

Replacing ξ by $\xi(p, s, \theta)$ in this equation, it follows 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

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

For any $\xi \in \mathbb{R}^J$, let $p(\xi, \theta)$ denote the vector of equilibrium prices, provided that there is a unique solution of the system generated by Equations (14)-(15). 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, Theorem 2 below shows 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 $p(\xi, \theta)$. This conclusion holds when each firm sells only one product and when, roughly speaking, there is not too much heterogeneity between groups.[5] This condition is related to the one we impose in Theorem 1, namely that the heterogeneity on price sensitivity is small. Here, we impose that θ lies in a neighborhood of Θ_0, where $\Theta_0 = \{(\alpha_0, ..., \alpha_0, \beta_0, ..., \beta_0), \alpha_0 \in A, \beta_0 \in B\}$, A and B being compact sets included in $(0, \infty)$ and $\mathbb{R}^k$ respectively. We also impose that $\xi \in K = [\underline{\xi}, \overline{\xi}]^J$, with $\underline{\xi} > -\infty$ and

$$\overline{\xi} \leq -\ln(2) - \inf_{(x,\tilde{c},\alpha,\beta) \in \mathrm{Supp}(X,c) \times A \times B} [x'\beta + \tilde{c}\alpha].$$

We show in the proof of Theorem 2 that this last restriction implies that market shares are always smaller than $1/3$, a restriction that is assumed by Aksoy-Pierson et al. (2013) when studying the related question of the uniqueness of price equilibria.

Theorem 2. *Suppose that Assumptions 1-3 hold, f_j is 1-Lipschitz for all j and firms sell only one product. Then there exists a neighborhood Θ_1 of Θ_0 such that for all $(\theta, \xi) \in \Theta_1 \times K$, 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 $p(\xi, \theta)$, the unique solution of Equation (15).*

Note that 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. We provide details on the estimation algorithm and the results of Monte Carlo simulations in Section 1.3 of the supplementary material.

In principle, the model above can be extended to the case of a continuum of consumer groups. Groups are then indexed by (α, β) and we impose that the distribution of (α, β) is parametric (so that θ remains finite-dimensional), with a compact support on α. Then the idea is to replace vectors by functions. Instead of considering $(p_j^1, ..., p_j^{n_D})$, we have $(\alpha, \beta) \mapsto p_j(\alpha, \beta)$, where $p_j(\alpha, \beta)$ denotes the price paid by consumers with parameters (α, β) for product j. Similarly, M_θ now takes functions as arguments, rather than vectors. Theorem 2 may then extend to this context, but computing ξ may be challenging. Instead, we suggest to use a discrete approximation of the distribution of heterogeneity, in the spirit of quadratures.

Finally, Theorem 2 does not directly apply to the case where observed prices are sales-weighted average prices, $\widetilde{p}_j = \sum_d (s_j^d / s_j) p_j^d$. This is because the s_j^d are unobserved here, so f_j is unknown by the econometrician. However, we can still apply the algorithm developed above. In this

---

[5]The case of multiproduct firms is left for future research. It may be dealt with the approach proposed by Nocke and Schutz (2016).

context,

$$M_{s,\theta,j}^d(p) = \widetilde{p}_j - \frac{1}{\alpha^d(1 - \sum_{k\in\mathcal{J}_j} s_k^d(p, \xi(p,s,\theta),\theta))} + \frac{1}{s_j}\sum_{d'=1}^{n_D} \frac{s_j^d(p, \xi(p,s,\theta))}{\alpha^{d'}(1 - \sum_{k\in\mathcal{J}_j} s_k^{d'}(p, \xi(p,s,\theta),\theta))},$$

which is known by the econometrician. Hence, we can define the sequence $(p_n)_{n\in\mathbb{N}}$ as above. Even though the proof of Theorem 2 does not easily extend to this case, simulations suggest that a similar result should hold in this case. For a similar DGP as the one presented in Section 5.2 below, we consider 100 values of θ and for each, 50 starting points p_0. For all the values of θ, the algorithm converged to the same vector of prices. We refer to Section 1.3 of the supplementary material for more details.

4.2 Discrimination based on unobserved individual characteristics, with proxy variables

We now consider an alternative to the previous extension, still in the case where the market shares s_j^d are unknown. Specifically, we show that It is possible to apply the methodology in Section 3.2 as long as a proxy for the variable D used by the seller to price discriminate is available. Let us suppose that we observe a discrete variable $\widetilde{D}$ such that (i) $(\zeta_i, \varepsilon_{ij}^d) \perp\!\!\!\perp \widetilde{D}$ and (ii) the matrix $\mathbf{P}$, which typical $(d,\widetilde{d})$ term is $P(D = d|\widetilde{D} = \widetilde{d})$, has rank n_D. Condition (i) is an exclusion restriction which imposes that consumers do not differ systematically in their taste across categories of $\widetilde{D}$, once we control for D. Condition (ii) is similar to the standard relevance condition in instrumental variable models and imposes that the proxy variable $\widetilde{D}$ is, basically, related to D. Let Y_i denote the product choice of consumer i. Under the first condition, we have

$$P(Y_i = j|\widetilde{D}_i = \widetilde{d}) = \sum_{d=1}^{n_D} P\left(D_i = d|\widetilde{D}_i = \widetilde{d}\right) P(Y_i = j|D_i = d, \widetilde{D}_i = d)$$
$$= \sum_{d=1}^{n_D} P\left(D_i = d|\widetilde{D}_i = \widetilde{d}\right) s_j^d.$$

Then, letting $\mathbf{s}_j = (s_j^1, ..., s_j^{n_D})'$, $\widetilde{\mathbf{s}}_j = (P(Y_i = j|\widetilde{D}_i = 1), ..., P(Y_i = j|\widetilde{D}_i = n_{\widetilde{D}}))'$, we have, for all $j = 1...J$,

$$\widetilde{\mathbf{s}}_j = \mathbf{P}\mathbf{s}_j$$

Because $\mathbf{P}$ has rank n_D, this equation in $\mathbf{s}_j$ admits a unique solution. This implies that $\mathbf{s}_j$ is identified. We can then apply the methodology above, using these market shares.

As an example of this proxy variable approach, consider a scenario where the econometrician observes the buyers' professions while sellers price discriminate based on buyers' income. In this context, we observe market shares of products by professional activity. The rank condition means that we know the probability of belonging to an income class conditional on the

professional activity. From this probability matrix, we are able to compute market shares of products by income class. The exclusion restriction imposes that the differences in preferences across professional activities only reflect the differences across income classes.

4.3 Combination of micro and macro data

In addition to market shares and product characteristics, we may observe, through survey data, additional information on purchasers. Berry et al. (2004) explain how such data can be used to improve the estimators, by including additional moments based on these data in the GMM program. This idea extends naturally to our context.

An interesting special case is when transaction prices are observed. Note that observing transaction prices is not sufficient to apply the usual BLP model, because we still do not observe counterfactual prices, i.e. the prices of the products that the consumers did not purchase. On the other hand, transaction prices can be helpful in our model for at least two purposes. First, they can help defining the demographic group variable D. Our model implies that the transaction prices of product j are identical among each group d. Hence, some candidates for D can be rejected on this ground. Similarly, in our extension to unobserved groups, our model implies that there are no more than n_D different transaction prices for each product. Transaction prices can therefore be useful to provide a lower bound on n_D. Second, observed transaction prices can be used to construct moment conditions aiming at improving estimation, in the same spirit as Berry et al. (2004). We may consider for instance covariances between transaction prices and the characteristics of the products or the purchasers. The idea is then to match the model-based covariances with their empirical counterpart. Another possibility is to use additional information on the overall distribution of transaction prices or discounts. Additional conditions would then take the form of differences between model-based moments of the price distribution and their empirical counterparts.

4.4 Allowing for cost differences

Our methodology relies on Assumption 1 which supposes constant marginal costs across groups of consumers. This assumption might not be valid in some settings. Specifically, when the demand segmentation is based on geographic variables, the costs can vary across groups because of local prices of production factors. In some cases, the cost of transporting the products from the factory to the retailers can be significant and vary within the territory. We can relax this 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 (9) 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 between groups, constant across j, by a marginal cost difference or a corresponding difference in the intercepts of the utility function.

4.5   Other functional forms on price effects

We have implicitly assumed up to now, following the common practice, that indirect utilities depend linearly on disposable income, namely on $\alpha_i(y_i - p_j)$, where y_i denotes the income before making one's choice. $\alpha_i y_i$ can then be removed, as it is constant across alternatives. To incorporate, for example, credit constraints as in BLP, the indirect utility may rather depend on $\alpha_i \ln(y_i - p_j)$. With such a specification, consumers cannot choose to buy products with prices above their annual income. Let us suppose, more generally, that the utility depends on disposable income through $q(y_i - p_j, \alpha_i)$ where q is known by the econometrician while $\alpha_i | D_i = d \sim \mathcal{N}(\alpha^d, \sigma_\alpha^{2d})$ with $(\alpha^d, \sigma_\alpha^{2d})$ unknown. Our methodology also applies to this setting. In such a case, one has to include entirely $q(y_i - p_j, \alpha_i)$ into $\mu_j^d(E_i, \zeta_i, p_j^d)$, with y_i as one component of E_i. Then Equations (9) and (11) remain unchanged. The only difference is that the terms entering into Ω^d are now different from Equation (10). But other than that, the construction of the moment conditions follows exactly the same methodology.

5   Estimation algorithm and Monte Carlo simulations

5.1   Implementation

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 (9) and (11) 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 (11). Here, we recommend to follow Lee and Seo (2016), who suggest a Newton's method improving upon the contraction mapping of BLP.

3. Given δ_n^d and p_n^d, compute the corresponding matrix Ω^d and update the transaction prices, using Equation (9).

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 for 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. Finally, in the absence of random coefficient on prices, the *delta-loop* does not involve transaction prices. We can then compute the mark-up terms using the correct δ^d, and the *price-loop* converges immediately to the true vector of prices.

We also use the following specifications for computing the GMM estimator. First, to approximate the aggregate market shares, we use in the simulations below 300 symmetric normal draws for each demographic group and market. We rely on Knitro derivative-based algorithm for minimization. Our initial values for the price sensitivity parameters are the estimates obtained with the simple logit model, while we use random draws from a uniform distribution $\mathcal{U}[-1/2, 1/2]$ for the value of the random coefficient σ^p. As suggested by Dubé et al. (2012) and Knittel and Metaxoglou (2014), we set a tight tolerance (10^{-12}) to solve numerically for the mean utilities and prices, while the tolerance levels are 10^{-6} for the parameters and 10^{-4} for the objective function. In the application where J is much larger, we use 1,000 Halton draws rather than 300 and rely on a tolerance of 10^{-6} for both the parameters and the objective function. Finally, we follow Nevo (2000) by setting the value of the objective function to a high value when the parameters imply non-defined values for δ or p.

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 $2Jn_D$.

5.2 Data Generating Process

To investigate the performance of our estimator and whether the algorithm produces reliable results, we perform Monte Carlo simulations. We construct 200 different datasets for T = 25 markets, J = 24 products and $n_D = 4$ demographic groups. For each market and product, we construct the vectors of observed characteristics $X_{jt} = (1, X_{1jt})$, unobserved characteristics ξ_{jt}^d, observed cost shifters $W_{jt} = (W_{1jt}, W_{2jt}, W_{3jt})$ and unobserved cost shifters ω_{jt}. The marginal cost of product j in market t then satisfies

$$c_{jt} = 0.7 + 0.7X_{1jt} + W_{1jt} + W_{2jt} + W_{3jt} + \omega_{jt}. \tag{17}$$

We suppose that $X_{1jt}, W_{1jt}, W_{2jt}, W_{3jt}, \omega_{jt}$ and ξ_{jt}^d are mutually independent. $X_{1jt} \sim \mathcal{U}[1, 2]$, $W_{kjt} \sim \mathcal{U}[0, 1](k = 1, 2, 3)$ while ξ_{jt}^d and ω_{jt} are two normal variables $\mathcal{N}(0, 0.1)$. The parameters of preferences are summarized in Table 1. Groups of consumers are heterogeneous in their price sensitivity. Group 1 is the less price sensitive group, but does not have the highest utility of holding a car nor the highest valuation for the exogenous characteristics (the valuation is set to 1.5, versus 2 for the three other groups). As in our application, the unobserved heterogeneity parameter σ^p is identical for the four demographic groups. Finally, we assume that the market includes 4 firms, each of them producing 6 products. Once we solve for prices and market shares $(s_{jt}^d, p_{jt}^d)_{d=1,2,3,4}$, we define for each product the posted price $\widetilde{p}_{jt}$ as the maximal price across demographic groups. We use product characteristics, cost shifters and functions of other product characteristics as instruments for the estimation.

	Proportion	Intercept	X_1	Price
Group 1	0.3	-1	1.5	-1.5
Group 2	0.3	-0.5	2	-2
Group 3	0.2	-0.1	2	-2.5
Group 4	0.2	-0.5	2	-3
σ^p				0.4

Table 1: Parameters of preferences in the Monte Carlo. simulations

5.3 Numerical Aspects

We first investigate the convergence of our algorithm for one synthetic dataset generated using the DGP described before. For that purpose, we compute the Lipschitz coefficient of the function g_θ for θ at its true value, except σ^p that we make vary in $\{0, 0.1, ..., 1\}$. As discussed above, the value of the Lipschitz coefficient is 0 when σ^p is set to 0. Then, for all values equal or below 0.9, we obtain Lipschitz coefficients that are increasing but remain

lower than 0.798. Conversely, when $\sigma^p = 1$ the Lipschitz coefficient becomes greater than 1, indicating that the algorithm may have problems to converge. This is consistent with Theorem 1: for σ^p close enough to zero, the algorithm is k-Lipschitz, with $k < 1$, but it may not be when σ^p is large. Note however that the system of equations may still be invertible in (δ, p). Given the values of the price sensitivities, a value of 1 for σ^p means that for some groups, more than 6% of the population have a positive price parameter, which implies that optimal prices may go to infinity.

We then evaluate the performances of our algorithm at the true parameter θ_0, by starting from 50 different initial values of prices equal to $R \times \widetilde{p}_j$, where $R \sim \mathcal{U}[0.25, 1]$. As expected, the algorithm always converges to the true value of the transaction prices. Besides, convergence occurs very quickly, in 13 or 14 iterations. We compute, at each iteration of the *price-loop*, the average and maximal absolute differences between the true prices and those obtained by the algorithm, across all products. We then take the average of these mean and maximal absolute differences over the 50 initial draws. The results, displayed in Table 2, show that the sequence of vectors of prices converges very quickly to the true vector.

Iteration	1	2	3	4	5	6	7
Mean	0.7313	0.0144	0.0004	1.38×10^{-5}	5.18×10^{-7}	2.65×10^{-8}	1.77×10^{-9}
Maximum	2.7729	0.0631	0.0029	1.9×10^{-4}	1.62×10^{-5}	1.48×10^{-6}	1.41×10^{-7}

Reading notes: "mean" (resp. "maximum") is the mean (resp. maximal) absolute difference between the true prices and those obtained by the algorithm across all products. The figures are averages over the 50 simulations. The average true price here is 3.91, with a range of [2.24; 5.99].

Table 2: Mean and maximum price difference across iterations.

We further check that the algorithm converges for values of the parameters different from θ_0, starting this time from the same initial price vectors, $p^d = \widetilde{p}$. We draw 50 different vectors of parameters from $\mathcal{U}[\theta_0/2, 3\theta_0/2)$ and investigate potential convergence issues. On the 50 different values of θ, 7 values do not lead to convergence as they imply a failure in the *price-loop*, with some prices tending to infinity. These cases of convergence failure do not cause any trouble for the estimation since the objective function is set to a high value whenever the price-loop does not converge. In practice, our algorithm always came back to regions of the parameter space where the *price-loop* converges. Over the 43 draws of θ for which convergence occurs, the *price-loop* converges in 28 iterations on average, with a minimum of 8 iterations and a maximum of 312 iterations. Our algorithm converges in 2.3 seconds on average.

Finally, following Knittel and Metaxoglou (2014), we carefully check the sensitivity of the estimation method to the initial values and the minimization algorithm. Regarding the effect of initial values, we draw 50 different initial values of $\theta = (\sigma^p, (\alpha^d)_{d=1,...,4})$. Specifically, σ^p is drawn from a $\mathcal{U}[0, 1/2]$ and α^d is drawn from $\mathcal{U}[\alpha^d/2, 3\alpha^d/2]$. The estimation always converged to the same parameter values in these 50 cases. This indicates that even if the

algorithm fails to converge for some values of the parameters, as indicated above, the global minimization algorithm does not display any problem of convergence. Roughly speaking, during the optimization, the parameter values for which the *price-loop* does not converge are discarded since they are associated to high values of the objective function. Regarding the choice of the minimization algorithm, we try a derivative-free minimization algorithm (namely the Neader-Mealde simplex) instead of Knitro. This algorithm does not perform as well as Knitro both in terms of convergence and in terms of the time spent in optimization. Specifically, we estimated the model using 50 different initial values and find that in 26% of the simulations, the objective function at the minimum obtained with the simplex is more than 1% higher than when using Knitro. Moreover, in the 74% remaining cases, the objective function with the simplex is never lower than with Knitro. Finally, the simplex is on average 7% slower than Knitro.

5.4 Simulation results

The Monte Carlo simulation results are displayed in Table 3. As in the application, we estimate separately the demand and supply parameters, using moment conditions from the demand- and supply-side, respectively. The estimation algorithm converges for every replication, and the GMM accurately estimates both demand and supply parameters. The pivot groups are exactly guessed and the estimated discounts are very close to the true underlying discounts. To get a sense of the computational burden of our estimation method, we also estimate the standard BLP model on the same simulated data. The standard BLP estimator is around 70 times quicker than the discriminatory model, with an average number of iterations 6 times smaller. This is partly due to the fact that for the standard BLP we optimize over σ^p only, while for our model, we optimize over $(\alpha^1, ..., \alpha^4, \sigma^p)$. Nevertheless, the computation time of our estimator remains decent because (i) it is possible to parallelize the computationally intensive part of the estimation algorithm and (ii) the contraction based on Newton's method converges more quickly than BLP's contraction.

	True	Estimators	
		Mean	Std. dev.
Price sensitivity			
Group 1	-1.5	-1.5	0.032
Group 2	-2	-2	0.04
Group 3	-2.5	-2.5	0.057
Group 4	-3	-3	0.067
sigma (σ^p)	0.4	0.4	0.021
Intercept			
Group 1	-1	-1	0.083
Group 2	-0.5	-0.51	0.098
Group 3	-1	-1.01	0.118
Group 4	-0.5	-0.51	0.139
Exogenous characteristic			
Group 1	1.5	1.5	0.016
Group 2	2	2	0.015
Group 3	2	2	0.017
Group 4	2	2	0.016
Marginal cost equation			
Intercept	0.7	0.7	0.03
X_1	0.7	0.7	0.015
W_1	1	1	0.017
W_2	1	1	0.017
W_3	1	1	0.016
Average discount (in %)			
Group 1	0	0	0
Group 2	7.28	7.26	0.26
Group 3	11.13	11.12	0.251
Group 4	13.69	13.66	0.236
% pivot well predicted			100
% simulations converging			100
Average number of iterations			30
Time (sec)			214
Test			
Average $\widehat{\kappa}$			1
Standard deviation of $\widehat{\kappa}$			0.16
% accept discrimination model			99.5

Reading notes: the results were obtained over 200 simulations. "Time" is the optimization time in seconds using our preferred starting point and on our desktop computer using 6 parallel workers (Intel® Core™, 6-Core Xeon E5, 3.5 GHz, 16Gb RAM). "% simulations converging" is the percentage of simulations for which our algorithm converged.

Table 3: Simulation results.

We also implement the test of the model and accept the discrimination model 99.5% of the time. Finally, we also investigated the performance of the test under uniform pricing. We ran 200 Monte Carlo simulations generated using the same DGP as above except that firms set uniform prices. Under this model, the right uniform pricing model was accepted 99.5% of the time.

6 Application to the French new car market

6.1 Data and methodology

We apply our methodology to estimate demand and supply together with unobserved discounts in the new automobile industry, using a dataset from the association of French automobile manufacturers (CCFA, Comité des Constructeurs Français d'Automobiles) that records all the registrations of new cars purchased by households in France between 2003 and 2008. Each year, we observe about one million vehicles registered and their main attributes: brand, model, fuel energy, car-body style, number of doors, horsepower, CO_2 emissions, cylinder capacity and weight. These characteristics have been complemented with fuel prices to compute the cost of driving (in euros for 100 kilometres).

Automobile sellers are well known to price discriminate, negotiate or to offer discounts over the sticker price to close the deal. But as in our theoretical model, we only observe here posted prices that come from manufacturers catalogues. As often, we do not observe the automobile options such as air conditioning, audio systems or metallic paint that are chosen by the purchasers. If the cost of the options is included in the marginal cost of the cars, our assumption that the marginal costs are constant across demographic groups could be violated. Rich purchasers may indeed purchase more of these options, for instance. Options choice can however be considered independent of car choices, as long as the same options are available to all products. We can then safely ignore option costs and option choices in our analysis.

We now turn to the construction of the consumer groups that are used by firms to price discriminate. Apart from car attributes, the date of the registration and some characteristics of the owner are provided in the CCFA database: municipality of residence and age. The age (or the age class) is presumably a strong determinant of purchase, and is easily observed by a seller even if he does not know the buyer before the transaction. We therefore assume that these characteristics are used by the automobile sellers to price discriminate. The income is also likely to affect preferences for different car attributes and price sensitivity. The income is, however, likely to be unobserved by the seller but instead inferred from the municipality the buyer lives in and the age class. We compute a predictor of buyer's income, namely the median household income in his age class and in his municipality using data from the French national institute of statistics (Insee).[6]

It seems reasonable to assume that the seller does not have a far better prediction of the buyer's income in such anonymous market, where buyers and sellers do not know each other before the transaction. It is crucial for our approach that buyers cannot lie about their individual characteristics, which implies in our application that buyers do not make geographical

---

[6]There are over 36,000 municipalities in France. Note also that Paris, Lyon and Marseille, the three largest cities, are split into smaller units ("arrondissement"). The heterogeneity in the median income across municipalities is therefore quite large.

arbitrage, i.e. buy the car in another municipality where discounts are higher. We believe that this assumption is reasonable since buyers have incentives to buy a new car at a close dealer to minimize transportation costs and take advantage of the after-sale services and guarantees. We thus 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 as the threshold for income. This income threshold corresponds roughly to the median yearly income in France in 2008. Note that we do not observe the owner's gender in our database. Even if this information was available, it would be hard to use it since the owner and the buyer can be different persons. Furthermore, many couples are likely to buy their car together. Nevertheless, we check in Section 6.5 below the sensitivity of our estimates to price discrimination with respect to gender.

As usual, when defining the groups of consumers, we face a trade-off between realism (it is likely that firms discriminate along several dimensions) and accuracy of the observed proportion of sales $\widehat{s}_j^d$ as estimators of the true market shares s_j^d. The six groups that we consider are large enough to avoid in most cases the problem of too many zero sales (see Table 4 in Section 2.1 of the supplementary material for the fraction of products with null market shares). Moreover, rather than discarding those products, we replace the proportion of sales by a predictor of s_j^d that minimizes the asymptotic bias, namely $\widehat{s}_j^d = \frac{n_j^d + 0.5}{N^d}$, n_j^d denoting the number of sales of product j in group d and N^d the number of potential buyers with characteristics d. Note that another simple correction of the basic market shares estimator has been proposed by Gandhi et al. (2013). We provide more details about our correction and show in Section 2.1 of our supplementary material that our results are robust to the choice of a correction for the market shares.

We define a product as a brand, model, segment, car-body style and fuel type. This results in a total of 3,205 products for the six years. Following BLP, we assume that each of these years corresponds to a different market. Table 4 presents the proportion of each consumer group in the population and the average characteristics of new cars purchased corresponding. We find significant heterogeneity across these groups. On average, the medium age, high income class purchases more expensive vehicles. They also choose larger and more powerful cars. Young purchasers are more interested in smaller cars (lighter and with three doors) whereas station wagons are more popular among the medium age class. The highest age group purchases lighter vehicles than medium age classes, but these vehicles are on average less fuel efficient.

Group	Freq.	Price	Fuel cost	HP	Weight	Three doors	Station wagon
A < 40, I < 27,000	15.7%	19,803	6.2	5.7	1182	19.0%	9.7%
A < 40, I ≥ 27,000	11.5%	20,911	6.5	6	1221	16.8%	12.9%
A ∈ [40,59], I < 27,000	16.3%	21,521	6.5	6.1	1231	14.3%	12.7%
A ∈ [40,59], I ≥ 27,000	22.3%	21,739	6.8	6.2	1236	14.8%	13.1%
A ≥ 60, I < 27,000	20.8%	20,117	6.9	5.9	1194	11.4%	8.9%
A ≥ 60, I ≥ 27,000	13.2%	20,831	7	6	1219	10.9%	10.5%

Reading notes: "A" represents the age and "I" the income. Prices are in constant (2008) euros, fuel cost is the cost of driving 100 kilometers, in constant (2008) euros, "HP" stands for horsepower, weight is in kilograms.

Table 4: Average characteristics of new cars purchased across groups of consumers.

The dataset does not contain any information on the distribution network, and thus the distribution part is not modeled in this application. We make the traditional assumption that manufacturers have only exclusive dealers and are perfectly integrated. As detailed in Nurski and Verboven (2016), exclusive dealing is still prevalent in most European countries, with 70% of car dealers being exclusive to one brand in Europe. As discussed in Section 3.3 above, adding vertical relations between manufacturers and dealers would be possible, provided that dealers compete à la Bertrand in the downstream market.

Hereafter, we estimate the random coefficient model with uniform pricing and with unobserved price discrimination. We also estimate nested logit models as robustness check. In all specifications, we control for the main characteristics of the cars such as horsepower, weight and the cost of driving 100 kilometers in the demand function. We also introduce dummies for station wagon body-style and three doors. Finally, we introduce year and brand dummies that are constrained to be identical for all demographic groups. For the two random coefficient models, we allow for unobserved heterogeneity of preferences inside groups of consumers in terms of price sensitivity.[7] Finally, to obtain more accurate results, we constrain the heterogeneity parameter to be identical for all demographic groups.

We estimate both models relying only on the moment conditions stemming from the demand.[8] The implementation of the estimation follows the method described in Section 5. In particular, we check that we reach the minimum by trying multiple initial values for the parameters and several minimization algorithms. We also verify that the system of equations defined by (9) and (11) has a unique solution at the estimated value of parameters by applying the tests performed in the simulation analysis and find that after drawing several initial values of transaction prices and mean utilities, the algorithm always converges to the same value of

---

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

[8] We also estimated the models using the moment conditions from the supply side. We supposed that the marginal cost depends on horsepower, fuel consumption (in liters for 100 kilometers) and car weight multiplied by a composite price index that aims at approximating the average input price. We also introduce brand dummies to control for manufacturer's specific unobserved quality of cars. We considered both linear and log-linear specifications for the marginal cost equation. Results on the demand side were very similar.

estimated transaction prices. We also compute the values of the Lipschitz coefficient for the price contraction, starting from the posted prices until the convergence to transaction prices. The *price-loop* converges in 8 iterations and the Lipschitz coefficient is always below 0.043.

In addition to exogenous characteristics we include the following instruments. The first is the car weight multiplied by a composite price index that aims at approximating the average input price. Specifically, we use a weighted average of steel, aluminum 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 continuous exogenous characteristics (namely weight, horsepower and fuel cost) of other brands' products. We also consider the sums of these characteristics over other brands' products of the same segment, supposed to be closer substitutes. Finally, we include the sums of these characteristics of the other products of the brand belonging to the same segment. Armstrong (2016) has recently 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 indeed relevant for prices. We use for that purpose the F-statistic of the joint nullity of the coefficients of these instruments in the linear regression of prices on the characteristics and these instruments. We obtain $F \simeq 24.1$, 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.

6.2   Parameter estimates and comparison with the standard model

The results for the models with uniform pricing and unobserved price discrimination are presented in Table 5. Results for the nested logit specification are displayed in Section 2.3.1 of the supplementary material. They are globally in line with the results we obtain for the models with random coefficients.

	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 ≥ 60, I $=$ L	-3.85**	0.149	-4.21**	0.133
Age ≥ 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 ≥ 60, I $=$ L	-5.73**	0.228	-6.48**	0.226
Age ≥ 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 ≥ 60, I $=$ L	1.11**	0.202	1.12**	0.178
Age ≥ 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 ≥ 60, I $=$ L	-3.51**	0.171	-3.46**	0.17
Age ≥ 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 ≥ 60, I $=$ L	4.55**	0.23	5.59**	0.243
Age ≥ 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 ≥ 60, I $=$ L	-0.52**	0.194	-0.53**	0.205
Age ≥ 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 ≥ 60, I $=$ L	-0.73**	0.128	-0.76**	0.14
Age ≥ 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 are constrained to be identical across groups of consumers.

Table 5: Parameter estimates for the standard uniform BLP model and our model with unobserved price discrimination.

The two random coefficients 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. The price sensitivity decreases with both age and income, leaving the young with low income the more price sensitive group.

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 valuated, reflecting that most of the consumers buy sedan or hatchback cars with five doors (four doors plus the trunk).

If qualitatively similar, the results we obtain with the two models exhibit some quantitative differences, as also illustrated below. It is therefore important to test between the two models. Recall that in the test developed in Section 3.4, $\kappa = 0$ 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 $p_1 = 0.52$. 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. We provide the same analysis for the nested logit models in Section 2.3.1 of the supplementary material. As Table 6 shows, price elasticities are, in absolute terms, lower for the model with uniform pricing for all groups, except for the pivot group. Hencen 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 1 of 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 2.2 of the supplementary material for more details.

Group of	Price elasticity		Average mark-up		Average surplus	
consumers	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	17916	21651

Reading notes: Mark-ups are in percentage.

Table 6: 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 hypothetical counterfactual simulation 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 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 display in Table 7. 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).

| | 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 $CS^d = \int \log \left(1 + \sum_{j=1}^{J} \exp(\delta_j^d + \mu_j(u, p_j^d)) \right) dF_\zeta(u)$. We thus take into account the substitution from the outside good.T he welfare is computed as the sum of manufacturers' profits and all consumers' surplus, minus the cost of the policy.

Table 7: Welfare analysis of the hypothetical purchase subsidy for young households and a hypothetical merger between PSA and Opel.

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).[9] Results displayed in the last two columns of Table 7 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.

6.3 Analysis of the discounts and the effect of price discrimination

Table 8 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 the one on price elasticity. The estimated pivot group (the group assumed to be paying the posted price) is identical for all the products 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 purchasers with a low income 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 age consumers get smaller discounts (11% for the low income group and 9.6% for the high income group). Finally, while 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. Note that these figures average vehicle specific discounts. Our methodology allows us to analyze further the heterogeneity across car models, since we estimate a discount value for each model and demographic group. Further results are displayed in Section 2.2 of the supplement.

---

[9]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.

	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 ≥ 60, I $= $ L	10.37	10.28	2,084	2,174
Age ≥ 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 8: 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 such a discrimination, 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%.[10] 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 appears to be a determinant of price dispersion nearly as important as second-degree price discrimination.

If third degree price discrimination is always profitable for a monopoly seller, this may not be the case in an oligopoly, because price discrimination may reinforce competition among firms. Under certain conditions, all firms may actually be worse off than if they could commit to a uniform pricing strategy (Holmes, 1989; Corts, 1998). The effect on consumers is also ambiguous since for some groups of consumers, some products may turn out to be cheaper without price discrimination. We investigate here the effect of price discrimination on firms and consumers by simulating, using our estimates of the model with price discrimination, the counterfactual prices and profits that would occur in equilibrium if firms could commit to set a single price for all the consumer groups.

---

[10]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.

Manufacturer	Profits with price discrimination	Profits without price discrimination	Gains from discrimination
PSA	1036.42	998.76	3.77%
RENAULT	691.42	665.92	3.83%
VOLKSWAGEN	340.13	341.11	-0.29%
FORD	170.24	165.27	3.01%
TOYOTA	162.78	158.3	2.83%
DAIMLER	156.53	144.76	8.13%
B.M.W.	137.31	132.87	3.34%
GM	113.08	112.25	0.74%
FIAT	86.27	85.3	1.14%
SUZUKI	52	51.77	0.43%
NISSAN	48.77	47.86	1.91%
HYUNDAI	46.38	45.68	1.53%
HONDA	29.62	28.62	3.48%
MAZDA	18.94	18.8	0.76%
CHRYSLER	16.1	15.94	0.99%
MITSUBISHI	9.89	9.67	2.28%
PORSCHE	8.82	9.32	-5.35%
SUBARU	1.85	1.88	-1.5%
SSANGYONG	1.79	1.83	-1.91%
ROVER	0.05	0.05	1.9%
Total industry	3128.39	3035.95	3.05%

Reading notes: Profits are annual profits, for the year 2007, in millions of euros. The gains from price discrimination represent the profits gains or losses of switching from the uniform pricing equilibrium to the price discrimination equilibrium.

Table 9: Gains and losses from price discrimination by brand.

Results on firms' profits are displayed in Table 9. Gains from price discrimination are rather small but heterogeneous. We observe that if price discrimination is profitable for most of the manufacturers, it makes 4 out of the 20 manufacturers worse off. The gains associated to price discrimination are particularly high for brands that commercialize powerful vehicles, such as Daimler group (that sells Mercedes, Dodge and Smart), with an increase of its profits by 8.1%. This makes sense, given that higher prices and horsepowers are associated to higher discounts or, put it another way, more price discrimination. Price discrimination appears to be also more profitable than average for the two French manufacturers (+3.77%, +3.83% for respectively PSA group (Peugeot and Citroen) and Renault group (Renault and Dacia). Conversely, Porsche is the manufacturer that is the most hurt by price discrimination since its profits are reduced by 5.4%. The total gains from price discrimination are rather small but significant, the industry profits increasing by 3.05% with price discrimination.

We also investigate the impact of price discrimination on consumers. In Table 10, we compute the average price differences between the uniform and the discriminatory prices for each group of consumers and report the number of products for which the discriminatory price is lower than the uniform one (see Column 2). We also compute average surplus for each group of consumers under the two pricing equilibria (see Columns 5-7). For the young groups, all products are more expensive under uniform pricing, and price discrimination makes them save around 700 euros. The situation is more contrasted for the 40-59 and 60+ groups. In

particular, all prices are lower under uniform pricing for the 60+ group, in which consumers would save on average the substantial amount of 1,900 euros. Overall, price discrimination is hardly beneficial for consumers as it increases the global average individual surplus by only 0.31%. Again, this average impact hides heterogeneous effects. The group experiencing the highest welfare gain is the group of young consumers with low income (+3.9%), while the pivot group is, not surprisingly, the one that suffers the most from price discrimination (-2.8%).

| Group of consumers | $\#\{j : p_j^d < p_j^{\text{uniform}}\}$ | Average gain in purchases | | Average surplus | | Δ surplus |
		S-weights	B-weights	Disc.	Unif.	(in %)
Age < 40, I = L	571	679	825	12,760	12,279	3.92
Age < 40, I = H	569	501	593	14,439	14,098	2.42
Age ∈ [40,59], I = L	553	350	371	14,714	14,477	1.64
Age ∈ [40,59], I = H	292	54	63	18,650	18,599	0.28
Age ≥ 60, I = L	483	240	190	15,367	15,200	1.1
Age ≥ 60, I = H	0	-1,857	-1,912	33,753	34,734	-2.82
Average	412	41	63	18,140	18,085	0.31

Reading notes: the second column indicates how many products (among the 571) have lower prices with the price discrimination regime. "S-weights" denotes the sales-weighted average while "B-weights" are those obtained by using the same artificial basket of cars for all groups. Average surplus are in euros. The last column measures the variation of average consumers' surplus due to price discrimination.

Table 10: Gains of price discrimination for groups of consumers.

For the total welfare, computed by simply summing manufacturers profits and consumers surplus, we find a net benefit of price discrimination of 301 million euros. Consumers gain 209 million euros, while the manufacturers make extra profits of 92 million euros.

6.4 Plausibility of the results

6.4.1 Comparison with French transaction prices

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.[11]

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

---

[11]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.

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

$$E(p^{BdF}|\widetilde{X}) = E(p^{CCFA}|\widetilde{X}). \tag{18}$$

To test the equality described by (18), we first compute $p(\widetilde{X}) \equiv E(p^{CCFA}|\widetilde{X})$ using the CCFA data and our estimates of the transaction prices. Then (18) 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.

Equality (18) 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 are displayed in Table 11. 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 (18) 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 while 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.

	$p^{CCFA} =$ 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 11: Comparison with BdF data.

6.4.2 Other evidence

Another confirmation of our results comes from a survey conducted by the French credit company, *Cetelem* (L'Observatoire Cetelem, 2013). First, it reveals that in 2012, 87% of the purchasers benefited from a discount from their car dealers, which is exactly what we estimate with our model (86.8%). Interestingly, a quarter of them also indicate that they did not even need to negotiate to obtain a rebate, which may be seen as evidence of price discrimination rather than a true bargaining process. Furthermore, for 68% of individuals who indicated that they negotiated the car price, the average discount was around 11%. This result is comparable to our average on the whole population, and also very close to the average discount we obtain on individuals below 60 years old (11.3%), who also represent around two third of the whole population. We were unable to find precise statistics on the dispersion of discounts, but we can report some anecdotal evidence. For example, when searching online using the keywords "*how much discount for new car*" (in French), the first website listed states that "discounts are generally between 5% and 20%".[12] The fourth website associated to the same key words search is a forum asking the question of how much discount one can expect to obtain on the purchase of a new car. One reply states that discounts do not exceed 20%, while another mentions an average discount of 6%.[13] Our estimations are overall consistent with these figures.

A recent study by Kaul et al. (2016) investigates the effect of the scrapping policy on the magnitude of discounts in Germany, using data collected from a sample of dealers. The study first reveals that some consumers do not obtain any discounts (see their Table 2 with summary statistics on discounts). When excluding demonstration cars and sales to employees, which are typically much more discounted, they obtain an average discount of 14%. This magnitude is broadly consistent with our estimate, though somewhat higher. Their study focuses on

---

[12]See `http://www.choisir-sa-voiture.com/concessionnaire/meilleur-prix-voiture.php`. We performed this search in November 2014 using Google search engine.

[13]See `http://forum.hardware.fr/hfr/Discussions/Auto-Moto/negocier-voiture-concession-sujet_15899_1.htm`.

the period 2007-2010, which corresponds to the beginning of the economic crisis. If posted prices did not adjust immediately, it is likely that car dealers reacted to this adverse economic climate by reducing their margins and increasing the discounts. In their regression analysis, they also find a positive link between discounts and posted prices, which is in line with the results displayed in Table 6 of the supplementary material.

In 2000, the UK Competition Commission investigated the competitiveness of the UK new car market and gathered data on average discounts by brand and segment (UK Competition Commission, 2000). The dataset is very reliable since it was collected directly from dealers. The report reveals that the average discount lies between 7.5% and 8%, also broadly in line with our estimated average discount. Once more, the difference may stem from differences between the two markets and the periods under consideration. This report also refers to a consumer survey conducted in 1995 asking automobile purchasers whether or not they obtained a discount over the posted price. This survey reveals that 17% of purchasers paid the posted price whereas 37% bargained and obtained a discount and 29% were automatically offered a discount. This figure of 17% is close to our estimation of 13%. Furthermore, the fact that some purchasers were "automatically offered a discount" corroborates our assumption that discounts are used as a tool to price discriminate because the posted price is not optimal for some consumers.

A direct comparison of the distribution of discounts we estimate and evidence on the U.S. market is more complicated. The two countries differ in particular in the characteristics of the retailing sector. In the U.S., dealers are all independent from the manufacturers, as opposed to France where only 10% of dealers are independent. Therefore, the pricing model we rely on seems less credible for the U.S. car market and we can then expect more spatial dispersion and price negotiation in the U.S. Despite these differences, Busse et al. (2012) report that the rebates represent on average 9.6% of the transaction prices, which is once more consistent with our estimated discounts.

Finally, few papers correlate the magnitude of discounts to age and income. Harless and Hoffer (2002) and Chandra et al. (2017) analyze price discrimination with respect to age and gender on the U.S. car market using dealers margins (see also Langer, 2016, focusing on discrimination by gender and marital status using transaction prices from survey data). They both report a positive correlation between the margins and purchasers' age. In the web appendix of the 2012 version of her paper, Langer documents significant price discrimination with respect to income, the high income groups of consumers (for both men and women) are associated with higher margins. These two results are in line with our findings on the estimated discounts and mark-up rates.

6.5 Other sources of price dispersion

There are other sources of price dispersion, apart from third-degree price discrimination, that we have not taken into account because of data limitation. While we discuss in detail below these other sources, we first provide a rough assessment of the importance of third-degree price discrimination in the overall dispersion of new car prices. For that purpose, we compute the variance of log transaction prices observed in the BdF survey. We then compare it with the variance of log posted prices observed in the CCFA database, restricting ourselves to the same subset of cars as in BdF (defined by brand, model and fuel type). Finally, we also compute the variance of log transaction prices estimated with our model, still for the same subset of cars. If third-degree price discrimination was the only reason for price dispersion, we would expect this latter variance to be equal to the variance obtained in BdF, and larger than the variance of list prices. The results are displayed in Table 12. While posted prices account for 80% of the variance of log transaction prices observed in BdF, our transaction prices account for 86.6% of this variance. In other words, third-degree price discrimination is able to capture around one third (33.4%) of the unexplained variance of observed log transaction prices in BdF.

$V\left(\ln(p^{BdF})\right)$	$V\left(\ln(\widetilde{p}_Y)\right)$	$V\left(\ln(p_Y^D)\right)$
0.121	0.097	0.105

Notes: $\widetilde{p}_Y$ (resp. p_Y^D) corresponds to the list price (resp. transaction price estimated with our model) associated to the consumer's choice of car Y.

Table 12: Variance of observed log transaction prices ($\ln(p^{BdF})$, log posted prices ($\ln(\widetilde{p}_Y)$) and log estimated transaction prices $\ln(p_Y^D)$.

Temporary promotions

Temporary promotions such as manufacturer rebates constitute another source of price dispersion. With exhaustive data on such promotions, we could take them into account in our model by modifying accordingly the list prices. While we we are not aware of any such exhaustive data, we obtained monthly data on rebates for a subset of cars from the consumer price index department of Insee. We document in Section 2.3.5 of the supplementary material the effect of accounting for such rebates. We find in particular that these rebates are on average around 3 times lower than our estimated average discount. Thus, while not negligible, they do not appear as important as discounts coming from third-degree price discrimination.

Price discrimination with respect to unobserved characteristics

An additional source of price dispersion is coming from other demographic characteristics that we do not observe but might be used to better price discriminate within the groups we consider.

Gender, race but also education (as a proxy of negotiation skills) are omitted from our analysis because we do not have those data. If there has been evidence of price discrimination against women in the U.S. (i.e. they pay more for the same car), recent results point towards a reduction of those differences over time (see, e.g. Chandra et al., 2017). Nevertheless, we performed a robustness check where we assess the effect of neglecting discrimination against women. We suppose that men receive on average a discount of the same amount as the one reported by Langer (2016). The effects on parameter estimates are small (see Section 2.3.3 of the supplementary material for details).

Price negotiation

Price negotiation could be another cause of price dispersion, as shown by, e.g., Scott Morton et al. (2011). However, to the extent that there is no search cost, price negotiation can be modeled similarly as price discrimination with respect to unobserved characteristics (see Huang, 2016, for such an approach). These characteristics would include, e.g., patience or bargaining disutility. Sellers would then discriminate between, say, patient and impatient consumers, offering lower prices to patient consumers. In such a set-up, our model would capture the benefit of negotiation net of the negotiation cost (in monetary terms). Now, search costs may matter as well. Scott Morton et al. (2011) report that in the U.S. consumers in the lowest search cost quartile pay on average 1.3% less than those in the highest quartile. This is significant but smaller than the magnitude of our discounts. Search costs may also be lower in France than in the U.S., since dealers are more spatially concentrated, with one dealer every 110 square kilometer versus one every 580 square kilometer in the U.S.

Spatial price dispersion

While we allow for price variation between different municipalities through the segmentation in 6 consumer groups, we neglect price variation that could occur because of differences in competition intensity between local markets (see Albuquerque and Bronnenberg, 2012; Murry and Zhou, 2016, for papers using local prices). Moreover, even under perfect integration between manufacturers and dealers, price variation may arise because of variations in marginal costs across dealers. These variations, which would violate Assumption 1, could be the consequence of heterogeneity in real estate prices. Using 2017 French notary data and a U.S. estimate of the average estate cost per car, we find that the marginal costs would be between €455 lower and €35 higher than the cost of the pivot group. The average cost difference represents on average 1.2% of the posted price of cars. To evaluate the sensitivity of our estimates to such heterogeneous costs, we simulate our model with our estimated parameters of demand but supposing that marginal costs exhibit such systematic differences across groups. We then estimate the model under the wrong assumption of equal marginal costs. The results suggest that the effect of neglecting cost heterogeneity are small and affect mainly the estimated coefficients of the intercept. We refer to Section 2.3.2 of the supplementary material for more

details.

Trade-in and financing

The purchase of a new car often involves the trade-in of an old car or the purchase of a financing plan. The characteristics of these ancillary transactions are rarely recorded but the purchase price of a trade-in car and the loan rate can be used by car sellers to do price discrimination. Therefore, even if we observed no price dispersion in transaction prices, sellers could still price discriminate through the trade-in value and the financing loan rate.

In the supplementary material, we investigate the effect of neglecting the trade-in by simulating the market equilibrium when a fraction of consumers use their old car as trade-in in each group of consumers we consider. We assume, in line with the results of Busse and Silva-Risso (2010), that sellers perfectly substitute the profits from new and used car sales and that they are able to sell used cars €500 higher than individuals. For simplicity, we also assume that the price at which buyers resell their car is constant and equal to €3,000. We then estimate the model neglecting that some consumers have a trade-in component in their transaction. We find that mark-ups and price elasticities are barely affected by the omission of the trade-in component in the model.

7 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 consumers to price discriminate. Adapting our methodology to such second-degree price discrimination does not seem obvious, and is left for future research.

References

Aksoy-Pierson, M., Allon, G. and Federgruen, A. (2013), 'Price competition under mixed multinomial logit demand functions', *Management Science* **59**(8), 1817–1835.

Albuquerque, P. and Bronnenberg, B. J. (2012), 'Measuring the impact of negative demand shocks on car dealer networks', *Marketing Science* **31**(1), 4–23.

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.

Berry, S., Levinsohn, J. and Pakes, A. (2004), 'Differentiated products demand systems from a combination of micro and macro data: The new car market', *Journal of political Economy* **112**(1), 68–105.

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.

Busse, M. R., Knittel, C. R., Silva-Risso, J. and Zettelmeyer, F. (2012), Did 'cash for clunkers' deliver? the consumer effects of the car allowance rebate system. Working Paper (November 2012 version).

Busse, M. R. and Silva-Risso, J. M. (2010), "'one discriminatory rent" or "double jeopardy": Multicomponent negotiation for new car purchases', *The American Economic Review* **100**(2), 470–474.

Caplin, A. and Nalebuff, B. (1991), 'Aggregation and imperfect competition: On the existence of equilibrium', *Econometrica* pp. 25–59.

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.

Ciliberto, F. and Williams, J. W. (2014), 'Does multimarket contact facilitate tacit collusion? inference on conduct parameters in the airline industry', *The RAND Journal of Economics* **45**(4), 764–791.

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

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

de Roos, N. and McKenzie, J. (2014), 'Cheap tuesdays and the demand for cinema', *International Journal of Industrial Organization* **33**, 93–109.

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. (2014), Identifying industry margins with unobserved price constraints: structural estimation on pharmaceuticals. CEPR Discussion Paper No. DP9881.

Einav, L. (2007), 'Seasonality in the us motion picture industry', *The Rand journal of economics* **38**(1), 127–145.

Einav, L., Leibtag, E. and Nevo, A. (2010), 'Recording discrepancies in nielsen homescan data: Are they present and do they matter?', *Quantitative Marketing and Economics* **8**(2), 207–239.

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.

Gale, D. and Nikaido, H. (1965), 'The jacobian matrix and global univalence of mappings', *Mathematische Annalen* **159**(2), 81–93.

Gandhi, A., Lu, Z. and Shi, X. (2013), Estimating demand for differentiated products with error in market shares. cemmap working paper.

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.

Holmes, T. J. (1989), 'The effects of third-degree price discrimination in oligopoly', *American Economic Review* **79**, 244–250.

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. (2014), Price negotiations and bargaining costs. Working paper.

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

Kaul, A., Pfeifer, G. and Witte, S. (2016), 'The incidence of cash for clunkers: Evidence from the 2009 car scrappage scheme in germany', *International Tax and Public Finance* **23**(6), 1093–1125.

Knittel, C. R. and Metaxoglou, K. (2014), 'Estimation of random-coefficient demand models: Two empiricists' perspective', *Review of Economics and Statistics* **96**(1), 34–59.

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/publications/2013/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'.

Murry, C. and Zhou, Y. (2016), 'Consumer search and automobile dealer co-location'.

Nevo, A. (2000), 'A practitioner's guide to estimation of random-coefficients logit models of demand', *Journal of Economics & Management Strategy* **9**(4), 513–548.

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

Nocke, V. and Schutz, N. (2016), Multiproduct-firm oligopoly: An aggregative games approach. Working paper.

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

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

Scott Morton, F., Silva-Risso, J. and Zettelmeyer, F. (2011), 'What matters in a price negotiation: Evidence from the us auto retailing industry', *Quantitative Marketing and Economics* **9**(4), 365–402.

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.

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/20111202195250/http://competition-commission.org.uk/rep_pub/reports /2000/439cars.htm#summary.

Warner, E. J. and Barsky, R. B. (1995), 'The timing and magnitude of retail store markdowns: evidence from weekends and holidays', *The Quarterly Journal of Economics* **110**(2), 321–352.

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

A Proofs

A.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 $K_r = [0, \bar{p}]^r$ for any $r \in \mathbb{N}$ and introduce the function

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

The proof is divided in three steps. First, we show that it suffices to prove 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 2, 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 $(\theta^d, p^d) \mapsto \delta_{\theta^d}^d(p^d)$ is C^1. Let $\mu_{\theta,j}^d(e, u, p_j^d)$ 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

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

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

$$Q_{\theta^d,j}(p^d, \delta^d) = \int s_{\theta^d}(e, u, p^d, \delta^d) dP_{E,\zeta}^d(e, u) - s_j^d.$$

Then $\delta_{\theta^d}^d(p^d)$ is defined by $Q_{\theta^d}(p^d, \delta_{\theta^d}^d(p^d)) = 0$. 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,

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

$$< \frac{\partial Q_{\theta^d, j}}{\partial \delta^d_j}.$$

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

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

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

$$\left\{\Omega^d(\theta^d, p^d, \delta^d), (\theta^d, p^d) \in \mathbb{R} \times K^J\right\} \subset \mathcal{A}^+ \text{ or } \left\{\Omega^d(\theta^d, p^d, \delta^d), (\theta^d, p^d) \in \mathbb{R} \times K^J\right\} \subset \mathcal{A}^-. \tag{19}$$

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, (19) 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

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

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 $\Omega^d(\theta_0^d, p^d, \delta^d)$ depend on p^d. Hence, $R_j^{d,d'}(\theta_0^d, \theta_0^{d'}) = 0$. By continuity of $R_j^{d,d'}$, there exists $\overline{\Sigma}^d$ such that for all $\theta^d \in [0, \overline{\Sigma}^d] \times \{0\}$, $R_j^{d,d'}(\theta^d, \theta_0^{d'}) < 1$. $\pi^{p,d} \mapsto$ $\max_{\Sigma^{p,d} \in [0, \overline{\Sigma}^d]} R_j^{d,d'}(\Sigma^{p,d}, \pi^{p,d})$ is also continuous and smaller than 1 at $\pi^{p,d} = 0$. Then there exists $\overline{\pi}^d$ such that for all $(\Sigma^{p,d}, \pi^{p,d}) \in [0, \overline{\Sigma}^d] \times [-\overline{\pi}^d, \overline{\pi}^d]$, $R_j^{d,d'}(\Sigma^{p,d}, \pi^{p,d}, \theta_0^{d'}) < 1$.

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 $(\theta^d, \theta^{d'}) \in ([0, \overline{\Sigma}^{d,d'}] \times [-\overline{\pi}^{d,d'}, \overline{\pi}^{d,d'}])^2$, $R_j^{d,d'}(\theta^d, \theta^{d'}) < 1$. In turn, this implies that $q_{\theta,j}^{d,d'}$ 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.

A.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 $a = \widehat{\sigma}_1^2 - \widehat{\sigma}_0^2$, $b = 2\sqrt{J}\widehat{\sigma}_0$ and $c = -J - \widehat{\sigma}_1^2 t$. The result follows using standard arguments on quadratic inequalities. The same reasoning applies for p_1.

A.3 Proof of Theorem 2

In the following, we let $\Theta = \{(\alpha^1, ..., \alpha^{n_D}, \beta^1, ..., \beta^{n_D}) \in A^{n_D} \times B^{n_D}\}$ and $\Pi = \prod_{j=1}^{J}[c_j, +\infty)$. The proof is divided in five steps. We first bound market shares under our assumption on ξ. Then we show that the solution $p(\xi, \theta)$ of the first-order conditions defined by Equation (15) are indeed well-defined (i.e., the system admits a unique solution) and regular. Existence and uniqueness are well-known in such a context (e.g. Caplin and Nalebuff, 1991) but we prove them for completeness. Third, we show that for any $\theta_0 \in \Theta_0$, $p(\xi, \theta_0)$ is an attractive fixed point of $M_{s(\xi, \theta_0), \theta_0}$. Fourth, we prove that $M_{s(\xi, \theta), \theta}$ is a contraction on a neighborhood of $p(\xi, \theta_0)$, for well chosen (ξ, θ). Finally, we prove the convergence of $(p_n)_{n \in \mathbb{N}}$ towards $p(\xi, \theta)$.

1. $\max_j s_j^d(p, \xi, \theta)/s_0^d(p, \xi, \theta) \leq 1/2$ **and** $\max_j s_j^d(p, \xi, \theta) < 1/3$ **for all** $(p, \xi, \theta) \in \Pi \times K \times \Theta.$

First, for all (d, j, ξ, p, θ), we have

$$\ln(s_j^d(p, \xi, \theta)/s_0^d(p, \xi, \theta)) = p\alpha^d + X_j'\beta^d + \xi_j.$$

The restriction $\xi \in K$ then implies

$$\ln(s_j^d(p, \xi, \theta)/s_0^d(p, \xi, \theta)) \leq -\ln 2.$$

Hence, $s_j^d(p,\xi,\theta)/s_0^d(p,\xi,\theta) \le 1/2$, which implies that $\max_j s_j(p,\xi,\theta) < 1/3$ for all $(p,\xi,\theta) \in \Pi \times K \times \Theta$.

2. $p(\xi,\theta)$ is well defined, C^1 and is a fixed point of $M_{s(\xi,\theta),\theta}$.

First, consider the function $R = (R_1,...,R_J)$ defined on Π by $R_j(p) = c_j - 1/(\alpha^d(1 - s_j^d(p,\xi,\theta)))$. For simplicity, we first let the dependence of R in d and $\theta \in \Theta$ implicit here. Let us consider the convex compact set $\mathcal{C} = \prod_{j=1}^J [c_j, \overline{p}_j]$, with $\overline{p}_j > c_j - 3/[2\alpha^d]$. By Step 1, $s_j^d(p,\xi,\theta) < 1/3$ for all $j = 1,...,J$. Therefore, for all $p \in \mathcal{C}$, $R_j(p) \le c_j - 3/[2\alpha^d]$ and thus, $R(p) \in \mathcal{C}$. Then, by Brouwer's theorem, R admits at least one fixed point on $\mathcal{C} \subset \Pi$, implying that (15) has at least one solution on Π.

To prove that this solution is unique, let $Q(p) = p - R(p)$ be defined on Π. We have

$$\frac{\partial Q_j}{\partial p_k}(p) = \mathbb{1}\{j = k\} - \frac{s_j^d(p,\xi,\theta)(\mathbb{1}\{j=k\} - s_k^d(p,\xi,\theta))}{(1 - s_j^d(p,\xi,\theta))^2}.$$

By Step 1, $s_j^d(p,\xi,\theta) < 1/2$. Therefore, $\partial Q_j/\partial p_j > 0$. Also,

$$\sum_{k \ne j} \left| \frac{\partial Q_j}{\partial p_k}(p) \right| = \frac{s_j^d(p,\xi,\theta)(1 - s_0(p,\xi,\theta) - s_j^d(p,\xi,\theta))}{(1 - s_j^d(p,\xi,\theta))^2}$$

$$< \frac{(1 - 2s_j^d(p,\xi,\theta))(1 - s_j^d(p,\xi,\theta))}{(1 - s_j^d(p,\xi,\theta))^2}$$

$$< \frac{\partial Q_j}{\partial p_j}(p),$$

where we have used in the last inequality $s_j^d(p,\xi,\theta) < 1 - 2s_j^d(p,\xi,\theta)$, again by Step 1. Thus, the Jacobian matrix of Q is diagonally dominant with positive diagonal elements. Hence, it is a P-matrix (see Example 2.3 in Gale and Nikaido, 1965). By Gale and Nikaido's Theorem 4, Q is injective on Π. Hence, there is a unique solution to $Q(p) = 0$, implying that $p(\xi,\theta)$ is well-defined.

Now, letting the dependence in (ξ,θ) explicit in Q, we have $Q(p(\xi,\theta),\xi,\theta) = 0$. Moreover, $Q(.,.,.)$ is C^1 and the matrix of (i,j) term $[\partial Q_i/\partial p_j(p,\xi,\theta)]$ is invertible by what precedes. Hence, by the implicit function theorem, $p(.,.)$ is also C^1.

Finally, $p(\xi,\theta)$ satisfies Equation (15). Therefore, by Assumption 2,

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

$$= c_j + f_j\left(\frac{1}{\alpha^1(1 - s_j^1(p,\xi,\theta))},...,\frac{1}{\alpha^{n_D}(1 - s_j^{n_D}(p,\xi,\theta))}\right)$$

$$= p_j^d(\xi,\theta) - \frac{1}{\alpha^d(1 - s_j^d(p,\xi,\theta))} + f_j\left(\frac{1}{\alpha^1(1 - s_j^1(p,\xi,\theta))},...,\frac{1}{\alpha^{n_D}(1 - s_j^{n_D}(p,\xi,\theta))}\right).$$

Moreover, $\xi(p(\xi,\theta),s(\xi,\theta),\theta) = \xi$. Thus, by definition of $M_{s,\theta}$, $p(\xi,\theta)$ is a fixed point of $M_{s(\xi,\theta),\theta}$.

3. $p(\xi,\theta_0) = p_0$ is a fixed point of $M_{s(\xi,\theta_0),\theta_0}$.

First, for all $\theta_0 \in \Theta_0$, Equation (15) is the same for all d and admits a unique solution. Therefore, $p_j^d(\xi,\theta_0)$ does not depend on d. Then, by Assumption 2, we have, for all ξ,

$$f_j\left(\frac{1}{\alpha^1(1 - s_j^1(p,\xi,\theta))}, ..., \frac{1}{\alpha^{n_D}(1 - s_j^{n_D}(p,\xi,\theta))}\right)$$
$$= \frac{1}{\alpha^d(1 - \sum_{k\in\mathcal{J}_{f_j}} s_k^d(p,\xi.\theta))}. \tag{20}$$

This implies that $p_j^d(\xi,\theta_0) = \widetilde{p}_j^d$. Hence, by definition of p_0, $p(\xi,\theta_0) = p_0$.

4. For all well-chosen θ,ξ, $M_{s(\xi,\theta),\theta}$ is a contraction on a neighborhood of $p(\xi,\theta_0)$.

Let us define

$$q_{j,\theta}^{d,d'}(p) = \frac{1}{\alpha^{d'}(1 - s_j^{d'}(p,\xi(p,s(\xi,\theta),\theta),\theta))} - \frac{1}{\alpha^d(1 - s_j^d(p,\xi(p,s(\xi,\theta),\theta),\theta))}.$$

By the same argument as in Step 1 of the proof of Theorem 1, it suffices to show that for all θ in the neighborhood of $\theta_0 \in \Theta_0$ and for all j,d,d', the function $q_{j,\theta}^{d,d'}(.)$ is a contraction on a neighborhood of $p(\xi,\theta_0)$. The result holds if we show that for appropriate θ,p,

$$\sum_{\ell,d''}\left|\frac{\partial q_{j,\theta}^{d,d'}}{\partial p_\ell^{d''}}(p)\right| < 1. \tag{21}$$

We first show the inequality for $\theta = \theta_0$ and $p = \widetilde{p}$. Here, we crucially rely on the fact that for all j,d,d',
$$s_j^d(\widetilde{p},\xi,\theta_0) = s_j^{d'}(\widetilde{p},\xi,\theta_0) = s_j(\xi,\theta_0),$$

where $s_j(\xi,\theta_0)$ is the j-th coordinate of $s(\xi,\theta_0)$. This implies in particular that for all j,k,d,d',

$$\frac{\partial s_j^d}{\partial \xi_k}(\widetilde{p},s(\xi,\theta_0),\theta_0) = \frac{\partial s_j^{d'}}{\partial \xi_k}(\widetilde{p},s(\xi,\theta_0),\theta_0).$$

This also implies that $\xi(.,s(\xi,\theta_0),\theta_0)$ is symmetric in p_k^d and $p_k^{d'}$, so that

$$\frac{\partial \xi_j}{\partial p_k^d}(\widetilde{p},s(\xi,\theta_0),\theta_0) = \frac{\partial \xi_j}{\partial p_k^{d'}}(\widetilde{p},s(\xi,\theta_0),\theta_0).$$

Therefore, all derivative terms related to $\xi(p, s(\xi, \theta), \theta)$ in $\partial q^{d,d'}_{j,\theta_0}/\partial p^{d'}_\ell$ simplify, and we get

$$\frac{\partial q^{d,d'}_{j,\theta_0}}{\partial p^{d'}_\ell}(\widetilde{p}) = \frac{s^d_j(\widetilde{p}, \xi, \theta_0)(\mathbb{1}\{j = \ell\} - s_\ell(\xi, \theta_0))}{(1 - s^d_j(\widetilde{p}, \xi, \theta_0))^2},$$

and similarly, $\partial q^{d,d'}_{j,\theta_0}/\partial p^{d'}_\ell(\widetilde{p}) = -\partial q^{d,d'}_{j,\theta_0}/\partial p^{d'}_\ell(\widetilde{p})$. Thus,

$$\sum_{\ell,d''}\left|\frac{\partial q^{d,d'}_{j,\theta_0}}{\partial p^{d''}_\ell}(\widetilde{p})\right| = \frac{2s^d_j(\widetilde{p}, \xi, \theta_0)}{(1 - s^d_j(\widetilde{p}, \xi, \theta_0))^2}[2(1 - s^d_j(\widetilde{p}, \xi, \theta_0)) - s^d_0(\widetilde{p}, \xi, \theta_0)].$$

The right-hand side is strictly smaller than 1 if and only if

$$s^d_j(\widetilde{p}, \xi, \theta_0)(6 - 2s^d_0(\widetilde{p}, \xi, \theta_0) - 5s^d_j(\widetilde{p}, \xi, \theta_0)) < 1. \tag{22}$$

By Step 1, $-2s^d_0(\widetilde{p}, \xi, \theta_0) \le -4s^d_j(\widetilde{p}, \xi, \theta_0)$. Moreover, $x \mapsto x(6 - 9x)$ reaches its maximum 1 at $x = 1/3$. Because $s^d_j(\widetilde{p}, \xi, \theta_0) < 1/3$ by Step 1 again, (22) holds.

Now, given the definition of $q^{d,d'}_{j,\theta}(.)$, it suffices to show that $(p, \theta) \mapsto \xi(p, s, \theta)$ is C^1. The function $(p, \xi, \theta) \mapsto s(p, \xi, \theta)$ is smooth, $s(p, ., \theta)$ is injective (Berry, 1994) and the Jacobian matrix $\partial s/\partial \xi(p, \xi, \theta)$ is diagonally dominant and therefore invertible. Thus, by the inverse function theorem, $(p, \theta) \mapsto \xi(p, s, \theta)$ is C^1.

5. There exists Θ_1 neighborhood of Θ_0 such that for all $(\xi, \theta) \in K \times \Theta_1$, $(p_n)_{n \in \mathbb{N}}$ converges towards $p(\xi, \theta)$.

First, by what precedes, there exists a neighborhood $\mathcal{V}_1 \subset \Theta$ of θ_0 and $r > 0$ such that for all $(p, p') \in B(p_0, r)^2$,
$$\|M_{s(\xi,\theta),\theta}(p) - M_{s(\xi,\theta),\theta}(p')\| \le C\|p - p'\|. \tag{23}$$
with $C < 1$.

Second, $p(., .)$ is C^1 by Step 2. Moreover, for all $\xi \in K$, $p(\xi, \theta_0) = p_0$. Hence,

$$\|p(\xi, \theta) - p_0\| = \|p(\xi, \theta) - p(\xi, \theta_0)\| \le \left[\max_{(\xi',\theta') \in K \times \Theta} \|\partial p/\partial \theta(\xi', \theta')\|\right] \|\theta - \theta_0\|.$$

Hence, there exists a neighborhood $\mathcal{V}_2 \subset \Theta$ of θ_0 such that for all $(\xi, \theta) \in K \times \mathcal{V}_2$, $\|p(\xi, \theta) - p_0\| \le r/2$.

Then, for all $(\xi, \theta) \in K \times \mathcal{V}_1 \cap \mathcal{V}_2$, we prove by induction that $p_n \in B(p_0, r)$ and $\|p_n - p(\xi, \theta)\| \le (r/2)C^n$. The result holds for $n = 0$ by what precedes. Suppose that it holds for n. Then,

because $p_n \in B(p_0, r)$, by Equation (23),

$$
\begin{aligned}
\|p_{n+1} - p(\xi, \theta)\| &= \|M_{s(\xi,\theta),\theta}(p_n) - M_{s(\xi,\theta),\theta}(p(\xi, \theta))\| \\
&\le C\|p_n - p(\xi, \theta)\| \\
&\le (r/2)C^{n+1}.
\end{aligned}
$$

Moreover, by the triangular inequality,

$$
\|p_{n+1} - p_0\| \le (r/2)C^{n+1} + r/2 \le r.
$$

Hence, $p_{n+1} \in B(p_0, r)$ and the result holds for $n + 1$. Therefore, it holds for all n, which shows that $(p_n)_{n\in\mathbb{N}}$ converges towards $p(\xi, \theta)$.