Supplement to "Automobile Prices in Market Equilibrium with Unobserved Price Discrimination"

Xavier D'Haultfœuille* Isis Durrmeyer† Philippe Février‡

September 13, 2018

Abstract

In this supplement, we first consider extensions of the model that are only briefly mentioned in the main paper. Second, we present Monte Carlo simulations and discuss numerical aspects of the estimation algorithm. Third, we display additional material on the application. Finally, we prove the convergence of the algorithm for the extension of the model to unobserved groups.

1 Extensions

1.1 Additional details on the case with unobserved groups

1.1.1 Motivation

We first discuss examples where price discrimination is done across groups that are unobserved by the econometrician. Note that we keep the same condition on price observability, namely Assumption 3. A first example is when both aggregate data on sales and 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 3 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 to estimate precisely the market shares $(s_j^d)_{d=1,\dots,n_D}$ 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

*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

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.[1] 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 $p_j^{d_j}$.

Another example where our methodology 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).

1.1.2 Convergence of the algorithm

We consider here the convergence of the sequence $(p_n)_{n\in\mathbb{N}}$, defined by $p_0 = \widetilde{p}_j$ and $p_{n+1} = M_{s(\xi,\theta),\theta}(p_n)$, towards $p(\xi,\theta)$ (the notation here is the same as in Section 3.4.2). This convergence holds when each firm sells only one product and when, roughly speaking, there is not too much heterogeneity between groups.[2] This condition is related to the one we impose in Theo-

[1] 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.

[2] The case of multiproduct firms is left for future research. It may be dealt with the approach proposed by Nocke and Schutz (2018).

rem 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

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

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 2-4 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* (13).

The proof is displayed in Section 4. 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 context,

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

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 2.1 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 2.6 below for more details.

1.2 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. 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 basically imposes that the proxy variable $\widetilde{D}$ is related to D. Let Y_i denote the product choice of consumer i. Under the first condition, we have

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

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.

1.3 Other supply-side models and moment conditions

So far, we have considered 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 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.

Related to this, since it is necessary to specify the nature of competition and pricing conduct, we can include moments corresponding to the supply side as in the standard BLP model. To do so, a relationship between marginal costs and cost shifter variables should be posited. Adding supply-side moment conditions can improve the accuracy of estimation.

1.4 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 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 within 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 counterparts. Another possibility is to use additional information on the overall distribution of transaction prices or discounts. Additional conditions then take the form of differences between model-based moments of the price distribution and their empirical counterparts.

1.5 Other functional forms on price effects

In our main setting, we have implicitly considered, following common practice, that indirect utilities depend linearly on disposable income. Namely, these utilities were supposed to depend on $\alpha_i(y_i - p_j)$, where y_i denotes the income before purchase. $\alpha_i y_i$ 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 a function 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. But Equations (8) and (10) remain unchanged. The only difference is that the terms entering into Ω^d are now different from Equation (9). But other than that, the construction of the moment conditions follows exactly the same methodology.

2 Monte Carlo simulations

2.1 Data Generating Process

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^d_{jt}, 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 $X_{1jt}, W_{1jt}, W_{2jt}, W_{3jt}$, ω_{jt} and ξ^d_{jt} are mutually independent. $X_{1jt} \sim \mathcal{U}[1,2]$, $W_{kjt} \sim \mathcal{U}[0,1](k = 1,2,3)$ while ξ^d_{jt} and ω_{jt} are two normal variables $\mathcal{N}(0, 0.1)$. The parameters of preferences are summarized in Table 1. Groups of consumers differ 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 characteristic (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 $(s^d_{jt}, p^d_{jt})_{d=1,2,3,4}$, 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.

2.2 Numerical Aspects

We use the following method to compute the GMM estimator. First, to approximate the aggregate market shares, we use 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 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 varies in $\{0, 0.1, ..., 1\}$. 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" and "maximum" are the mean and 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 maximal price differences 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 comes 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 converges to the same parameter values in all 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.

2.3 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 sides, 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 investigate the performance of the test under uniform pricing. We run 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 is accepted 99.5% of the time.

2.4 Effects of ignoring price discrimination

We finally investigate the bias from ignoring price discrimination and compute the relative errors in estimated parameters and economic variables when the econometrician uses other prices instead of the true transaction prices. We consider the three examples of observed prices satisfying Assumption 3 discussed in the paper: maximal prices, (sales-weighted) average prices and one price observed for each product. As explained in the paper, ignoring price discrimination results in bias in the estimators because using the observed prices instead of the true prices constitutes a non-classical measurement error problem. The results of Table 4 quantify this bias.

The price sensitivity parameters always display substantial bias, with relative errors ranging between -6.4% and -11.2%, depending on the prices used and the group of consumers. The errors in the parameter of heterogeneity in price sensitivity σ^p lie between -19.2% and -17.4% of the true value. The intercept, which represents the utility of buying a product relative to choosing the outside good, displays very large bias from -115% to almost 70%. The signs of the bias vary with the group and the prices observed but we obtain similar bias when using average prices and one transaction price. The coefficients for the non-price characteristic are estimated with a very small bias under the first scenario while under the second and third models, the bias is of around -3% for all the groups of consumers. The parameters of the marginal cost equation are also estimated with large bias for the intercept, and smaller errors for the cost shifters.

Ignoring price discrimination has also large effects on economic variables because those are direct functions of prices and parameters. We see indeed that price elasticities, mean consumer surplus, average mark-ups and marginal costs are estimated with large errors when using the observed prices instead of the true prices. Errors in price elasticities are between -6.8% and 7.7%. As expected, these errors are rather small for the pivot group (Group 1) when we use maximal prices but the biases are large for the other groups and under Specifications (2) and (3). Again, errors in average surplus are small for the group for which prices are observed. On the other hand, the bias in average consumer surplus reaches 13.3% for the most price sensitive group in Specification (3). Biases in the estimated mark-ups are also substantial. We underestimate the mark-up for Group 1 and overestimate the mark-ups of all the other groups. The bias goes up to 78% for the most price sensitive group in Specification (3). Finally, the errors in estimated marginal costs are very small when the econometrician uses the average prices or one transaction price but they reach almost 6% when using the maximal prices. At the end, those simulation results show that the errors on variables of interest can be large and quite unpredictable as they depend on the consumer group and the prices used to estimate the demand.

	(1)	(2)	(3)
	$\widetilde{p} = \max_d p^d$	$\widetilde{p} = \frac{\sum s^d \phi^d p^d}{\sum s^d \phi^d}$	$\widetilde{p} = p^{d_j}$
Price sensitivities			
α^1	-6.4	-8.5	-7.8
α^2	-8.3	-10.4	-9.9
α^3	-8.8	-11	-10.2
α^4	-8.9	-11.2	-10.4
σ^p	-19.2	-17.6	-17.4
Intercept			
β_I^1	24.8	50.5	50.9
β_I^2	-0.8	66.9	69.8
β_I^3	-29.4	12.1	13.5
β_I^4	-115	-14.7	-11.3
Product characteristic			
β_X^1	-0.1	-3.5	-3.7
β_X^2	-0.1	-3.4	-3.6
β_X^3	1.2	-2.9	-3.2
β_X^4	2	-3.1	-3.4
Cost equation			
γ^I	18.5	-5.6	-8.2
γ^X	5.6	1.7	1.4
γ_1	0	1.9	2.1
γ_2	0	1.9	2.1
γ_3	0.1	2	2.2
Price elasticities			
ϵ^1	0.2	-6	-6.8
ϵ^2	2.5	-3.8	-4.6
ϵ^3	5.9	-0.6	-1.5
ϵ^4	7.7	1.2	0.3
Mean consumer surplus			
group 1	0.7	1.7	5
group 2	6.4	7.6	10
group 3	9.5	11	12.5
group 4	10.4	12.2	13.3
Average mark-up			
group 1	-19.2	-14	-13
group 2	8.3	15.3	16.6
group 3	35.7	44.5	46.2
group 4	65.6	76.3	78.3
Marginal cost	5.8	0.2	-0.3

Reading notes: All values are in %. Relative bias in parameter estimates are computed as the difference between the average estimated parameters and the true parameters divided by the true parameters. Bias in average price elasticities, mark-ups and consumer surplus are computed as the relative difference in the average variable over products, markets and simulations. We use the DGP presented in Section 2.1. The first column corresponds to the case where we observe the maximal prices. The second column corresponds to the case where only average prices are observed while in the third column only one transaction price is observed and it is drawn randomly using the sales as probability weights.

Table 4: Relative bias in the estimated parameters when ignoring price discrimination.

2.5 The logit and nested logit cases

We explore the computational aspects of the GMM in the logit and nested logit cases, for which we optimize on $(\alpha^1, ..., \alpha^{n_D})$ and $(\alpha^1, ..., \alpha^{n_D}, \sigma^1, ..., \sigma^{n_D})$, respectively. We consider nearly the same DGP as in Section 2.1. The utilities and marginal costs are set in the same way, except that because n_D varies hereafter, the coefficients of the characteristics are fixed in a slightly different manner. Specifically, the intercept and the coefficient of the non-price characteristic are equal to -0.75 and 2 for all groups, while the price sensitivity coefficient is set on an equally spaced grid ending at -1.5 and with a width of -0.5 (e.g., -3, -2.5, -2 and -1.5 with 4 demographic groups). As in Section 2.1, we consider 25 markets on which 4 firms sell 6 products each. In the nested logit, we consider 8 products per nest, and fix the coefficient of correlation of products within segment $\sigma^d = 0.5$ for each group d. We then make the number of firms vary from 4 to 16. Hence, the number of nests varies as we change the number of firms. We suppose that the econometrician does not observe transaction prices p_j^d but only the list prices, which are supposed to satisfy $\widetilde{p}_j = \max(p_j^1, ..., p_j^{n_D})$. Finally, we use product characteristics, cost shifters and the sums of the characteristics of the other products of the firm as instruments for the estimation. In the nested logit, we also rely on the additional instrumental variables defined as the sum of the characteristics of the other products in the segment.

To minimize the GMM function, we first have to choose a starting point. In both the logit and nested logit models, we choose the 2SLS estimator of that model, supposing that the transaction prices are equal to $\widetilde{p}$. Regarding the minimization itself, we use the BFGS quasi-newton algorithm. This choice may seem surprising, given that the objective function is not differentiable everywhere because of the maximum function appearing in the price equation. However, the algorithm works very well in practice and is much faster than the simplex algorithm. Such a good behavior is documented in the optimization literature, see e.g. Lewis and Overton (2013). To assess whether the algorithm converges to the global minimum, we draw randomly, for each simulation, 10 other initial points. We then consider to have reached the global minimum if for our preferred estimator, based on the initial 2SLS estimator, the value function is smaller or equal to the minimum of these other 10 optimizations, with a tolerance of 10^{-3}.

Table 5 reports the computational aspects of the optimization: the average time (in seconds) to compute our estimator based on the 2SLS starting point, and the proportion of simulations for which this estimator reaches the global minimum of the value function. The algorithm always converges. Moreover, in the vast majority of the simulations the estimator is the global minimum of the objective function according to the criterion above. Table 5 also reports the statistical properties of our GMM estimator. As expected, the root mean squared error (RMSE) decreases with the number of products in both models. We also observe that the estimator of α is more imprecise in the nested logit than in the logit model. Note however that $\alpha/(1-\sigma)$, which is the main term of price elasticities, is much better estimated, with a RMSE comparable to the RMSE of α in the logit model. Intuitively, it is difficult to disentangle the effect of prices (α) from the effect of intra-group log market shares (σ) in the nested logit model, because the projections of these two variables on instruments are strongly correlated.

		$n_D=2$				$n_D=4$			
		Time	% reaching	RMSE		Time	% reaching	RMSE	
Model	J	(sec)	global min.	α	σ	(sec)	global min.	α	σ
Logit	24	0.014	100	0.02	–	0.29	100	0.03	–
	48	0.017	100	0.016	–	0.33	100	0.024	–
	96	0.022	100	0.012	–	0.39	100	0.016	–
Nested logit	24	1.4	100	0.53	0.16	9.3	95.5	0.24	0.062
	48	2.1	100	0.41	0.12	14	97.5	0.19	0.05
	96	7.0	100	0.27	0.08	190	99.5	0.17	0.042

Reading notes: 200 simulations for each setting. "Time" is the optimization time in seconds on our desktop computer (Intel® Core™, i3-4160, 3.6 GHz, 8Gb RAM) and using our preferred starting point. "% reaching global min." is the percentage of simulations for which the estimator reaches the global minimum of the value function. We display the average root mean squared errors (RMSE) of $\widehat{\alpha}^d$ and $\widehat{\sigma}^d$ across the different demographic groups d.

Table 5: Computational and statistical aspects of the GMM optimization for the logit and nested logit models.

2.6 Algorithm and simulations in the case of unobserved groups

We consider, for the model with unobserved groups studied in Section 3.4.2, an algorithm similar to the one developed for the model with observed groups. Namely, for each value of the vector of parameters θ, we compute $p(\xi,\theta)$ as the limit of the sequence $p_{n+1} = M_{s,\theta}(p_n)$, through the following steps:

1. Start from initial values for p_0^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, invert Equation (12) to compute ξ_n, using the algorithm of Lee and Seo (2016). Compute the corresponding market shares $s_j^d(p_n^d, \xi_n, \theta)$, for all (j,d).

3. Compute p_{n+1}^d using Equation (14).

4. Repeat steps 2 and 3 until convergence of prices.

We evaluate the computational and statistical properties of our GMM estimator in this case through 200 Monte Carlo simulations. We generate market equilibrium for $T=50$ markets with $J=48$ products offered by 4 firms, each of them offering 12 products. As in the

DGP of the simulations for the main model (see Section 2.1), the product characteristics are $X_{jt} = (1, X_{1jt})$ and the cost shifters are $(X_{jt}, W_{1jt}, W_{2jt}, W_{3jt})$. However, in this model the unobserved characteristics ξ_{jt} are not group-specific. All the variables follow the same distributions as before. We consider 4 groups of consumers that differ only in their price sensitivity parameter, and set $\alpha^d = (-1, -1.5, -2, -2.5)$. The preference for holding a car is set to $\beta_0 = -6$ and the preference for the characteristic X_1 is set to $\beta_1 = 3.5$. For each product, we suppose to observe the maximal price over the 4 consumer groups.

The proportion ϕ_t^d of each group varies across markets and is such that $\phi_t^d \sim \mathcal{U}(0.1, 0.3)$ for $d = 1, 2, 3$ and $\phi_t^4 = 1 - \sum_{d=1}^{3} \phi^d$. Though we could include the proportions ϕ^d in the vector of parameters θ, we assume here that they are known. This is the case when the econometrician knows the groups that are used for price discrimination and their proportion in the population, but does not observe their specific demand (e.g. male/female in our application, if there is price discrimination with respect to gender).

We estimate the model using demand-side moments only. We use an approximation of optimal instruments that follows Reynaert and Verboven (2014). We compute them using a numerical approximation of the Jacobian evaluated at the true parameter values. Finally, we use a tolerance of 10^{-16} in the inner loop where we compute ξ_n and a tolerance of 10^{-12} in the outer loop where we compute $p(\xi, \theta)$.

The simulation results are displayed in Table 6. As with the main model, and consistent with the result of Theorem 2, our algorithm always converges. Perhaps surprisingly, the average time to compute the estimator is much smaller than with the main model (27 seconds versus 214 seconds, see Table 3 above), despite a much larger average number of iterations (148 versus 31). This is because we have to perform the inner loop, where we compute ξ_n for given prices, only once rather than n_D times in the main model. Also, the optimization is faster because the dimension of θ is smaller than with the main model ($\dim(\theta) = 6$ versus 18). The estimator displays good performances and the average discounts are very precisely estimated.

	True	Estimation	
		Mean	Std. dev.
Price sensitivity			
Group 1	-1	-1	0.006
Group 2	-1.5	-1.5	0.01
Group 3	-2	-2	0.032
Group 4	-2.5	-2.5	0.055
Exogenous characteristics			
β_0	-6	-6	0.028
β_1	3.5	3.5	0.008
Mean objective function value		1.6×10^{-5}	
% simulations converging		100	
Number of iterations		148	
Time (sec)		27	
Average discount			
Group 1	0	0	0
Group 2	9.44	9.45	0.16
Group 3	13.65	13.64	0.22
Group 4	16.06	16.05	0.18

Reading notes: Mean and standard deviations of parameters for the converging replications of the 200 simulations. "Std. dev." stands for the standard deviation across simulations. "% simulations converging" is the percentage of simulations for which our algorithm converged. "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).

Table 6: Simulation results for the model with unobserved groups.

As discussed in Section 3.4.2 of the main paper, Theorem 2 does not cover the case where average prices are observed. We nevertheless investigate through simulations whether our algorithm converges in this case. Using one synthetic dataset generated from the DGP described above, we assume that only the sales-weighted average prices are observed instead of maximal prices. We check that our algorithm still solves the system of equation in (ξ, p). For this, we draw 100 different values of $(\alpha^d)_{d=1,...,n_D}$ from $\mathcal{U}(\alpha^d/2, 3\alpha^d/2)$ and compute prices using $p_j^{0,d} = \tilde{p}_j$ as initial prices. We then compare these prices with those we obtain using 50 different sets of initial prices. Specifically, we draw independently across d and j the $p_j^{0,d}$, with $p_j^{0,d} \sim \mathcal{U}[1/2\tilde{p}_j, 2\tilde{p}_j]$. For each of the 100 different values of $(\alpha^d)_{d=1,...,n_D}$, we find that the 50 initial vectors of prices always lead to the same prices. This strongly suggests that Theorem 2 also applies to the case where average prices are observed.

3 Additional material on the application

3.1 Additional details on data

The dataset we use was provided from the association of French automobile manufacturers (CCFA, Comité des Constructeurs Français d'Automobiles). Each year, we observe about one million vehicles registered and their main attributes: brand, model, fuel type, fuel efficiency, car-body style, number of doors, horsepower, cylinder capacity and weight. These characteristics have been complemented with fuel prices to compute the cost of driving (in euros for 100 kilometers). 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. 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 use data from the French national institute of statistics (Insee) to obtain individuals' expected income. There are over 36,000 municipalities in France and the three largest cities, are split into smaller units. The heterogeneity in the median income across municipalities is therefore quite large. We choose 27,000 euros per year as the threshold for income since it roughly corresponds to the median yearly income in France in 2008. Note that we do not observe the owner's gender in our database. Even if we did, it would be hard to use it since the owner and the buyer can be different persons. Also, many couples are likely to buy their car together. Nevertheless, we check below that our results are robust to price discrimination with respect to gender.

Table 7 presents the proportion of each consumer group in the population and the average characteristics of the new cars purchased by these groups. 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 $\geq 27,000$	11.5%	20,911	6.5	6	1221	16.8%	12.9%
A $\in$ [40,59], I $<$ 27,000	16.3%	21,521	6.5	6.1	1231	14.3%	12.7%
A $\in$ [40,59], I $\geq 27,000$	22.3%	21,739	6.8	6.2	1236	14.8%	13.1%
A $\geq$ 60, I $<$ 27,000	20.8%	20,117	6.9	5.9	1194	11.4%	8.9%
A $\geq$ 60, I $\geq 27,000$	13.2%	20,831	7	6	1219	10.9%	10.5%

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

Table 7: Average characteristics of new cars purchased by the different consumer groups.

It is crucial for our approach that buyers cannot lie about their individual characteristics. This rules out any geographical arbitrage, namely that some consumers buy the car in another municipality because discounts are higher. We believe that this assumption is reasonable in this context 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.

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 prevalent in most European countries, with 70% of car dealers being exclusive to one brand in Europe. As discussed above, adding vertical relations between manufacturers and dealers is possible, provided that dealers compete à la Bertrand in the downstream market.

3.2 Correction for null market shares

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 8). Yet, 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.

Group	Frequency of null sale
Age <40, Income $<27{,}000$	11.6%
Age <40, Income $\geq 27,000$	10.3%
Age $\in$ [40,59], Income $<27{,}000$	7.5%
Age $\in$ [40,59], Income $\geq 27,000$	4%
Age ≥ 60, Income $<27{,}000$	7.8%
Age ≥ 60, Income $\geq 27,000$	7.6%

Table 8: Fraction of products with null market shares in the final sample.

The idea of the correction is to consider simple estimators of s_j^d of the form $(n_j^d + c)/N^d$, and fix c such that $\ln((n_j^d + c)/N^d)$ is asymptotically unbiased. The reason we are looking for such a c is that $\ln(s_j^d)$ plays an important role at least in the logit or nested logit models. With an unbiased estimator of $\ln(s_j^d)$, we can estimate consistently and as usually the demand parameters. However, in our framework where individuals choose independently from each others, so that $n_j^d \sim \text{Binomial}(N_d, s_j^d)$, it is well-known that only polynomials of s_j^d of degree at most N_d can be estimated without bias. Our aim is then to find instead an estimator that is asymptotically unbiased at the first order.

For that purpose, we consider an asymptotic approximation where s_j is small but $\lambda_j^d \equiv N_d s_j^d \to \infty$. Let $Z_j^d = (n_j^d - \lambda_j^d)/\sqrt{\lambda_j^d}$. A second-order Taylor expansion of $(n_j^d + c)/N^d$ around s_j^d yields

$$\sqrt{\lambda_j^d}\left[\ln((n_j^d + c)/N^d) - \ln\left(s_j^d\right)\right] = Z_j^d + \frac{c}{\sqrt{\lambda_j^d}} - \frac{s_j^{d2}}{2\widetilde{s}_j^{d2}}\frac{1}{\sqrt{\lambda_j^d}}\left(Z_j^d + \frac{c}{\sqrt{\lambda_j^d}}\right)^2,$$

where $\widetilde{s}_j^d$ is between $(n_j^d + c)/N^d$ and s_j^d. The first order term, Z_j^d, is asymptotically standard normal and thus asymptotically centered. Now, considering the second-order term,

$$\sqrt{\lambda_j^d}\left\{\sqrt{\lambda_j^d}\left[\ln((n_j^d + c)/N^d) - \ln\left(s_j^d\right)\right] - Z_j^d\right\} = c - \frac{s_j^{d2}}{2\widetilde{s}_j^{d2}}\left(Z_j^d + \frac{c}{\sqrt{\lambda_j^d}}\right)^2.$$

Moreover, $s_j^{d2}/\widetilde{s}_j^{d2} \xrightarrow{\mathbb{P}} 1$ and $\left(Z_j^d + \frac{c}{\sqrt{\lambda_j^d}}\right)^2 \xrightarrow{L} \chi_1^2$. Hence,

$$\sqrt{\lambda_j^d}\left\{\sqrt{\lambda_j^d}\left[\ln((n_j^d + c)/N^d) - \ln\left(s_j^d\right)\right] - Z_j^d\right\} \xrightarrow{L} c - \frac{1}{2}\chi_1^2.$$

Choosing $c = 0.5$ therefore ensures that this second-order term is asymptotically centered around 0.

We examine the robustness of the estimation results to the correction of the null shares adopted. We re-estimate the nested logit model using the Laplace transformation of the market share equation used by Gandhi et al. (2013). This correction replaces the market share by:

$$\tilde{s}_j^d = \frac{N^d \widehat{s}_j^d + 1}{N^d + J + 1}.$$

As Table 16 in Section 3.4.1 below suggests, the estimation results are robust to the choice of a correction to deal with products with null market shares. The estimated parameters are very close using the two alternative corrections. As a consequence, subsequent results (not displayed here) on, e.g., discounts, are also close under the the two correction methods.

3.3 Additional results

3.3.1 Differences with the uniform pricing model

We first present additional results on the difference between our model and the standard BLP. We display in Table 9 the relative differences between the main parameters of preferences. More specifically, we compute $(\theta^{unif} - \theta^{disc})/\theta^{disc}$, where θ^{unif} represents the vector of parameters estimated under the standard model and θ^{disc} the vector of parameters estimated with our price discrimination model. We observe that for all the groups except for the pivot group, the price sensitivities are underestimated with the standard BLP model. Conversely, the price sensitivity of the pivot group is overestimated. We also overestimate the importance of within group heterogeneity in price sensitivities. The coefficients of the intercept are underestimated for all the groups except the pivot group for which the difference in the parameters is very small (-1%). The differences in the preference for horsepower are very large and positive, except for the group of old consumers with low income. Errors in the sensitivity to fuel cost can be up to 10.5%. The preferences for car models with three doors display also large differences, whereas we obtain similar estimates for the parameters of preference for station wagon cars.

	Age < 40		Age $\in$ [40,59]		Age ≥ 60	
	I = L	I = H	I = L	I = H	I = L	I = H
Price	-5.4	-4	-7.9	-5.7	-8.5	18.1
Std. dev. (σ^p)			8.2			
Intercept	-13.5	-9.6	-10.7	-7.7	-11.7	-1
Horsepower	44.7	49.8	21	20.5	-0.6	128.3
Fuel cost	10.5	10.1	5.4	3.7	1.6	5
Weight	-14.4	-11.9	-14.9	-11.7	-18.5	12.5
Three doors	-44.1	-364.9	28.4	11.7	-2.5	14.9
Stat. Wagon	-1.5	-0.1	-3.6	-0.7	-3.6	10.7

Reading notes: All values are in % of the estimated parameters under the price discrimination model, $(\theta^{unif} - \theta^{disc})/\theta^{disc}$.

Table 9: Relative difference in estimated parameters between the uniform and the price discrimination models.

Figure 1 displays the distribution of the difference in estimated marginal costs between the two models. We compute here the relative difference $(\hat{c}_j^{unif} - \hat{c}_j^{disc})/\hat{c}_j^{disc}$, where $\hat{c}^{unif}$ stands for the marginal cost of product j implied by the uniform pricing model and $\hat{c}_j^{disc}$ the one implied by the price discrimination model. The costs are always overestimated in the uniform pricing model, with an average relative difference of 9.5%. The relative cost difference even exceeds 18% for 2.9% of the products. These differences stem from the fact that, in the uniform pricing model, the marginal costs are deduced from the difference between the posted prices and the average mark-ups. In contrast, in the price discrimination model, the marginal costs are equal to the difference between the posted prices and the mark-ups of the pivot group. The pivot group mark-ups are higher than the average mark-ups estimated in the standard model, resulting in lower marginal costs. Ultimately, the errors in the estimation of marginal costs translate into errors in counterfactual simulation exercises.

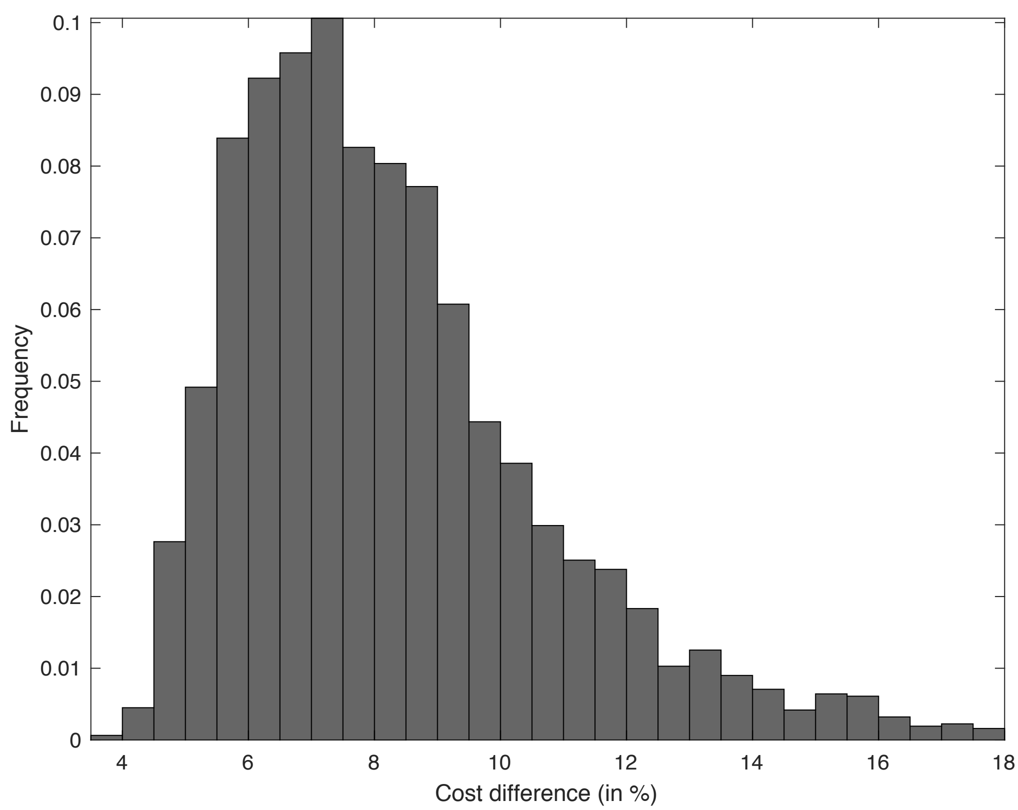

Figure 1: Distribution of the relative difference between estimated marginal costs.

3.3.2 Additional results on discounts

Figure 2 displays the resulting distribution of discounts across products for those having a discount lower than 20% (which represent 98% of the products). The corresponding average discount, averaged by product rather than by consumers, is equal to 9.6%, with some heterogeneity. For 10% of the products, the average discount is smaller than 6.7%, while for the 10% most discounted cars, the rebate is larger than 11.9%, and it even exceeds 32% for 1% of the cars. To understand better the source of this heterogeneity, we regress these discounts on the characteristics of the cars. The results are displayed in Table 10. Discounts increase with the list price and horsepower but decrease with weight and fuel cost. These results reflect both the differences in sales between consumer groups (e.g. products mostly sold to the pivot group tend to have a small average discount) and differences in the pricing strategy. Results with basked-weighted discounts are however similar, showing in particular that it is profitable for firms to offer large discounts for their most expensive cars.

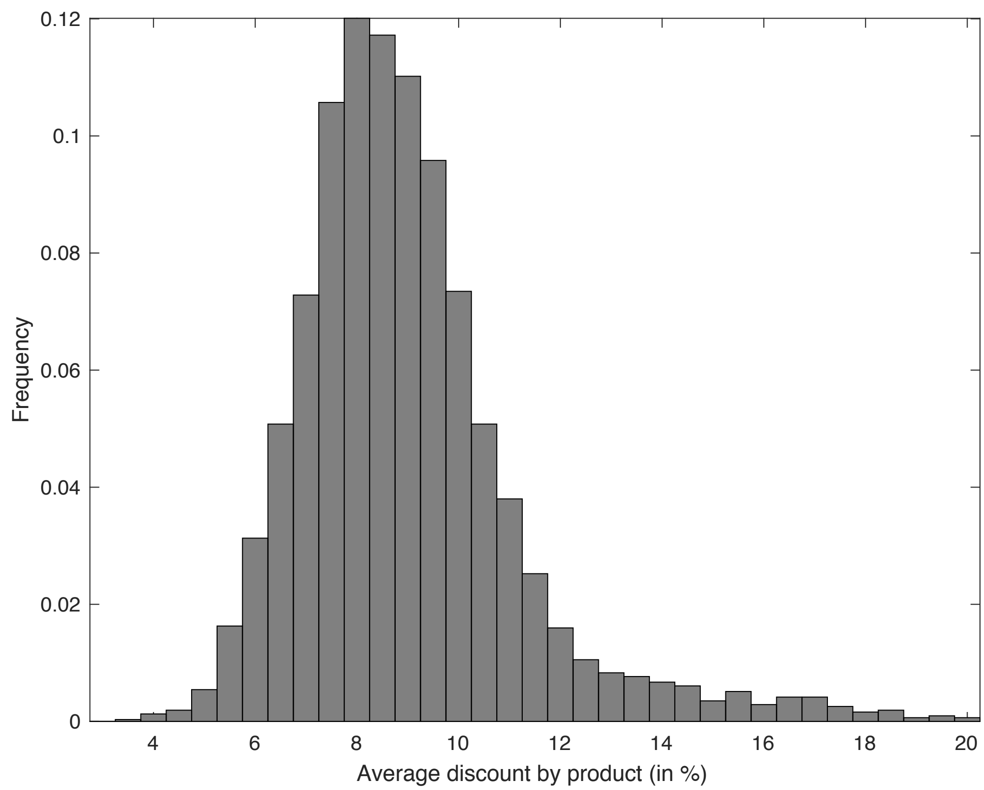

Figure 2: Distribution of estimated discounts across products.

Variable	Parameter	Std-err
Intercept	13.59**	2.4
List price	3.99	7.94
Horsepower	2.63**	0.83
Fuel cost	-2.78**	0.39
Weight	-8.98	13.7
Three doors	0.77	0.47
Station wagon	0.26**	0.09
Executive	-6.21	7.7
Small Family	-2.57*	1.25
Large Family	-4.83	5.98
Small MPV	-1.34**	0.4
Large MPV	-1.65	1.01
Sports	-4.56	4.06
Allroad	-1.88	1.43
R^2	0.65	

Reading notes: the standard-errors are robust to heteroscedasticity and account for the first-step errors in the estimated discounts. Significance levels: †: 10%, *: 5%, **: 1%.

Table 10: Regression of average product discount on cars characteristics.

3.3.3 Importance of third-degree price discrimination for firms and consumers

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.

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 11: Gains and losses from price discrimination by brand.

Results on firms' profits are displayed in Table 11. 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 12, 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 price 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. Consumers in this group 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 $\in$ [40,59], I = L	553	350	371	14,714	14,477	1.64
Age $\in$ [40,59], I = H	292	54	63	18,650	18,599	0.28
Age $\geq$ 60, I = L	483	240	190	15,367	15,200	1.1
Age $\geq$ 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 under price discrimination. "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 12: 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.

3.3.4 Comparison of our results with other evidence

In addition to the comparison with BdF transaction price data, we confront our results to 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 indicated 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 negotiating the car price, the average discount was around 11%. This result is comparable to our average on the whole population (9.6%), and is very close to the average discount we obtain on individuals below 60 years old (11.3%). Interestingly, the latter population also represents 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%".[3] The fourth website associated to the same key words search is a forum asking the question of

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

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%.[4] 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 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 10 above.

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

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

(2002) and Chandra et al. (2017) analyze price discrimination with respect to age and gender on the U.S. car market using dealer 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.

3.4 Robustness checks

3.4.1 Nested logit model

We consider here the nested logit model as an alternative to the random coefficient model. We show that our results are basically confirmed with this specification. The nested logit approach requires to define a segmentation of the market in homogeneous groups of products. Our segmentation, based on the main use of the vehicle, is close to the one of The European New Car Assessment Program one (Euro NCAP). Table 13 displays the 8 segments that we consider and their market shares over the period. Note in particular that sports cars include all convertible cars as well as vehicles with a high ratio horsepower/weight, whereas the small multi-purpose vehicle segment (MPV) includes small vans such as Renault Kangoo. The entire classification is presented in Table 14.

Segment	Market shares (in %)	Segment	Market shares (in %)
Supermini	45.14	Small MPV	17.56
Executive	1.17	Large MPV	1.07
Small Family	17.01	Sports cars	5.11
Large Family	8.67	Allroad	4.77

Table 13: Segments and their market shares.

Table 14: Segmentation of the automobile market.

Group	**Make**	Supermini	Small family	Large family	Executive	Sports car	Small MPV	Large MPV	Allroad/SUV
PSA	*Citroen*	C1, C2, C3, Saxo	Xsara	C5	C6	-	Berlingo, C4, Nemo, Xsara	C8	C-Crosser
	Peugeot	106, 107,1007,206, 207	306, 307, 308	406, 407	607	-	Bipper, Partner	807	4007
Renault	*Renault*	Clio, Modus, Twingo	Megane	Laguna	Vel Satis	-	Kangoo, Megane	Espace	Koleos
	Dacia	Sandero	Logan	-	-	-	-	-	-
B.M.W	*B.M.W*	-	1-Series	3-Series	5, 6, 7-Series	Z4	-	-	X3, X5, X6
	Mini	Mini	-	-	-	-	-	-	-
Chrysler	*Chrysler*	-	-	Sebring	300C, 300M, Crossfire	-	PT Cruiser	Voyager, G.Voyager	-
	Jeep	-	-	-	-	-	-	-	Compass, Cherokee, Commander, G.Cherokee, Wrangler
	Dodge	-	Caliber	Journey	Viper	-	-	-	Durango, Nitro
Daihatsu	*Daihatsu*	Cuore, Sirion, YRV	-	-	-	Copen	-	-	Terios
Daimler	*Mercedes*	-	A-Class	C, CLK-Class	E, CL, R, S, SL, CLS, SLR-Class	SLK-Class	B-Class, Vaneo	Viano	G, GL, GLK, ML-Class
	Smart	Fortwo, Forfour	-	-	-	Coupe, Roadster	-	-	-
Fiat	*Alfa Romeo*	Mito	147	156, 159, GT	166, Brera	GTV, Spider	-	-	-
	Fiat	500, Palio, Panda, Punto, Seicento	Bravo, Stilo	Croma	-	Barchetta	Doblo, Fiorino, Idea, Multipla	Ulysse	Sedici
	Lancia	Y	-	Lybra	Thesis	-	Musa	Phedra	-
Ford	*Ford*	Fiesta, Ka,	-	Mondeo	-	Puma	Focus, Fusion, T.Connect, Tourneo	Galaxy, S-Max	Kuga
	Jaguar	-	-	X-Type	S-Type, XJ, XK	-	-	-	-
	Land Rover	-	-	-	-	-	-	-	Freelander, Defender, Discovery, R.Rover
	Volvo	-	C30, V50	C70, S40, S60, V70	V40, S80	-	-	-	XC60, XC70, XC90
GM Europe	*Chevrolet*	Kalos, Matiz	Aveo, Lacetti, Nubira	Epica, Evanda	-	Corvette	Rezzo, Tacuma	-	Captiva, Tahoe
	Daewoo	Kalos, Matiz	Lanos	-	Evanda	-	Rezzo	-	Korando
	Opel	Corsa	Astra	Insigna, Signum, Vectra	Omega	Tigra, Speedster	Agila, Combo, Meriva, Zafira	-	Antara, Frontera
	Saab	-	-	9-3	9-5	-	-	-	-
Honda	*Honda*	Jazz	Civic	Accord	-	S2000	FR-V, Stream	-	CR-V, HR-V
Hyundai	*Hyundai*	Atos, Getz, I10	Accent, Coupe, I30	Elantra, Sonata	-	-	Matrix	Trajet	Tucson, Santafe, Terracan
	Kia	Picanto, Rio	Cee-d, Cerato	Magentis	-	-	Carens, Soul	Carnival	Sorento, Sportage
Lada	*Lada*	-	111, 112	-	-	-	-	-	Niva
Mazda	*Mazda*	2	3	6	RX8	MX5	5, Premacy	MPV	-
Mitsubishi	*Mitsubishi*	Colt	Lancer	Carisma	-	-	Spacestar	Grandis	Outlander, Pajero
Nissan	*Nissan*	Micra, Note	Almera, Qashqai	Primera	350Z, Maxima-Q	-	Almera	-	X-Trail, Murano, Pathfinder, Patrol, Terrano
Porsche	*Porsche*	-	-	-	911, Boxter, Cayman	-	-	-	Cayenne
Rover	*Rover*	25, Streetwise	45	75	-	-	-	-	-
Ssangyong	*Ssangyong*	-	-	-	-	-	-	Rodius, Stavic	Actyon, Korando, Kyron, Rexton
Subaru	*Subaru*	Justy	Impreza	Legacy	-	-	-	-	Forester, B9Tribeca
Suzuki	*Suzuki*	Alto, Ignis, Splash, Swift, SX4	Liana	-	-	-	Wagon-R	-	G. Vitara, Jimny, Samurai, Vitara
Toyota	*Toyota*	Aygo, IQ, Yaris	Auris	Avensis, Prius	-	Celica, MR	Corolla	Previa	RAV4, L.Cruiser
	Lexus	-	-	IS	GS, LS	-	-	-	RX
VW Group	*Audi*	A2	A3	A4, A5	A6, A8, R8	S3, S4, S6, S8 TT	-	-	Allroad, Q5, Q7
	Seat	Arosa, Ibiza	Cordoba, Leon	Toledo	-	-	Altea	Alhambra	-
	Skoda	Fabia	-	Octavia, Superb	-	-	Roomster	-	-
	Volkswagen	Fox, Lupo, Polo	Eos, Golf, Jetta, New-beetle	Scirocco, Passat	-	Phaeton	Caddy, Touran	Sharan	Tiguan, Touareg

	(1) Uniform model Our correction		(2) Discrimination model Our correction		(3) Discrimination model Gandhi et al. correction	
	Parameter	Std-err	Parameter	Std-err	Parameter	Std-err
Price sensitivity						
Age < 40, I = L	-2.66**	0.061	-2.66**	0.061	-2.5**	0.058
Age < 40, I = H	-2.52**	0.06	-2.51**	0.06	-2.36**	0.058
Age $\in$ [40,59], I = L	-2.21**	0.051	-2.21**	0.052	-2.08**	0.049
Age $\in$ [40,59], I = H	-2.13**	0.053	-2.12**	0.053	-1.98**	0.048
Age ≥ 60, I = L	-1.93**	0.066	-1.95**	0.065	-1.77**	0.06
Age ≥ 60, I = H	-1.77**	0.07	-1.81**	0.067	-1.81**	0.063
Intra-segment correlation						
Age < 40, I = L	0.17**	0.039	0.18**	0.04	0.11**	0.039
Age < 40, I = H	0.29**	0.042	0.3**	0.042	0.22**	0.041
Age $\in$ [40,59], I = L	0.21**	0.034	0.23**	0.034	0.17**	0.033
Age $\in$ [40,59], I = H	0.28**	0.04	0.3**	0.04	0.19**	0.039
Intercept						
Age < 40, I = L	-6.53**	0.256	-7.08**	0.255	-7.5**	0.245
Age < 40, I = H	-6.47**	0.268	-7.06**	0.262	-7.44**	0.252
Age $\in$ [40,59], I = L	-6.96**	0.245	-7.26**	0.243	-7.59**	0.233
Age $\in$ [40,59], I = H	-6.54**	0.265	-6.88**	0.255	-7.41**	0.237
Age ≥ 60, I = L	-7.86**	0.313	-7.94**	0.296	-7.89**	0.287
Age ≥ 60, I = H	-8.2**	0.317	-8.21**	0.301	-8.29**	0.291
Horsepower						
Age < 40, I = L	5.74**	0.221	5.74**	0.219	5.38**	0.207
Age < 40, I = H	5.15**	0.198	5.13**	0.196	4.82**	0.188
Age $\in$ [40,59], I = L	4.25**	0.18	4.25**	0.178	4**	0.169
Age $\in$ [40,59], I = H	3.92**	0.171	3.91**	0.17	3.59**	0.162
Age ≥ 60, I = L	2.88**	0.228	2.95**	0.225	2.57**	0.213
Age ≥ 60, I = H	2.44**	0.228	2.55**	0.227	2.71**	0.205
Fuel cost						
Age < 40, I = L	-6.04**	0.242	-5.99**	0.243	-5.85**	0.226
Age < 40, I = H	-4.86**	0.217	-4.82**	0.216	-4.76**	0.204
Age $\in$ [40,59], I = L	-5.07**	0.2	-4.99**	0.199	-4.94**	0.188
Age $\in$ [40,59], I = H	-4.15**	0.186	-4.1**	0.184	-4.23**	0.183
Age ≥ 60, I = L	-4.15**	0.184	-4.18**	0.183	-3.83**	0.17
Age ≥ 60, I = H	-3.51**	0.18	-3.54**	0.178	-3.46**	0.167
Weight						
Age < 40, I = L	4.13**	0.243	4.08**	0.243	4**	0.231
Age < 40, I = H	4.03**	0.242	3.99**	0.242	3.9**	0.231
Age $\in$ [40,59], I = L	4.18**	0.23	4.12**	0.23	4.03**	0.22
Age $\in$ [40,59], I = H	3.87**	0.231	3.82**	0.232	3.86**	0.217
Age ≥ 60, I = L	3.49**	0.249	3.53**	0.248	3.19**	0.237
Age ≥ 60, I = H	3.52**	0.252	3.58**	0.249	3.49**	0.239
Three doors						
Age < 40, I = L	-0.08	0.178	-0.09	0.176	-0.05	0.173
Age < 40, I = H	-0.25	0.167	-0.25	0.166	-0.21	0.163
Age $\in$ [40,59], I = L	-0.22	0.172	-0.22	0.17	-0.19	0.167
Age $\in$ [40,59], I = H	-0.35*	0.169	-0.35*	0.167	-0.31†	0.168
Age ≥ 60, I = L	-0.6**	0.187	-0.61**	0.186	-0.56**	0.178
Age ≥ 60, I = H	-0.65**	0.184	-0.66**	0.184	-0.64**	0.175
Station wagon						
Age < 40, I = L	-0.6**	0.127	-0.59**	0.126	-0.6**	0.123
Age < 40, I = H	-0.42**	0.121	-0.41**	0.12	-0.43**	0.117
Age $\in$ [40,59], I = L	-0.45**	0.121	-0.44**	0.12	-0.46**	0.117
Age $\in$ [40,59], I = H	-0.47**	0.121	-0.45**	0.12	-0.51**	0.119
Age ≥ 60, I = L	-0.7**	0.126	-0.7**	0.126	-0.65**	0.119
Age ≥ 60, I = H	-0.67**	0.125	-0.68**	0.125	-0.65**	0.119

Reading notes: Standard-errors are robust to heteroscedasticity and computed using the standard formula for GMM. Significance levels: †: 10%, *: 5%, **: 1%.

Table 16: Estimation of parameters: nested logit model with uniform pricing and price discrimination, with our correction and Gandhi et al.'s correction of market shares.

The estimated parameters are generally similar to the ones for the random coefficient models presented in the paper. Note that for two groups (the old purchasers with low and high income), we obtain negative intra-segment correlations whereas this parameter should belong to $[0, 1]$. Thus, we constrain these two parameters to be equal to zero in the estimation, which amounts to consider the logit specification for these two groups of consumers.

We then present the same results as those given in Tables 2 and 4 of the main paper and Figures 1 and 2 and Table 10 of this supplement, but for the nested logit specification. Table 17 first shows that the average price elasticities are very similar to those obtained with the random coefficients model. Under price discrimination, they range from -6.4 to -3.7, almost identical to the range [-6.4, -3.9] that we obtain with the random coefficient model. Here again, older people are much less price sensitive than the other groups. Perhaps surprisingly, on the other hand, high-income individuals below 60 appear to be more price sensitive than the low-income ones, both under price discrimination and uniform pricing. The pivot group is nevertheless still the older, high-income group of consumers.

Group of	Price elasticity		Average mark-up		Average surplus	
consumers	Disc.	Unif.	Disc.	Unif.	Disc.	Unif.
Age $<$ 40, I = L	-5.54	-6.23	22.1	22.3	585	585
Age $<$ 40, I = H	-6.39	-7.24	19.9	21.5	663	662
Age $\in$ [40,59], I = L	-5.52	-5.94	23	21.6	631	631
Age $\in$ [40,59], I = H	-5.8	-6.32	22.5	21.6	998	996
Age $\leq$ 60, I = L	-3.83	-3.87	30.9	23.6	836	848
Age $\leq$ 60, I = H	-3.77	-3.68	31.7	22.9	1245	1275
Average	-5.1	-5.49	25.2	22.3	873	879

Table 17: Comparison of average price elasticities for the nested logit models with uniform pricing and unobserved price discrimination.

We also observe that the model without price discrimination overestimates price elasticities for all groups except the pivot and always overestimates the marginal costs as Figure 3 shows. The average difference in marginal costs is 11.6%, with important heterogeneity. In particular, the difference exceeds 20% for 12.9% of the products.

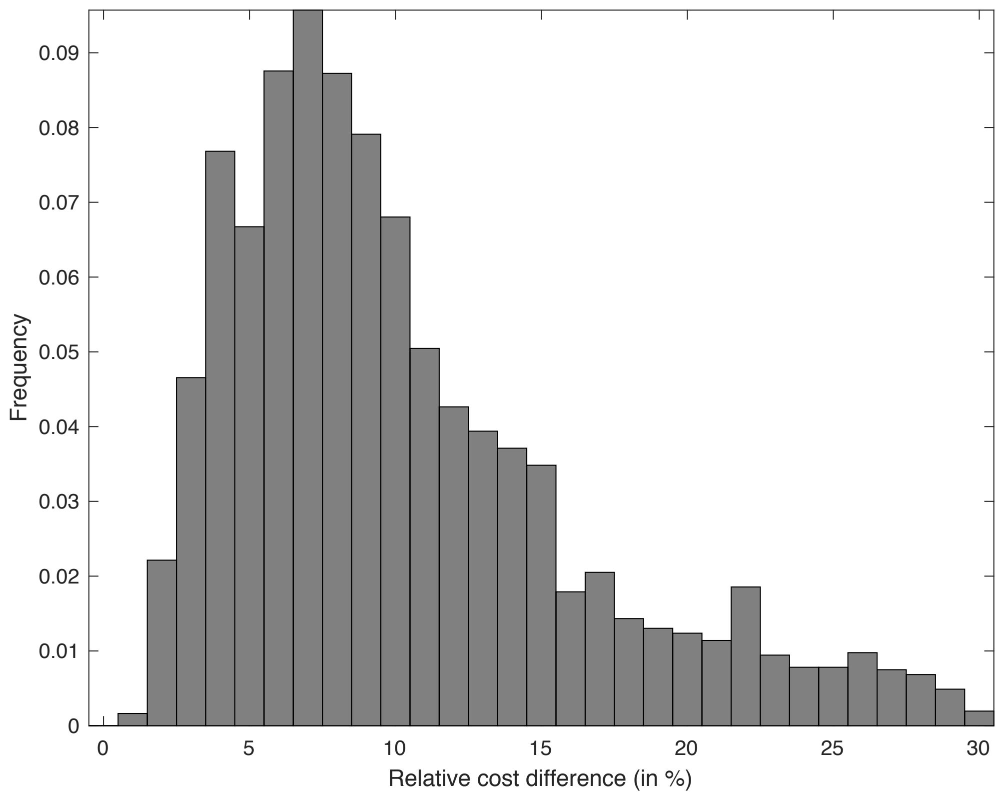

Figure 3: Distribution of the relative difference between estimated costs $\left(\frac{\hat{c}^{unif}-\hat{c}^{disc}}{\hat{c}^{disc}}\right)$ under the nested logit specification.

Turning to the discounts, we obtain again that the youngest purchasers obtain the highest discounts, though such discounts are on average higher for the high income young purchasers. Interestingly, the oldest individuals with low income get a very small discount in average (2.6%) compared to an average of 10.3% with the random coefficient model. The average discount is nevertheless very close to the one obtained with the random coefficient model.

	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	13.96	13.62	2,447	2,473
Age $<$ 40, I = H	14.71	14.9	2691	2,706
Age $\in$ [40,59], I = L	10.76	11.04	1,994	2,005
Age $\in$ [40,59], I = H	11.37	11.61	2,111	2,111
Age $\geq$ 60, I = L	2.59	2.5	458	453
Age $\geq$ 60, I = H	0	0	0	0
Average	8.74	8.79	1,590	1,596

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

Table 18: Average discount by group of consumers for the nested logit model.

Finally, we display in Figure 4 the distribution of average discounts over car models. Both the average (7.2%) and the standard deviation (3.7%) are lower than the figures obtained with the random coefficient model (9.6% and 4.7%, respectively). For 10% of the cars the discount exceeds 17.6% (versus 13.9% unfder the random coefficient model). Finally, the regression of the discounts on cars' characteristics shows, as before, that large fuel costs and heavy vehicles are associated with lower discounts, while horsepower is associated to greater discounts. On the other hand, the list price has a negative rather than positive effect on discounts in this specification, no longer in lines with the results of Kaul et al. (2016).

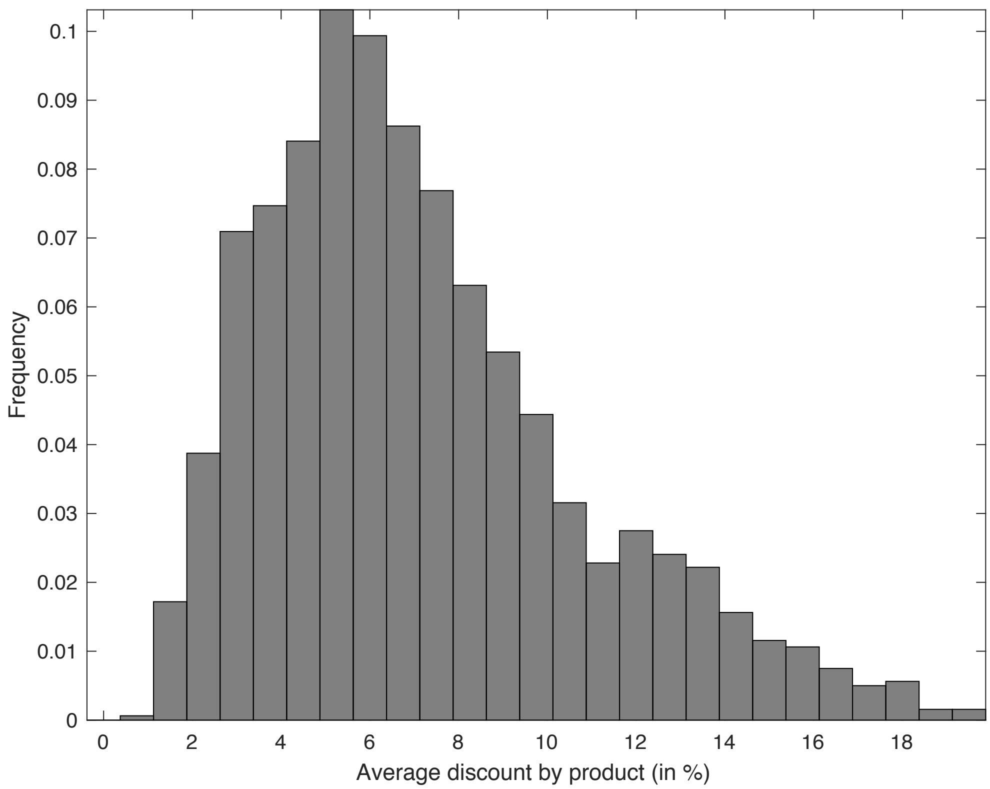

Figure 4: Distribution of estimated discounts for the nested logit model.

Variable	Parameter	Std-err
Intercept	18.06**	2.07
Posted price	-1.36**	0.18
Horsepower	2.3**	0.35
Fuel cost	-2.34**	0.32
Weight	-3.25**	0.45
Three doors	0.71**	0.13
Station wagon	0.22**	0.08
Executive	-3.65**	0.44
Small Family	-3.24**	0.41
Large Family	-4.25**	0.5
Small MPV	-2.68**	0.34
Large MPV	-1.85**	0.3
Sports	-3.22**	0.41
Allroad	-2.16**	0.32
R^2	0.76	

Reading notes: The regression includes segment dummies. Standard-errors are robust to heteroscedasticity and account for the first-step errors in the estimated discounts. Significance levels: †: 10%, *: 5%, **: 1%.

Table 19: Regression of average product discount on car characteristics.

3.4.2 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 20. 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

Reading 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 20: 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 are not aware of any such exhaustive data, we obtained data on rebates for a subset of cars from the consumer price index department of Insee. We have monthly data on temporary promotions for a sample of around 200 cars, 97 model names and 23 brands over the period 2004-2006. We match this dataset with monthly sales data from CCFA (2003-2006) on the brand and model name, the fuel type and annual average price. When available in the promotion dataset, we also use the cylinder capacity, the horsepower and the body style. We match each car model × fuel type from the promotion data to its nearest neighbor in the CCFA dataset using the sum of the squared difference between the following standardized variables: horsepower, cylinder capacity and price. At the end, we obtain an unbalanced panel of 194 cars over 36 months.

We find that the level of promotions is rather small compared to our estimated average discounts. The average sales weighted rebate is €666, which represents 3.2% of the average posted price. This is 3 times lower than our estimated average discount (2,023 euros). We also find that temporary rebates do not display important seasonality patterns. They tend to be the largest in March (€200 more than the average) and in July (€127 more than the average) and the lowest in May (€152 less than the average) and April (€123 less than the average). Finally, we investigate whether promotional activity is driven by low past sales. For that purpose, we regress the temporary promotion on the sales in the past three months using different fixed effects. In Specification (1) of Table 21, we control for the month and the year while in Specification (2) we use the date (month × year) as control. In Specification (3) we control for the month, the year and the model name. Finally in Specification (4) we add car model age dummies as controls. In the first two specifications only the sales three month before appear to be positively correlated with very low levels of statistical significance. However the sales at the past periods do not seem to drive the magnitude of promotions.

	(1)		(2)		(3)		(4)	
	Estimate	Std err	Estimate	Std err	Estimate	Std err	Estimate	Std err
Sales t-1	-0.008	0.121	-0.003	0.122	-0.165	0.103	-0.153	0.101
Sales t-2	0	0.141	-0.02	0.143	-0.043	0.117	-0.038	0.114
Sales t-3	$0.216^{\dagger}$	0.11	0.224^{*}	0.111	0.015	0.093	0.047	0.091
Month, year FE	X							
Month × year FE			X					
Month, year, model FE					X			
Month, year, model, model age FE							X	

Reading notes: Significance levels: †: 10% *: 5% **: 1%. The model age fixed effects in Specification (4) include every age dummy from 1 to 6 and the reference are the models older than 6 year-old. All the specifications estimated using 3,548 observations.

Table 21: Regression of the rebate on past sales and some controls.

Price discrimination with respect to unobserved characteristics Additional price dispersion may come from other demographic characteristics that we do not observe but that are used to 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. One may in particular be worried of our omission of gender, since there has been evidence of price discrimination against women in the U.S. (i.e. they pay more for the same car), though recent results point towards a reduction of those differences over time (see, e.g. Chandra et al., 2017).

We therefore check the robustness of our results to price discrimination with respect to gender, in addition to age and income. For that purpose, we use our estimated parameters of demand and marginal costs and modify the primitives by introducing some discrete unobserved heterogeneity within groups. Specifically, we consider that each group of consumers is equally composed of men and women, and that men have higher price sensitivities than women. We use our estimates for the price sensitivities of women and then calibrate the men price sensitivities so that the average transaction price for men is €250 lower. The value of the average price difference is inspired by Langer (2016), who finds a difference of $250 using transaction price data. All other parameters of preference are assumed to be identical for men and women. We then solve for the new market equilibrium using this new set of parameters.

We then analyze how neglecting the gender in this setting affects the main results (see Table 22). The differences in mark-ups and price elasticities are quite small, in particular the total average mark-up and price elasticity differ by 0.1 to 0.2 points. As expected, the average discount is underestimated, but not by much since price discrimination with respect to gender is very small compared to price discrimination with respect to the six groups we consider. Additionally, the mean absolute relative error in the estimated marginal costs is very small (0.24%) and the price sensitivity parameters we obtain are a convex combination of the men's and women's parameters.

Group	Average discount (%)		Average mark-up (%)		Average price elasticity	
	True	Estimated	True	Estimated	True	Estimated
Age < 40, I $=$ H	13.6	12.2	19.6	18	-6.86	-6.22
Age $\in$ [40,59], I $=$ L	12.7	11.5	20.2	18.7	-6.53	-6.02
Age $\in$ [40,59], I $=$ H	8.9	9.7	18.1	20.3	-5.02	-5.55
Age ≥ 60, I $=$ L	11.6	10.5	22	20.5	-6.02	-5.55
Age ≥ 60, I $=$ H	0.9	0	26.1	28.5	-3.77	-3.93
Average	10.4	9.8	20.5	20.4	-5.8	-5.63

Reading notes: "True" stands for the model with the calibrated parameters for the gender. "Estimated" stands for the estimated parameters in the model that neglects price discrimination on gender.

Table 22: Effect of neglecting the gender on discounts, mark-ups and price elasticities.

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, 2017, 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 2, could be the consequence of heterogeneity in local production factors (e.g., labor or real estate cost).

To check how such heterogeneous local costs affect the robustness of our estimates, we collected estate costs for a sample of 1,395 municipalities through notary data in 2017.[5] We then compute the average estate prices for each demographic group as $\bar{\rho}^d = \sum_m q_m^d \rho_m / \sum_m q_m^d$ with q_m^d the number of car sales in the municipality m from purchasers of group d and ρ_m is the average estate price in municipality m). According to the National Automobile Dealer

[5]Municipalities were drawn without replacement from the database of all French municipalities, with probability proportional to their size.

Association in the U.S. the estate cost represents on average \$961 per car in 2009, which corresponds to €689 using the average exchange rate for 2009.[6] We use this figure to compute the average additional cost of estate by demographic group taking the pivot group as the reference. Average estate prices and incremental costs are displayed in the first two columns of Table 23. We then simulate the new market shares and prices in market equilibrium with heterogeneous marginal costs and estimate a model that neglects the cost heterogeneity. Results, presented in Table 23, show that the effect of neglecting such cost heterogeneity is very small. Whereas the average discounts differ by 1.4 points, average mark-ups and price elasticities are almost identical. Furthermore, neglecting such cost heterogeneity affects mainly the coefficients of the intercept.

	Estate	Additional	Discount (%)		Mark-up (%)		Price elasticity	
Group	price	cost	True	Est.	True	Est.	True	Est.
Age < 40, I $=$ L	2,871	-337	15.3	13.4	17.6	17.2	-6.28	-6.42
Age < 40, I $=$ H	4,622	35	11.8	12	17.8	17.8	-6.19	-6.18
Age $\in$ [40,59], I $=$ L	2,315	-455	14	11.5	19.2	18.7	-5.86	-6.01
Age $\in$ [40,59], I $=$ H	3,843	-131	10.3	9.6	20.4	20.2	-5.5	-5.54
Age ≥ 60, I $=$ L	2,529	-410	12.9	10.5	21.1	20.5	-5.42	-5.55
Age ≥ 60, I $=$ H	4,459	0	0	0	28.2	28.3	-3.94	-3.94
Average	3,247	-257	11.1	9.7	20.6	20.3	-5.54	-5.62

Reading notes: Estate prices are in euros per square meter for houses. "True" corresponds to the values calibrated with heterogeneous costs. "Est." corresponds to the values estimated when neglecting the cost heterogeneity.

Table 23: Effect of neglecting cost differences across consumer groups.

Trade-in and financing The existence of multiple components such as the trade-in of an old car or the purchase of a financing plan occurring at the same time as the car purchase can be source of further unobserved price heterogeneity within consumer groups. Some individuals may use their old car as trade-in, while some others may prefer to keep it or sell it themselves. This leads to an opportunity for the sellers to use the trade-in value to further price discriminate within groups. 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.

Furthermore, to the extent that both operations enter additively in the sellers' profit function, the introduction of the trade-in is equivalent, from the seller's point of view, to a constant marginal cost difference between the transactions with trade-in and those without. This in turns violates Assumption 2 of identical costs, but this time the cost difference is within each group of consumers.

[6] See nada.org/dealershipfinancialprofile/.

To investigate the robustness of our results to neglecting the trade-in of an old car, we generate the equilibrium prices and market shares of a model based on our parameter estimates, but with a fraction of buyers selling their used cars in each of the 6 consumers groups. We use the fraction of transactions that involved a trade-in observed on BdF within each age class.[7] We obtain that 80%, 76% and 83% of transactions involve the trade-in of an old car for respectively the young, middle age and old purchasers. We then set the resale value of the car to €3,000, which is approximately the median trade-in value observed in BdF. This resale value is added to the utility associated to the outside option which implies that the outside option has a greater value for the subgroup of traders than for the subgroup of non-traders. Finally, we set the margin of the seller on the traded car to be €500.

Using this DGP, we estimate our model again, neglecting the trade-in. Table 24 shows that the estimated price parameters are almost identical to the true parameters. The coefficients of the intercept are overestimated. This could be expected, as it is more profitable to sell to consumers with a trade-in car. This translates, when neglecting the trade-in, into larger mean utilities of holding a car. On the other hand, the average mark-ups and price elasticities are very close to their true values (see Table 25).

	Estimated		"True"	
Group	Parameter	Std. error	Parameter	Std. error
Price parameters				
Age $<$ 40, I = L	-4.84**	0.117	-4.83**	0.12
Age $<$ 40, I = H	-4.52**	0.116	-4.52**	0.119
Age $\in$ [40,59], I = L	-4.32**	0.115	-4.32**	0.118
Age $\in$ [40,59], I = H	-3.96**	0.114	-3.96**	0.116
Age $\geq$ 60, I = L	-4.22**	0.131	-4.21**	0.133
Age $\geq$ 60, I = H	-3.06**	0.13	-3.05**	0.134
σ^p	0.99**	0.083	0.98**	0.086
Intercept				
Age $<$ 40, I = L	-6.03**	0.207	-6.24**	0.208
Age $<$ 40, I = H	-6.72**	0.206	-6.92**	0.207
Age $\in$ [40,59], I = L	-6.67**	0.206	-6.85**	0.208
Age $\in$ [40,59], I = H	-6.73**	0.205	-6.9**	0.207
Age $\geq$ 60, I = L	-6.29**	0.225	-6.48**	0.226
Age $\geq$ 60, I = H	-6.17**	0.274	-6.31**	0.282

Reading notes: "Estimated" stands for the estimated parameters in the model that neglects the trade-in of an old car. "True" stands for the model with a fraction of individuals with trade-ins.

Table 24: Effect of neglecting the trade-in part of the transaction on the price parameters and the coefficients of the intercept.

[7] Since we do not have this information by income class, we use the same fraction for each income subcategory.

Group	Average mark-up (%) "True"	Estimated	Average price elasticity "True"	Estimated
Age $<$ 40, I = L	17.8	17.2	-6.26	-6.41
Age $<$ 40, I = H	18.4	17.8	-6.06	-6.2
Age $\in$ [40,59], I = L	19.2	18.7	-5.87	-6
Age $\in$ [40,59], I = H	20.8	20.2	-5.43	-5.54
Age $\geq$ 60, I = L	21.1	20.5	-5.41	-5.55
Age $\geq$ 60, I = H	28.8	28.1	-3.88	-3.97
Average	20.9	20.3	-5.5	-5.63

Reading notes: "True" stands for the model with the calibrated parameters for the model with trade-in. "Estimated" stands for the estimated parameters in the model that neglects the trade-in part the transaction. For the groups with trade-in cars, mark-ups include the margin on the trade-ins while price elasticities are calculated using the price gross of the trade-in resale value.

Table 25: Effect of neglecting the trade-in part of the transaction on average mark-ups and price elasticities.

4 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 (13) is 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, $s_j^d(p, \xi, \theta)/s_0^d(p, \xi, \theta) \leq 1/2$, 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 $p \in \mathcal{C}$, $R_j(p) \leq c_j - 3/[2\alpha^d]$ 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 (13) 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,

$$\begin{aligned}\sum_{k\neq j}\left|\frac{\partial Q_j}{\partial p_k}(p)\right| =& \frac{s_j^d(p,\xi,\theta)(1 - s_0(p,\xi,\theta) - s_j^d(p,\xi,\theta))}{(1 - s_j^d(p,\xi,\theta))^2} \\ <& \frac{(1 - 2s_j^d(p,\xi,\theta))(1 - s_j^d(p,\xi,\theta))}{(1 - s_j^d(p,\xi,\theta))^2} \\ <& \frac{\partial Q_j}{\partial p_j}(p),\end{aligned}$$

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 (13). Therefore, by Assumption 3,

$$\begin{aligned}\widetilde{p}_j(\xi,\theta) &= f_j(p_j^1(\xi,\theta), ..., p_j^{n_D}(\xi,\theta)) \\ &= c_j + f_j\left(\frac{1}{\alpha^1(1 - s_j^1(p,\xi,\theta))}, ..., \frac{1}{\alpha^{n_D}(1 - s_j^{n_D}(p,\xi,\theta))}\right) \\ &= p_j^d(\xi,\theta) - \frac{1}{\alpha^d(1 - s_j^d(p,\xi,\theta))} + f_j\left(\frac{1}{\alpha^1(1 - s_j^1(p,\xi,\theta))}, ..., \frac{1}{\alpha^{n_D}(1 - s_j^{n_D}(p,\xi,\theta))}\right).\end{aligned}$$

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 (13) 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 3, we have, for all ξ,

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

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,

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

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 $\partial q_{j,\theta_0}^{d,d'}/\partial p_\ell^{d'}$ simplify, and we get

$$\frac{\partial q_{j,\theta_0}^{d,d'}}{\partial p_\ell^{d'}}(\widetilde{p})=\frac{s_j^d(\widetilde{p},\xi,\theta_0)(\mathbb{1}\{j=\ell\}-s_\ell(\xi,\theta_0))}{(1-s_j^d(\widetilde{p},\xi,\theta_0))^2},$$

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

$$\sum_{\ell,d''}\left|\frac{\partial q_{j,\theta_0}^{d,d'}}{\partial p_\ell^{d''}}(\widetilde{p})\right| = \frac{2s_j^d(\widetilde{p},\xi,\theta_0)}{(1-s_j^d(\widetilde{p},\xi,\theta_0))^2}[2(1-s_j^d(\widetilde{p},\xi,\theta_0)) - s_0^d(\widetilde{p},\xi,\theta_0)].$$

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

$$s_j^d(\widetilde{p},\xi,\theta_0)(6-2s_0^d(\widetilde{p},\xi,\theta_0)-5s_j^d(\widetilde{p},\xi,\theta_0)) < 1. \tag{18}$$

By Step 1, $-2s_0^d(\widetilde{p},\xi,\theta_0) \leq -4s_j^d(\widetilde{p},\xi,\theta_0)$. Moreover, $x \mapsto x(6-9x)$ reaches its maximum 1 at $x = 1/3$. Because $s_j^d(\widetilde{p},\xi,\theta_0) < 1/3$ by Step 1 again, (18) holds.

Now, given the definition of $q_{j,\theta}^{d,d'}(.)$, 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 $C < 1$ and a neighborhood $\mathcal{V}_1 \subset \Theta$ of θ_0 and $r > 0$ such that for all $(p,p') \in B(p_0,r)^2$,

$$\|M_{s(\xi,\theta),\theta}(p) - M_{s(\xi,\theta),\theta}(p')\| \leq C\|p-p'\|. \tag{19}$$

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

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

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

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 $\|p_n - p(\xi,\theta)\| \leq (r/2)C^n$. 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 (19),

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

Moreover, by the triangular inequality, $\|p_{n+1}-p_0\| \leq (r/2)C^{n+1} + r/2 \leq r$. 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 $p(\xi,\theta)$.

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.

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

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. (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.

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).

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

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.

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.

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.

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.

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.

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.

Lewis, A. S. and Overton, M. L. (2013), 'Nonsmooth optimization via quasi-newton methods', *Mathematical Programming* pp. 1–29.

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

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

Murry, C. and Zhou, Y. (2017), Consumer search and automobile dealer co-location. Working paper.

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.

Nocke, V. and Schutz, N. (2018), 'Multiproduct-firm oligopoly: An aggregative games approach', *Econometrica* **86**(2), 523–557.

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.

Reynaert, M. and Verboven, F. (2014), 'Improving the performance of random coefficients demand models: the role of optimal instruments', *Journal of Econometrics* **179**(1), 83–98.

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.

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

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.