Blurred boundaries: a flexible approach for segmentation applied to the car market

Laura Grigolon*

December 31, 2020

Abstract

Prominent features of differentiated product markets are segmentation and product proliferation blurring the boundaries between segments. I develop a tractable demand model, the Ordered Nested Logit, which allows for asymmetric substitution between segments. I apply the model to the automobile market where segments are ordered from small to luxury. I find that consumers, when substituting outside their vehicle segment, are more likely to switch to a neighboring segment. Accounting for such asymmetric substitution matters when evaluating the impact of new product introduction or the effect of subsidies on fuel-efficient cars.

JEL No. D11, D12, L62, M3

* *Laura Grigolon:* University of Mannheim, Department of Economics and MaCCI, L3, 5-7 Mannheim, Germany. (Email: laura.grigolon@uni-mannheim.de). I am particularly grateful to Frank Verboven for his helpful comments and to Penny Goldberg for her support. I also thank Victor Aguirregabiria, Geert Dhaene, Tobias Klein, Laura Lasio, Chang Liu, Jean-Marc Robin, Nicolas Schutz, Pasquale Schiraldi, Michelle Sovinsky, Johannes Van Biesebroeck, Patrick Van Cayseele, Mike Veall, and participants at various seminars and conferences. I gratefully acknowledge financial support from SSHRC (Insight Development Grant) and the German Research Foundation (DFG) through CRC TR 224.

1 Introduction

In most differentiated product markets, products can be partitioned into segments according to shared common features. Segmentation is not only a descriptive process, but also a practice used by firms to develop targeted marketing strategies and decide the placement of their products. Often, segments can be ordered in a natural way. Cars can be ordered from small (subcompact) to luxury according to price, size, engine performance, comfort and prestige; hotels and restaurants can be ordered on the basis of their ratings (number of stars); retail brands can be ordered in tiers according to quality and price.

In parallel with segmentation, the variety of products has also dramatically increased over time: cars, computers, printers, and smartphones are just a few examples of industries in which product proliferation is visibly prevalent. Broadening the product line has blurred the boundaries between segments, thus decreasing the distance between them: a premium subcompact car can be a potential substitute for a compact car. As a consequence, segments tend to overlap with their neighbors. Correlation between segments has important implications when we want to measure the impact of competitive events, such as the introduction of varieties combining features from different segments. Environmental policies aimed at encouraging the adoption of cleaner cars can also affect sales across segments differently.

I propose a new discrete choice model, the Ordered Nested Logit model, that captures ordered segmentation in differentiated product markets and allows for asymmetric substitution toward proximate neighbors. This model is a new member of the Generalized Extreme Value (GEV) model family developed by McFadden (1978). I construct the Ordered Nested Logit in the context of market level data. The GEV family is consistent with random utility theory and yields a tractable closed-form for choice probabilities. Berry (1994) has provided a framework to estimate two special members of this family with market level data: the Logit and the Nested Logit model. The Ordered Nested Logit model generalizes the Nested Logit model by incorporating an extra parameter that measures the correlation in preferences between neighboring segments: the Nested Logit model implicitly sets such correlation to zero. Hence, the Ordered Nested Logit has the Nested Logit and the Logit as special cases: it can serve as a test for the validity of the constraints imposed by the Nested Logit and, a fortiori, the Logit model. Apart from these two models, only a few other members of the GEV model family have been exploited so far with market level data: notable examples are the principles of differentiation model by Bresnahan et al. (1997), the flexible coefficient multinomial logit by Davis and Schiraldi (2014), and the inverse product differentiation logit model by Fosgerau et al. (2019).

Is asymmetric substitution toward neighboring segments captured by the demand models we currently use? In the Nested Logit model, neighboring segment effects are ruled out by construction. The model requires the stochastic components of utility attached to the segment choice to be independent. Therefore, while preferences can be correlated across products within the same segment (or nest), substitution outside a segment is symmetric to all other segments. In contrast, the random coefficients logit model by Berry et al. (1995) has the potential to generate more flexible substitution patterns, where products tend to be closer substitutes as they share similar observed continuous characteristics. Grigolon and Verboven (2014) simulate the effect of a joint 1% price increase of all cars in a given segment and show that the random coefficients logit model yields more intense substitution toward neighboring segments. But flexibility is achieved only if the parameters of the models, which determine how the random coefficients govern substitution patterns, are correctly and precisely identified. Berry and Haile (2014) clarify that the identification of those parameters poses a distinct empirical problem from price endogeneity and provide general results for identification in differentiated product markets, showing that those parameters are identified by standard exclusion restrictions. Reynaert and Verboven (2014) and Gandhi and Houde (2019) study practical instrumentation strategies for empirical work. With market share data, we can only use the mean choice probabilities (the market shares) as moments that identify the parameters measuring heterogeneity. Good instruments would mimic the ideal experiment of random variation in the characteristics or number of products to identify the response in terms of market shares; in practice, identification can prove difficult in complex set-ups, especially with four or more random coefficients, as documented by Reynaert and Verboven (2014). The Ordered Nested Logit relies on the same variation in the data to identify the nesting and neighboring nesting parameters; by assuming and estimating a correlation structure based on the proximity of product groups, the model can be a parsimonious alternative to the Random Coefficient Logit model. Finally, the Random Coefficients Logit model does not produce a closed-form for the choice probabilities. Earlier work documented sources of numerical issues (e.g. Knittel and Metaxoglou, 2008) and recent articles (Kalouptsidi, 2012, Dubé et al., 2012, Lee and Seo, 2015) have proposed methods that improve the performance of random coefficients models. Berry et al. (2004b) derive the properties of the nested fixed point estimator and show that the simulation error in the approximation of the market shares is bounded only if the number of draws rises with the sample size.[1] Avoiding the simulation of market shares altogether may alleviate some of those difficulties.

First, I formally derive the Ordered Nested Logit model and relate it to commonly used discrete choice models, focusing on the comparison of substitution patterns implied by the Nested Logit and the Ordered Nested Logit model. Using simulated data, I document the flexibility of the Ordered Nested Logit in producing asymmetric substitution patterns and handling misallocation of products into nests. I also provide guidance on the design of the nesting structure.

Next, I apply the Ordered Nested Logit model to a unique dataset on the car market covering three major European countries between 1998 and 2011. The process of purchasing a car is modelled as a nested sequence, with the choice between the segments (including the

[1] Brunner et al. (2017) document that the simulation errors in the approximation of the market shares can generate failure to convergence to a local minimum, numerical instabilities and unreliable identification of the parameters governing the substitution patterns.

outside good segment) at the upper node level and the choice of the specific vehicle at the lower node. I estimate the degree of correlation in consumer preferences both within each segment, as in the Nested Logit model, and in neighboring segments. The demand estimates of the Ordered Nested Logit model clearly indicate a rejection of the simpler Nested Logit model: correlation in car choices is present not only within a segment, but also between neighboring segments.

The demand estimates have striking implications for the substitution patterns. While the Nested Logit model yields symmetric and very low substitution toward other segments, the Ordered Nested Logit model shows a large substitution effect to the neighboring segments. I look at the impact of the introduction of premium subcompact cars on sales by vehicle class. The Nested Logit model predicts that only sales of other subcompact cars are affected by the introduction of those vehicles, while the Ordered Nested Logit model shows, more plausibly, that the segment immediately above (compact cars) is affected as well. Next, I simulate a subsidy to clean vehicles: such policy is clearly asymmetric because it favours mainly subcompact and compact cars. The Nested Logit model predicts, again, that sales of non-subsidized cars do not notably change after the policy, while the Ordered Nested Logit model shows a sizeable decrease in sales of the upper segments, especially the standard segment which has cars that are just above the eligibility threshold.[2]

1.1 Related Literature

The model I propose takes inspiration from the literature on Nested Logit models (Williams, 1977; Daly and Zachary, 1977; McFadden, 1978) and from the Ordered Generalized Extreme Value (OGEV) model by Small (1987). The OGEV model was the first closed-form GEV

[2] Green subsidies are usually temporary and naturally call for a dynamic approach to model consumers' decisions over time, which can be implemented only with additional information on the secondary market and the patterns of ownership (see Schiraldi, 2011). The Ordered Nested Logit model could also be useful in a dynamic framework, as it avoids the need of simulating the market share integral thus potentially alleviating some numerical difficulties.

model to allow for taste correlation between neighboring products. However, it has been developed in settings where a limited number of alternatives have a natural order so that correlation in unobserved utility between two alternatives depends on their proximity in the ordering. With market level data, such as a dataset on the car market, ordering hundreds of products in each market would prove impossible, while ordering groups of products, the segments, is a sensible strategy to obtain a tractable model and flexible substitution patterns. Several other authors have tried to relax the hierarchical structure imposed by the Nested Logit, especially in the transportation literature; see Chu (1989); Vovsha (1997); Ben-Akiva and Bierlaire (1999). The most flexible model in this literature is the generalized Nested Logit model by Wen and Koppelman (2001), where an alternative can be a member of more than one nest to varying degrees. Bresnahan et al. (1997) develop a principles of differentiation model which is an example of a closed-form GEV model applied to market-level data. Grigolon and Verboven (2014) show in their empirical application to the car market that sources of market segmentation may not be captured by the continuously measured characteristics in the random coefficient logit model. They do so by adding a nested logit structure to BLP's random coefficients model. This paper builds on that work by considering a new tractable model from the GEV family to capture an additional feature of such heterogeneity: ordering in market segmentation. The Ordered Nested Logit accounts explicitly for this form of vertical differentiation by estimating a parameter that measures such correlation in preferences between neighboring segments. Davis and Schiraldi (2014) propose an analytic model capable of generating flexible substitution patterns combining elements of the Paired Combinatorial Logit and the Cross Nested Logit. Their model is fully flexible as it can potentially avoid the restrictiveness of Logit and Nested Logit models allowing the second cross derivatives between any pair of goods to be non-zero. In the Ordered Nested Logit, the second cross derivative between pairs of products belonging to different nests and neighboring segments is instead zero. In practice, estimating all the parameters in the model proposed by Davis and Schiraldi (2014) would prove unfeasible, as it is impossible to estimate all alternative specific constants in a Logit model. The authors impose a form of structure to avoid proliferation of parameters that arises with a large number of products: they parametrize the correlation parameters to be function of the distance between products, following Pinkse et al. (2002). In the same spirit, the Ordered Nested Logit explicitly models the idea of varying degrees of distance between nests.[3] Finally, Fosgerau et al. (2019) propose an empirical model specified in terms of inverse demand: their model extends the nested logit by allowing segmentation to be non-hierarchical, while maintaining tractability.

The remainder of the article is organized as follows. Section 2 puts forward the Ordered Nested Logit model. A study using simulated data illustrates the flexibility of the model. Section 3 describes the application dataset and the econometric procedure, including the identification issues. Section 4 provides the empirical results and the implied price elasticities. Section 5 presents the policy counterfactuals. Section 6 concludes.

2 Modelling correlation between neighboring segments

The GEV family Demand is modelled within the discrete choice framework. Consider T markets, $t = 1, .., T$, with L_t potential consumers in each market. Markets are assumed to be independent, so I suppress the market subscript t to simplify notation. Each consumer i chooses a specific product j, $j = 1, \dots J$. Consumer i's indirect utility is:

$$\begin{aligned} U_{ij} &= x_j\beta - \alpha p_j + \xi_j + \varepsilon_{ij} \\ &\equiv \delta_j + \varepsilon_{ij}, \end{aligned}$$

[3] There is a long tradition of estimating demand in product space assuming weak separability across product groups when defining consumer preferences, which reduces the dimensionality of the problem but imposes mutually exclusive product groupings. Blundell and Robin (2000) break weak separability by developing the concept of latent separability, in which products from different groups can interact through sub-utilities stemming from latent activities. While firmly in the discrete choice literature in characteristics space, my work echoes Blundell and Robin (2000) in its attempt of breaking the rigidity of nesting structure.

where x_j is a vector of observed product characteristics, p_j is price, and ξ_j is the unobserved product characteristic. Following Berry (1994), I decompose U_{ij} into two terms: δ_j, the mean utility term common to all consumers, and ε_{ij}, the utility term specific to each consumer.

The consumer-specific error term ε_{ij} is an individual realization of the random variable ε. The distribution of ε determines the shape of demand and the implied substitution patterns. McFadden (1978) has proposed a family of random utility models, the Generalized Extreme Value (GEV) family, in which those patterns can be modeled in different ways according to the specific behavioral circumstances. A GEV model is derived from a generating function $G = G(e^{\delta_0,...,\delta_J})$, a differentiable function defined on $\mathbb{R}^J_+$: (i) which is non-negative; (ii) which is homogeneous of degree 1; (iii) which tends toward $+\infty$ when any of its arguments tend toward $+\infty$; (iv) whose n^{th} cross-partial derivatives with respect to n distinct e^{δ_j} are non-negative for odd n and non-positive for even n.

According to the GEV postulate, the choice probability of buying product j is:

$$s_j = \frac{e^{\delta_j} \cdot G_j(e^{\delta_0,...,\delta_J})}{G(e^{\delta_0,...,\delta_J})}, \tag{1}$$

where G_j is the partial derivative of G with respect to e^{δ_j}.

The Ordered Nested Logit model Assume that the set of products j is partitioned into N mutually exclusive and collectively exhaustive nests. In addition, assume that those N nests are naturally ordered, with n increasing along its natural ordering: $n = 1, ..., N$. The ordering may correspond to an increasing value of important characteristics such as price and quality. I define the Ordered Nested Generalized Extreme Value model (in short, Ordered Nested Logit) as the model resulting from the following G function within the GEV class:

$$G = \sum_{r=1}^{N+M} \left(\sum_{n \in B_r} w_{r-n} \left(\sum_{j \in S_n} \exp\left(\frac{\delta_j}{1-\sigma_n}\right)\right)^{\frac{1-\sigma_n}{1-\rho_r}} \right)^{1-\rho_r}, \tag{2}$$

where n is the nest to which the products belongs; M is a positive integer; $w_m \geq 0$ and $\sum_m w_m > 0$. The weight w_m is the allocation weight of a nest into a "neighborhood of nest", the set of nests B. The parameters σ_n and ρ_r are constants satisfying $0 \leq \rho_r \leq \sigma_n < 1$. Those conditions are sufficient to satisfy the four properties of GEV generating functions; Appendix A provides the proof for each condition.[4] Finally, define the set of N nests as $B_r = \{S_n \in \{S_1, ..., S_N\} | r - M \leq n \leq r\}$.[5] Each of the $(N + M)$ sets contains up to $M + 1$ contiguous nests (and *all* the alternatives in those nest).

Consider a simple example with five nests S, ten alternatives and $M = 2$:

$$j = \underbrace{1; 3}_{S_1}; \underbrace{2, 5}_{S_2}; \underbrace{4, 6}_{S_3}; \underbrace{7; 9}_{S_4}; \underbrace{8; 10}_{S_5}$$

Alternatives within a nest need not to be ordered, but nests are. In our example the sets of nests are: $B_1 = \{S_1\}$, $B_2 = \{S_1, S_2\}$, $B_3 = \{S_1, S_2, S_3\}$, $B_4 = \{S_2, S_3, S_4\}$, $B_5 = \{S_3, S_4, S_5\}$, $B_6 = \{S_4, S_5\}, B_7 = \{S_5\}$. where each nest S_n belongs to $M + 1$ different sets. The degree of proximity between neighboring nests can be modelled flexibly as each set of nests can have its own parameter ρ_r. The shape of the demand function crucially depends on the two parameters, σ_n and ρ_r, that parameterize the cumulative distribution of the error term ε. The first one, σ_n, corresponds to a pattern of dependence in ε across products sharing the same nest (as in the Nested Logit). The second one, ρ_r, corresponds to a pattern of dependence in ε across products belonging to neighboring nests. Consider, for example, the effect of a price shock to alternative one belonging to segment S_1. The dependence in ε

[4] Small (1987), Vovsha (1997), and Bresnahan et al. (1997) impose the condition that the sum of the weights is equal to one. Those weights are then interpreted as allocation parameters of nests to sets of nests. As long as weights are non-negative and at least one of the weights is strictly positive, the generating function G belongs to the GEV family, as shown in Bierlaire (2006) and in Appendix A. The condition that the sum of the weights equals one is empirically useful to ensure that the estimation of the model is feasible: I make use of it in the empirical application. Using simulated data, I expand on the role of the weights and show that possible misspecifications in weights do not seem to affect the parameter estimates of interest and the resulting substitution patterns (see Section 2.3).

[5] Although B_r was defined as a nest of nest indices, I will sometimes write, with a slight abuse of notation, $S_n \in B_r$.

measured by σ_n determines that a share of consumers, who had initially chosen alternative one in S_1, will switch to another alternative in segment S_1. The dependence in ε measured by ρ_r determines that a share of consumers will switch to the neighboring segments: in our example, with $M = 2$, the neighboring segments are S_2 and S_3.

If the random components follow the G function in Equation (2), by the GEV postulate in Equation (1) the choice probability of buying product $j \in S_n$ is:

$$s_j = \sum_{r=n}^{n+M} s(j|n) \cdot s(n|B_r) \cdot s(B_r), \tag{3}$$

where:

$$\begin{aligned}
s(j|n) &= \frac{\exp\left(\frac{\delta_j}{1-\sigma_n}\right)}{Z_n}, \\
s(n|B_r) &= \frac{w_{r-n} Z_n^{\frac{1-\sigma_n}{1-\rho_r}}}{\exp\left(I_r\right)}, \\
s(B_r) &= \frac{\exp\left((1-\rho_r)I_r\right)}{\sum_{s=1}^{N+M} \exp\left((1-\rho_s)I_s\right)}, \\
Z_n &= \sum_{j \in S_n} \exp\left(\frac{\delta_j}{1-\sigma_n}\right), \\
I_r &= \ln \sum_{n \in B_r} w_{r-n} Z_n^{\frac{1-\sigma_n}{1-\rho_r}}.
\end{aligned}$$

Equation (3) involves algebraic rearrangements from the choice probabilities expressed according to the GEV postulate in Equation (1): Appendix A provides a derivation of this expression.

The Nested Logit model To clarify the logic of the modeling strategy for the Ordered Nested Logit, consider the G function associated with a traditional specification, the Nested Logit model, in which the ordering of the segments is not explicitly modeled. The model incorporates potential correlation among products only within a nest (segment), not between nests. The J alternatives are grouped into N nests labeled $S_0, ..., S_N$. The G function takes the form:

$$G = \sum_{n=1}^{N} \left(\sum_{j \in S_n} e^{\frac{\delta_j}{1-\sigma_n}} \right)^{1-\sigma_n}, \tag{4}$$

where σ_n captures correlation among products within the same nest. Consistency with random utility maximization requires σ_n to lie in the unit interval. In the Nested Logit model only alternatives belonging to the same nest have stochastic terms that are correlated, and such correlation is directly related to σ_n. The generating function G of the Ordered Nested model in Equation (2) reduces to the Nested Logit model in Equation (4) if $\rho_r = 0$. In addition, if $\sigma_n = 0$ for all nests, the model becomes the standard Logit in which each element of ε is independent.

Following Berry (1994), I can write the choice probability of a product j for the Nested Logit model as follows:

$$s_j = s(j|n) \cdot s(n) \tag{5}$$

where:

$$\begin{aligned} s(j|n) &= \frac{\exp\left(\frac{\delta_j}{1-\sigma_n}\right)}{Z_n}, \\ s(n) &= \frac{Z_n^{1-\sigma_n}}{\exp\left(I_n\right)}, \\ Z_n &= \sum_{j \in S_n} \exp\left(\frac{\delta_j}{1-\sigma_n}\right), \\ I_n &= \ln \sum_{n=1}^{N} Z_n^{1-\sigma_n}. \end{aligned}$$

Compare the market shares of the Ordered Nested Logit model in Equation (3) with the market shares of the one-level Nested Logit model in Equation (5): similarly to the one-level Nested Logit model, in the Ordered Nested Logit model, a change in the attributes of alternative j (say, a price increase) will determine that s_j is diminished by the presence of attractive alternatives within a nest n. Differently from the Nested Logit model, in the Ordered Nested Logit s_j is also diminished by the presence of attractive alternatives in neighboring nests B_r. Ceteris paribus, this effect is increasing in ρ_r: one may expect that if the values of σ_n and ρ_r are sufficiently high, products belonging to the same segment or to neighboring segments will be closer substitutes compared to products belonging to further segments. The substitution patterns will be more precisely illustrated in the next paragraph.

2.1 Substitution patterns

The flexibility introduced by the Ordered Nested Logit model is easily assessed by looking at the matrix of own- and cross- price elasticities, as presented in Corollary 1, page 38 in Davis and Schiraldi (2014):

$$\frac{\partial \ln s_i}{\partial \ln p_j} = \left(I(j = i) + \frac{e^{\delta_j} G_{ij}}{G_i} - s_j \right) (-\alpha p_j).$$

Table 1 compares the substitution patterns implied by the Nested Logit and the Ordered Nested Logit model. The elasticities of the Ordered Nested Logit reduce to the ones of the Nested Logit if $\rho_r = 0$. More generally, the elasticities of the Ordered Nested Logit model depend not only on the conditional probability of choosing alternative i in nest n, but also on the conditional probability of choosing nest n in a set of nests B_r. In Appendix A, I provide a derivation of the expressions for the first and second derivatives (G_i and G_{ij}).

I focus on the cross-price elasticities. In the Nested Logit, when two products belong to different nests, we see that a price reduction reduces the probabilities for all the other alternatives by the same percentage, a pattern of substitution that is a manifestation of the Independence from Irrelevant Alternatives (IIA) property (case b and c in Table 1). In the Ordered Nested Logit, if two products belong to different nests but to the same set of nests B_r (case b in the Table), proportionate shifting does not hold. Another way to look at this is to focus on the ratio of probabilities between alternatives. In the Nested Logit model the second cross-partial derivative, G_{ij}, is equal to zero for product i in a different nest than product j. In the Ordered Nested Logit, $G_{ij} \neq 0$ for j in a different nest than j but in the same set of nests B_r. In both the Nested Logit and the Ordered Nested Logit models, the IIA property holds for two products in the same nest, so the ratio of probabilities of alternative i and j is independent of the attributes or existence of the other alternatives.[6] The Nested Logit model relaxes the IIA property across nests only to a certain extent: the ratio of probabilities of products in different nests will only depend on the attributes of products in nests that contain i and j, but not on all other nests: Train (2009) describes this property as *'independence from irrelevant nests'*. In contrast, in the Ordered Nested Logit this form of IIA is weakened as the ratio of probabilities of two products will depend not only on the attributes and existence of the alternatives in the two nests, but also all the alternatives in the neighboring nests.

[6] Note that in the Random Coefficients Logit model the IIA property remains present at individual level, as the individual-level choice probabilities are a multinonomial logit.

Table 1: Segment Elasticities: Ordered Nested Logit vs Nested Logit

Nested Logit	Ordered Nested Logit
	GEV Generating function G
$G=\sum_{n=0}^{N}\left(\sum_{j\in S_n}\exp\left(\frac{\delta_j}{1-\sigma_n}\right)\right)^{1-\sigma_n}$	$G=\sum_{r=1}^{N+M}\left(\sum_{n\in B_r}w_{r-n}\left(\sum_{j\in S_n}\exp\left(\frac{\delta_j}{1-\sigma_n}\right)\right)^{\frac{1-\sigma_n}{1-\rho_r}}\right)^{1-\rho_r}$
	Own Elasticities $\frac{\partial\ln s_i}{\partial\ln p_i}$
$\left(\frac{1}{1-\sigma_n}-\frac{\sigma_n}{1-\sigma_n}\cdot s(i\vert n)-s_i\right)(-\alpha p_i)$	$\left(\frac{1}{1-\sigma_n}-\sum_{r=n}^{n+M}\left(\frac{\rho_r}{1-\rho_r}s(i\vert n)s(n\vert B_r)-\frac{\rho_r-\sigma_n}{(1-\rho_r)\cdot(1-\sigma_n)}s(i\vert n)\right)-s_i\right)(-\alpha p_i)$
	Cross elasticities $\frac{\partial\ln s_i}{\partial\ln p_j}$
	a. same segment $i,j\in S_n$
$\left(\frac{\sigma_n}{1-\sigma_n}s(j\vert n)+s_j\right)\alpha p_j$	$\left(\left(\sum_{r=n}^{n+M}\frac{\rho_r}{1-\rho_r}s(j\vert n)\cdot s(n\vert B_r)-\frac{\rho_r-\sigma_n}{(1-\rho_r)\cdot(1-\sigma_n)}s(j\vert n)\right)+s_j\right)\alpha p_j$
	b. different nest, same set of nests $i,j\notin S_n$; $i,j\in B_r$
$\alpha s_j p_j$	$\left(\sum_{r=n}^{n+M}\left(\frac{\rho_r}{1-\rho_r}s(j\vert n)\cdot s(n\vert B_r)\right)+s_j\right)\alpha p_j$
	c. different nest, different set of nests $i,j\notin S_n$; $i,j\notin B_r$
$\alpha s_j p_j$	$\alpha s_j p_j$

The table compares the substitution patterns generated in the Nested Logit and the Ordered Nested Logit generated according to the GEV generating functions G (first row).

2.2 The Ordered Nested Logit versus other GEV models

The OGEV model The OGEV model derived by Small (1987) is based on the following G function (see Definition 1 in Small, 1987):

$$G = \sum_{r=1}^{J+M} \left(\sum_{j \in B_r} w_{r-j} \exp \left(\frac{\delta_j}{1-\rho_r} \right) \right)^{1-\rho_r},$$

where M is a positive integer; the weights w_m are overlapping parameters for alternatives; the parameter ρ_r is a measure of correlation between alternatives, rather than nests as in our model, and B_r is a set of alternatives, not nests.

The OGEV model responds to different modelling needs with respect to the Ordered Nested Logit: the OGEV is designed for situations where individual-level data are available, with a limited number of alternatives can be naturally ordered. Instead, the Ordered Nested Logit model is designed for situations in which numerous alternatives are present. Groups of those alternative can be naturally ordered, while alternatives in each group do not need to be ordered.[7]

The Generalized Nested Logit model The Ordered Nested Logit model can be viewed as a special case of the Generalized Nested Logit (GNL) by Wen and Koppelman (2001). Recall the generating function of the GNL model:

$$G = \sum_{k=1}^{K} \left(\sum_{j \in S_k} \left(\alpha_{jk} \exp \left(\delta_j \right) \right)^{\frac{1}{1-\rho_k}} \right)^{1-\rho_k},$$

where S_k is the set of all alternatives included in nest k, α_{jk} is the allocation parameter which is the portion of alternative j assigned to nest k.

The Ordered Nested Logit model can be written as a special case of the GNL if (i)

[7] The Ordered Nested Logit model also differs with respect to the nested version of the OGEV model described by Small (1994) and Bhat (1998), which is similar to a nested logit except that at the lower node the alternatives (not segments) are grouped according to the OGEV model rather than the standard logit.

alternatives are positioned in the nest to which they originally belong, so $S_n = \{j \in S_n\}$; (ii) all the alternatives in neighboring nests are put together in a nest B_r formed by combinations of nests in ordered position: $B_r = \{S_n \in \{1, ..., N\} | r - M \leq n \leq r\}$; (iii) the weights or allocation parameters α_{jk} are equal for all alternatives in nest B_r. Hence, weights are associated to the nest B_r rather than its alternatives.

Summary The Ordered Nested Logit model generalizes the Nested Logit model by capturing asymmetric interactions across nests. It differs from the OGEV model by Small (1987) because it is designed to capture asymmetric interactions across *nests,* not across *alternatives.* Hence, it does not impose an order across alternatives, but across groups of alternatives (nests). The Generalized Nested Logit model by Wen and Koppelman (2001) is the most general instance of a GEV model, but the complexity requires normalization assumptions to identify the parameters and constraints to make the estimation feasible: see Bierlaire (2006).[8] The Ordered Nested Logit includes an ordered nesting structure motivated by features commonly found in differentiated product markets: those restrictions render the model easy to handle for estimation while retaining flexibility.

2.3 Simulated Data

The Ordered Nested Logit model is appealing for its closed form formulation and its ability to capture more complicated substitution patterns than the Nested Logit. As a first step to test the benefits of the Ordered Nested Logit model, I consider two main sets of Monte Carlo experiments. In the first experiment, I generate data according to the Ordered Nested Logit model and fit the Nested Logit. In the second experiment, I generate data according to a Random Coefficient Logit and fit the Ordered Nested Logit model. All the details on

[8] As the Ordered Nested Logit is a special case of the Generalized Nested Logit or the Cross-Nested Logit model proposed by Bierlaire (2006), one can also follow the proof offered in Bierlaire (2006) to verify the properties of the generating function G.

the experiments and the tables are reported in Appendix B.

The experiments have three objectives: (i) to assess the flexibility of the Ordered Nested Logit in approximating the correct substitution patterns under various models; (ii) to evaluate the consequences of product misallocation to nests; (iii) to provide guidance on the design of the nesting structure, in particular on the number of nests (N), neighboring nests (M), and the role of weights. A summary of the results follows.

(i) Substitution patterns When comparing the segment-level elasticities between the correctly specified Ordered Nested Logit and the Nested Logit model in Table B.2, we clearly see that the Nested Logit model delivers symmetric substitution patterns outside a segment: asymmetries are ruled out by construction. When comparing the substitution patterns delivered by the correctly specified Random Coefficient Logit versus the misspecified Ordered Logit in Table B.5, the Ordered Nested Logit approximates well the asymmetric substitution pattern generated by the Random Coefficient Logit model even if the model is misspecified, with a slight overestimation of substitution toward the most immediate neighbor and underestimation toward the distant ones.

(ii) Product misallocation Both the Nested Logit and the Ordered Nested Logit models require partitioning the products into nests: in Table B.3, I show that the Ordered Nested Logit model is less sensitive to misclassification of products into nests with respect to the Nested Logit. The bias in the own- and cross-price elasticities resulting from a misspecified Ordered Nested Logit is always smaller than the one resulting from a misspecified Nested Logit model.

(iii) Design of the nesting structure I provide guidance on the nesting structure, with a focus on (i) the choice of the number of nests (N); (ii) the choice of the number of neighboring nests (M); (iii) the nesting weights. The results can be summarized as follows.

(i) If the researcher specifies nests too narrowly, both the nesting parameter σ and the neighboring nesting parameter ρ present an upward bias, so much so that the neighboring nesting parameter may be even greater than nesting parameters ($\rho > \sigma$), which is inconsistent with random utility maximization.

(ii) By introducing the parameter M governing which nests are correlated, the Ordered Nested Logit model gives another dimension of choice to the researcher. If only the immediately proximate neighbor matters ($M = 1$), while the researcher allows for $M = 2$, the nesting parameter σ presents a downward bias, and the neighboring nest parameter ρ an upward bias. This may result in $\rho > \sigma$. The pattern is reversed if the correct DGP suggests more flexibility in terms of number of neighbors ($M = 2$), while the researcher uses $M = 1$: the estimated nesting parameter presents an upward bias and the neighboring nesting parameter a downward bias. In general, a researcher may want to pursue as much flexibility as possible ($M > 1$), but doing so may determine situations where the neighboring nesting parameter is greater than nesting parameters ($\rho > \sigma$), which is, again, inconsistent with random utility maximization;

(iii) Direct estimation of the weight coefficients requires the use of additional instruments and the estimates tend to be rather imprecise. I also assess the role of the weight choice by estimating a model in which fixed weights are intentionally misspecified (not estimated). I find that the demand parameters are hardly impacted by the misspecification; the substitution patterns are close to the true ones. In conclusion, possible misspecifications in weights do not seem to affect the parameter estimates of interest and the resulting substitution patterns.

3 Empirical study

3.1 Data

I now turn to the application of the Ordered Nested Logit to the automobile market. For the empirical study, I combine different datasets. The main one is a dataset on the automobile market provided by a marketing research firm, JATO: it includes essentially all transactions of passenger cars sold between 1998 and 2011 in the three largest European car markets: France, Germany, and Italy. The data is highly disaggregated, and I aggregate it at the level of the car model (nameplate), e.g. Volkswagen Golf. For each car model/country/year, I have information on sales, prices and various characteristics such as vehicle size (curb weight, width and height), engine attributes (horsepower and displacement), fuel consumption (liter/100 km or €/100 km), emissions, the brands' specific perceived country of origin, and, for models introduced or eliminated within a given year, the number of months with positive sales. The dataset is augmented with macro-economic variables including the number of households for each country, fuel prices and GDP. Low-sold car models, which are more susceptible to recording or measurement errors, as well as non-passenger cars, such as pickups and large vans, are removed. I also exclude minivans, sports cars and sport utility vehicles because they do not naturally fit in a univocal ordering of the segments: for example, sports cars are on average more powerful but not more expensive than luxury cars. The resulting dataset consists of 5,788 model/country/year observations or, on average, about 138 models per country/year.[9]

Prices are list prices including value added taxes and registration taxes which differ across

[9] The model could be extended to incorporate multidimensionality in ordering. One could parametrize the potential correlation among products along two (or more) dimensions of ordering by taking the weighted sum of two Ordered Nested Logit generating function $G(\cdot)$ as follows:

$$G(e^{\delta}) = \alpha_{d1}G_{d1} + \alpha_{d2}G_{d2},$$

where $d1$ and $d2$ denote dimension1 and dimension 2 of the ordering.

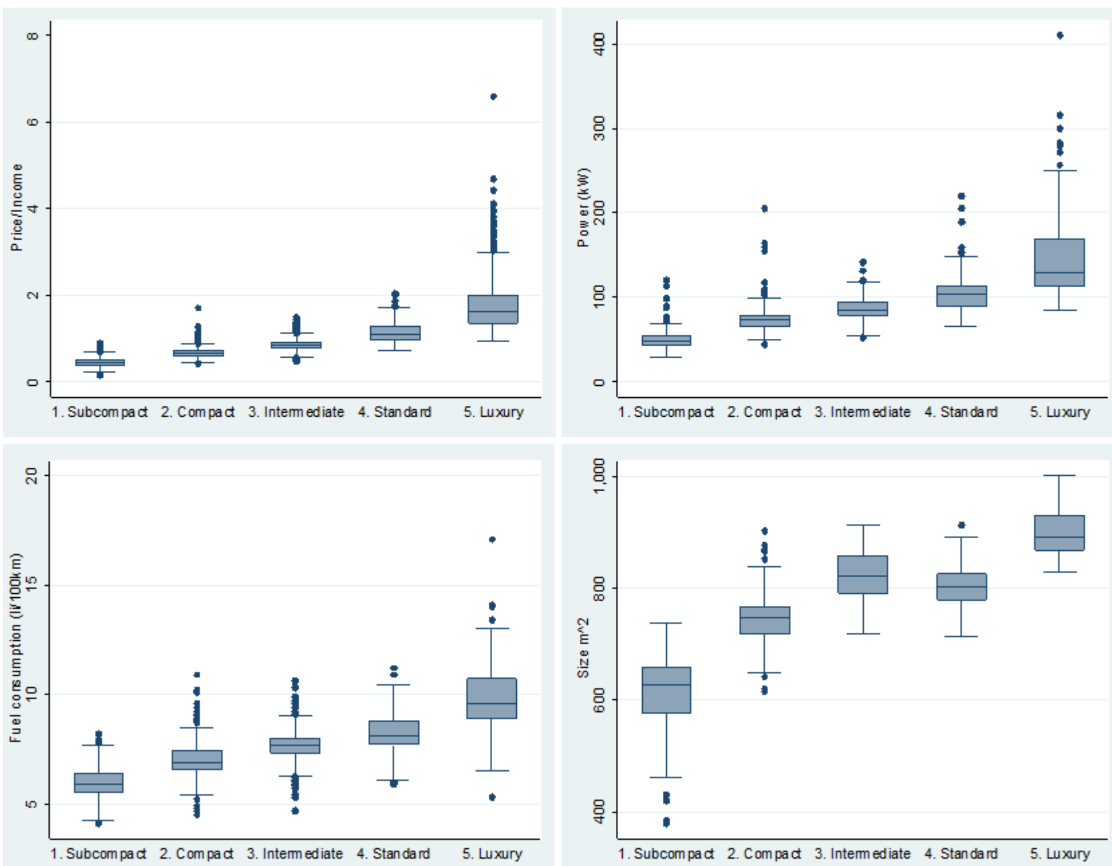
countries and engines: such information comes from the European Automobile Manufacturers Association. Prices are also corrected to account for active scrapping schemes and feebate programs according to the eligibility criteria for each vehicle: information on those programs comes from IHS Global Insight (an automotive consultant) and the European Automobile Manufacturers Association. Finally, the dataset is augmented with information on the location of the main production plant for each car model (from PWC Autofacts), and three input prices by country of production: unit labor costs, steel prices, and a producer price index. Table 2 presents summary statistics for sales, price, and vehicle characteristics used in demand estimation.

Starting from JATO's classification, I attribute each model to a marketing segment. I define five segments: subcompact, compact, standard, intermediate, and luxury.[10] Cars belonging to the same segment share similar characteristics in terms of price, horsepower, fuel consumption and size. Segmentation is used by carmakers to position their vehicle in the market place: they often advertise their vehicle as the cheapest or best performing in its class. Leading automotive magazines, such as *Auto motor und sport*, award a 'best car' prize for each segment. Comparison websites and consumer reports also feature the classification into segments as a prominent search tool. But the boundaries between segments are blurred by the presence of cars with some characteristics, including price, image and extra accessories, which would position those cars in an upper segment. Audi A1 or BMW Mini are examples of 'luxury subcompacts' designed to compete across segments. Table 3 and Figure 1 provide a descriptive illustration of segmentation in the car market. The top panel of the table presents the mean and standard deviation of price, horsepower, fuel consumption, and size by segment. Figure 1 represents also the median, the minimum and maximum values, and the values of the lower and upper quartiles of those characteristics. The table and the figure illustrate that the mean and median values of all characteristics increase from subcompact

[10] For example, a Volkswagen Golf belongs to the compact segment. The smaller Polo belongs one segment below the Golf (subcompact), while the bigger Passat is located one segment above (intermediate).

to luxury (with the exception of size from the intermediate to standard segment). At the same time, the large variability displayed by those characteristics within a segment suggest that some overlap across segments is plausible and depends on the proximity of the ordering. The bottom panel of Table 3 shows how well characteristics predict to which segment each car model belongs. Classifications are reasonably accurate (always above 80% with one exception), but the prediction power is not perfect and confirms the need to quantify the presence of neighboring segment effects.

Figure 1: Characteristics by Segment



The figure reports the median, the minimum and maximum values, and the values of the lower and upper quartiles by segment of the following vehicle characteristics: price/income, power, fuel consumption, size.

Table 2: Summary Statistics

	Mean	Std. Dev.
Sales (units)	13,821	24,312
Price/Income	0.84	0.50
Power (in kW)	82.19	35.72
Fuel efficiency (€/100 km)	7.27	1.46
Size (m^2)	7.46	1.14
Foreign (0-1)	0.78	0.42
Months present (1-12)	9.66	2.61

The table reports means and standard deviations of the main variables. The total number of observations (models/markets) is 5,788, where markets refer to the 3 countries and 14 years.

Table 3: Summary Statistics by Segment

		Subcomp.	Compact	Interm.	Standard	Luxury
Price/Income	Mean	0.45	0.68	0.86	1.13	1.80
	Std. Dev.	(0.10)	(0.12)	(0.13)	(0.22)	(0.67)
Power	Mean	50.37	73.84	88.52	104.91	145.47
(kW)	Std. Dev.	(10.83)	(14.16)	(12.96)	(19.48)	(42.44)
Fuel consumption	Mean	5.90	6.98	7.69	8.21	9.65
(li/100km)	Std. Dev.	(0.71)	(0.72)	(0.63)	(0.88)	(1.34)
Size	Mean	6.10	7.46	8.23	8.06	9.00
(m2)	Std. Dev.	(0.71)	(0.43)	(0.41)	(0.34)	(0.38)
Number of obs.		1,802	1,409	1,131	716	730
		Correct classifications into segments (percent)				
Subcompact		-	92.37	97.28	98.89	100.00
Compact			-	74.59	89.80	96.27
Intermediate				-	81.20	90.74
Standard					-	84.72
Luxury						-

The top panel of the table reports means of the main variables per segment in the top panel. The bottom panel of the table reports the percentage of correctly classified car models, based on binary logit of a segment dummy per pair on four continuous characteristics (i.e. power, fuel efficiency, width and height). Subcomp.=subcompact, Interm=intermediate.

3.2 Specification

To estimate the demand for cars in France, Germany, and Italy, I modify the Ordered Nested Logit specified above. In each period (year) and country t, L_t potential consumers choose one alternative, either the outside good $j = 0$ or one of the J cars. Following Berry (1994) and the subsequent literature, price is treated separately because it is an endogenous characteristic. The outside good includes the option 'do not buy a product', $j = 0$ for which consumer i's indirect utility is $u_{i0} = \varepsilon_{i0}$. For cars $j = 1, ..., J$, the utility specification is:

$$U_{ijt} = x_{jt}\beta - \alpha_i p_{jt} + \xi_{jt} + \varepsilon_{ijt} \equiv \delta_{jt} - \alpha_i p_{jt} + \varepsilon_{ijt},$$

where x_{jt} is a $1 \times K$ vector of characteristics including price, horsepower, fuel consumption, size measures (width and height), and a dummy variable for the country of origin. For the potential market size (L_t), I follow the literature and use the total number of households in each year and market.

In estimation, the coefficient of price, α_i, is specified in two ways: (i) $\alpha_i = \alpha/y$, where y is equal to income per capita; (ii) $\alpha_i = \alpha/y_i$, a specification in which I exploit information on income distribution. Both specifications imply that households with higher income are less sensitive to price.[11]

The error term ε_{ij} is the individual realization of the random variable ε : as discussed above, its distribution determines the substitution patterns. I assume that the $5 + 1$ nests (segments) are ordered as follows: S_0, the outside good; S_1, subcompact; S_2, compact; S_3, standard; S_4, intermediate; and S_5, luxury. The ordering corresponds to an increasing value

[11] Specifically, I take advantage of the easily available information on the percentage of the total income attributable to each decile of population. I couple this information with aggregate income by country (y) to obtain the average consumer income for the 10 segments of the population. I can then compute the purchase probability by population segment and sum up those probabilities to obtain the market share for each car in each country and year. I specify the consumers' distaste for higher prices as $\alpha_i = \frac{\alpha}{y_i}$ where α is the estimated parameter. As noted by Berry et al. (2004a) and Brenkers and Verboven (2006), this functional form can be derived as a first-order Taylor series approximation of the Cobb-Douglas utility function used in Berry et al. (1995) as long as price is small relative to capitalized income.

of observable and unobservable characteristics such as price, size, comfort, and performance. The outside good nest is the nest with the 'inferior quality' good. [12]

The distribution of the error term ε_{ij} thus follows the assumptions of the Ordered Nested Logit as defined in Equation (2). To obtain as much flexibility as possible, I assume that $M = 2$, so that each segment S has two contiguous segments as neighbors, or, in other words, each segment belongs to 3 different sets of segments B. In sum, I have $6+2$ sets containing up to 3 contiguous nests (and *all* the alternatives in those nest):

$$\begin{aligned}
B_0 &= \{S_0\} \\
B_1 &= \{S_0, S_1\} \\
B_2 &= \{S_0, S_1, S_2\} \\
B_3 &= \{S_1, S_2, S_3\} \\
B_4 &= \{S_2, S_3, S_4\} \\
B_5 &= \{S_3, S_4, S_5\} \\
B_6 &= \{S_4, S_5\} \\
B_7 &= \{S_5\}
\end{aligned}$$

The nesting parameter σ_n differs across the 5 nests. While the model theoretically has 8 neighboring nest parameters ρ_r (one for each set of nests $B_r = 0, ..., 7$), I impose the following structure to avoid the proliferation of parameters and issues in identification. As the decision to choose the outside option is fundamentally different from the decision to choose one of the alternatives in the choice set, I estimate one neighboring nest parameter for the sets containing the outside option (B_0, B_1, B_2), and one neighboring nest parameter

[12] The industry and the European Commission have at times used more detailed classifications, for example by distinguishing the subcompact segment between city/mini cars and small cars (segment A and B). When using more detailed classifications, I found that the model was not supported in the data $(\rho_r > \sigma_n)$: the result is consistent with the Monte Carlo analysis when nests are narrowly defined.

for the sets containing the inside good $(B_3, B_4, B_5, B_6, B_7)$. This implies the estimation of 7 random coefficients: 5 nesting parameters σ and 2 parameters determining the degree of proximity between groups of nests, $\rho_0 = \rho_1 = \rho_2$ and $\rho_3 = \rho_4 = \rho_5 = \rho_6 = \rho_7$. Finally, for simplicity all nests are assigned the same weight $1/(M+1) = 1/3$.

3.3 The estimation procedure

The estimation procedure for the Ordered Nested Logit model follows the methodological lines of Berry (1994), Berry et al. (1995) and the subsequent literature. I exploit the panel features of the dataset to specify the product-related error term as follows: $\xi_{jt} = \xi_j + \xi_t + \Delta\xi_{jt}$, where ξ_j is a fixed-effect for each car model, ξ_t is a full set of country/year fixed effects and a set of dummy variables for the number of months each model was available in a country within a given year (for models introduced or dropped within a year). $\Delta\xi_{jt}$ is the remaining product-related error term.

The estimation procedure is standard in the literature. First, I numerically solve for the error term $\Delta\xi_{jt}$ as a function of the vector of parameters. Second, I interact $\Delta\xi_{jt}$ with a set of instruments to form a generalized method of moments (GMM) estimator.

Consider the solution of $\Delta\xi_{jt}$ first. In the Nested Logit model $\Delta\xi_{jt}$ has an analytic solution. In the Ordered Nested Logit model $\Delta\xi_{jt}$ is the numerical solution of the system $s = s(\delta, \alpha, \sigma_n, \rho_r)$. I use a modified version of Berry et al.'s (1995) contraction mapping: $\delta^{k+1} = \delta^k + [1 - \max(\widehat{\sigma_n}, \widehat{\rho_r})] \cdot [\ln(s_t) - \ln(s_t(\delta_t^k))]$. If one does not weigh the second term by $[1 - \max(\widehat{\sigma_s}, \widehat{\rho_r})]$ the procedure may not lead to convergence; see Appendix A in Grigolon and Verboven (2014).

Let $\widehat{\Delta\xi}$ be the sample analogue of the vector $\Delta\xi$, and Z the matrix of instruments. The GMM estimator is defined as:

$$\min_{\alpha, \sigma_n, \rho_r} \widehat{\Delta\xi}'(Z\Omega Z')\widehat{\Delta\xi},$$

where Ω is the weighting matrix. I follow a two step-procedure: first I use the weighting matrix $\Omega = (Z'Z)^{-1}$. Then I re-estimate the model with the optimal weighting matrix. To minimize the GMM objective function with respect to the parameters α, σ_n, ρ_r, I first concentrate out the linear parameters β. Also, I do not directly estimate more than 200 car model fixed effects ξ_j, but instead use a within transformation of the data (Baltagi, 1995). Standard errors are computed using the standard GMM formulas for asymptotic standard errors. Following Dubé et al. (2012), I use a tight tolerance level to invert the shares using the contraction mapping ($1e-12$), check convergence for 10 starting values at each step, and check that the first order conditions are satisfied at convergence.

3.4 Identification

The GMM estimator requires an instrumental variable vector Z with a rank of at least $K+8$ (K is the dimension of the β vector; the price parameter α; the five nesting parameters σ_n and the two parameters characterizing correlation between neighboring nests ρ_r). The interpretation of $\Delta\xi_{jt}$ as unobserved product quality disqualifies price p_{jt} as an instrument since it could imply a positive correlation with $\Delta\xi_{jt}$. There are two main reasons for such correlation. First, if an unobservable characteristic, for example comfort, rises with price, consumers will avoid expensive cars less than they would without that characteristic. Second, if adding comfort is costly for the manufacturer, the price of the car is expected to reflect this cost. A similar argument holds for the correlation between the shares within a segment or within neighboring segments and $\Delta\xi_{jt}$: parameters σ_n and ρ_r are special kinds of random coefficients (Cardell, 1997). Berry and Haile (2014) clarify that, even abstracting from price endogeneity, identification of random coefficients requires instrumentation for the endogenous market shares: this calls for instrumentation of the share terms.

Following Berry et al. (1995), I assume that the observed product characteristics x_{jt} are uncorrelated with the unobserved product characteristics $\Delta\xi_{jt}$, so product characteristics x_{jt} are included in the matrix of instruments. Note that this assumption is weaker than the often adopted assumption that x_{jt} is uncorrelated with ξ_{jt}.

I include three sets of moment conditions. The first set focuses on the identification of the price coefficient. Armstrong (2016) suggests the use of cost-shifters, especially when the number of products is large, to identify price effects. I use input prices derived from the country of production of each car: a steel price index interacted with car weight (as a proxy for material costs) and unit labor costs in the country of production.

The second set of instruments, often used in the literature, includes interactions of the exogenous characteristics. In particular, I use (i) counts and sum of the characteristics of other products of competing firms by segment; (ii) counts and sum of the characteristics of other products of the same firm by segment; (iii) counts and sum of the characteristics of other products of competing firms by a set of segments B_r; (iv) counts and sum of the characteristics of other products of the same firm by a set of segments B_r. These instruments originate from supply side considerations, where I assume that firms set prices according to a Bertrand-Nash game. When the number of products in one segment, or in the neighboring segments increases, demand should become more elastic and this should affect prices and shares. Similarly, if one firm produces a large share of the products in one segment or in neighboring segments, sales and prices for each product of that particular firm should be higher.

Following Gandhi and Houde (2019), the third set of instruments is the difference in car attributes to capture the relative position of each product in the characteristic space. Those instruments approximate the optimal instrumental variables I used with simulated data without requiring initial estimates.[13] In particular I construct the sum of square of characteristic differences within each segment and within each set of segments, B_r.

[13] With simulated data, I did not need to use any approximation because I constructed the optimal instruments from the parameters and the functional form assumptions of the true data generating process.

4 Results

4.1 Demand estimates

Table 4 shows the parameter estimates for four specifications. The first one (Nested Logit I) is the one-level Nested Logit model, which imposes $\rho_r = 0$. The second specification (Ordered Nested Logit I) is an Ordered Nested Logit with $M = 2$; both σ_s and ρ_r are estimated and the coefficient of price, α_i, is specified as α/y, where y is equal to income per capita of each country. The third specification (Nested Logit II) is a Random Coefficient Nested Logit, which again imposes $\rho_r = 0$ but allows for heterogeneity in price sensitivities, so that $\alpha_i = \alpha/y_i$. Finally, the fourth specification (Ordered Nested Logit II) is an Ordered Nested Logit identical to the second specification, in which we have both σ_s and ρ_r, with the addition of heterogeneity in price sensitivities: $\alpha_i = \alpha/y_i$.

Allowing heterogeneity in price sensitivity is useful for two reasons. First, if a researcher believes that coefficients on both market segmentation and some continuously measured characteristics are random, a 'mixed' Ordered Nested Logit model can represent well such situation. Including a random coefficient on price is useful to illustrate the flexibility of the model. Second, modelling heterogeneity in price sensitivities is important because of the focus on price elasticities, especially when considering demand for large budget share products, such as cars. The Ordered Nested Logit model is well suited to capture heterogeneity attributable to market segmentation, and does so more flexibly than the Nested Logit while retaining tractability. At the same time, price sensitivity is a particularly relevant aspect of heterogeneity that may not be completely captured by market segmentation alone. Adding a random coefficient to the Ordered Nested Logit is a tractable solution to flexibly account for heterogeneity in both dimensions. As the random coefficient on price is based on the income distribution, it also accounts for differences in prices and market shares attributable to differences in the distribution of income across countries. Of course, this comes at the cost of losing the closed-form solution for market shares, but can be reasonable to capture the features of the market under study.

In all four models, the price parameter (α_i) and the parameters of the characteristics (β) have the expected sign and are all significantly different from zero. Most parameter estimates have also roughly the same magnitude. In the Nested Logit model I (Column 1), the nesting parameters are all precisely estimated; their magnitude is consistent with random utility maximization ($0 \leq \sigma_n < 1$) and (non-monotonically) decreases from subcompact to luxury: consumer preferences are more homogeneous for subcompact cars ($\sigma_1 = 0.95$) with respect to luxury cars ($\sigma_5 = 0.35$). This is consistent with earlier findings by Goldberg and Verboven (2001) and Brenkers and Verboven (2006). Higher values of σ_n also imply stronger within group substitution relative to substitution to the outside option.

In the second specification, the Ordered Nested Logit I (Column 2), parameters σ_n are again precisely estimated and non-monotonically decreasing. The first neighboring nesting parameter, capturing correlation between proximate nests when the outside nest option is included, is low in magnitude and imprecisely estimated ($\rho = 0.13$); the second neighboring nesting parameter is precisely estimated and indicates that correlation between neighboring segments is strongly supported by the data: $\rho = 0.68$ with a standard error of 0.08. The null hypothesis of $\rho_r = 0$ assumed by the Nested Logit is rejected against the alternative hypothesis of a more general Ordered Nested Logit model; in other words, the Nested Logit I is rejected against the more general Ordered Nested Logit I.

The third and fourth specifications are identical to the first and the second, with the addition of heterogeneity in price sensitivity. Again, the estimates of σ_n are significantly different from zero in both models. For the Ordered Nested Logit II, the null hypothesis of $\rho_r = 0$ assumed by the Nested Logit II is again rejected. In sum, both heterogeneity in price and in market segmentation are important as demonstrated by the Ordered Nested Logit II: when adding the dimension of price heterogeneity, correlation between neighboring segments remains relevant.[14]

The Ordered Nested Logit I and the Nested Logit II are non-nested models. I use the test proposed by Rivers and Vuong (2002) for selection among misspecified non-nested models, where the selection criteria is based on the value of the GMM objective function. In particular, I use values of the first-step objective function, employing the same estimator of the weighting matrix, based on the same set of instruments. The test statistic is $T = \frac{(Q_1 - Q_2)}{\widehat{\sigma}}$, where Q_i is the value of the GMM function of model i, and $\widehat{\sigma}$ the estimated value of the standard deviation of the difference between Q_i's.[15] The test shows that the Ordered Nested Logit I is asymptotically "better" (less misspecified) with respect to the Nested Logit II. I interpret the result as evidence that correlation between neighboring segments matters more in this data with respect to the another dimension of heterogeneity, price sensitivity.

Finally, all models imply similar own-price elasticities; demand is always elastic, which is consistent with oligopolistic profit maximization.

---

[14] I formally test (i) the Nested Logit I against the Ordered Nested Logit I, and (ii) the Nested Logit II against the Ordered Nested Logit II by using the likelihood ratio test adapted to the GMM context, where the likelihood ratio statistic is defined as the difference between the value of the objective function of the restricted model and the value of the objective function of the unrestricted model (Hayashi (2000)). Under the null hypothesis, the statistic is asymptotically χ^2 distributed with degrees of freedom equal to the number of restrictions. Each restricted model is rejected against the more general model.

[15] The variance of the difference is estimated using bootstrap simulations, with resampling of the independent markets t. The null hypothesis is that the two non-nested models are asymptotically equivalent. The first alternative hypothesis (H_1) is that model 1 is asymptotically "better" (less misspecified) than model 2; the second alternative hypothesis (H_2) is that model 2 is asymptotically better than model 1. The value of T is compared to the critical values of a standard normal; with α denoting the size of the test and $t_{\alpha/2}$ the value of the inverse standard normal distribution evaluated at $1 - \alpha/2$. If $T < -t_{\alpha/2}$, H_0 is rejected in favor of H_1; if $T > t_{\alpha/2}$, H_0 is rejected in favor of H_2. Otherwise, H_0 is not rejected.

Table 4: Parameter Estimates for Alternative Demand Models

	Nested Logit I (1)		Ordered NL I (2)		Nested Logit II (3)		Ordered NL II (4)	
	Estimate	St.Err.	Estimate	St.Err.	Estimate	St.Err.	Estimate	St.Err.
	Mean valuations for the characteristics in $x_{jt}(\beta)$							
Price/Income	-1.43	0.17	-1.31	0.13	-1.11	0.28	-1.02	0.20
Power (kW/100)	0.80	0.12	0.68	0.09	0.43	0.11	0.31	0.08
Fuel consumpt. (€/10,00km)	-0.72	0.10	-0.44	0.08	-0.84	0.10	-0.53	0.07
Width (cm/100)	0.52	0.18	0.45	0.14	0.60	0.18	0.54	0.14
Height (cm/100)	1.13	0.16	0.93	0.12	1.20	0.17	1.00	0.12
Foreign (0/1)	-0.44	0.02	-0.30	0.02	-0.48	0.02	-0.34	0.02
	Nesting parameters (σ_n)							
Subcompact	0.95	0.02	0.96	0.02	0.91	0.02	0.92	0.02
Compact	0.77	0.02	0.82	0.01	0.77	0.02	0.82	0.01
Intermediate	0.80	0.02	0.83	0.02	0.79	0.02	0.83	0.02
Standard	0.78	0.03	0.85	0.02	0.77	0.03	0.86	0.02
Luxury	0.35	0.07	0.68	0.05	0.33	0.07	0.68	0.05
	Neighboring Nesting Parameters (ρ_r)							
$\rho_0 = \rho_1 = \rho_2$	-		0.13	0.11	-		0.20	0.12
$\rho_3 = \rho_4 = \rho_5 = \rho_6 = \rho_7$	-		0.68	0.08	-		0.68	0.08
Model FE	Yes		Yes		Yes		Yes	
Year·Country FE	Yes		Yes		Yes		Yes	
Income distr.	No		No		Yes		Yes	
Own Elasticity	-6.931		-8.300		-4.181		-4.944	

The table shows the parameter estimates and standard errors for the three demand models: (i) the Nested Logit model, which assumes homogenous income distribution ($\alpha_i = \alpha/y$) and set the neighboring segmentation parameter at zero ($\rho = 0$); (ii) the Ordered Nested Logit I with homogenous income distribution ($\alpha_i = \alpha/y$) and two neighboring nest parameters, one for the sets containing the outside option (B_0, B_1, B_2), and one for the sets containing the inside good (B_3, B_4, B_5, B_6, B_7); (iii) the Nested Logit with heterogeneous income distribution ($\alpha_i = \alpha/y_i$); (iv) the Ordered Nested Logit with heterogeneous income distribution ($\alpha_i = \alpha/y_i$) and two neighboring nest parameters, one for the sets containing the outside option (B_0, B_1, B_2), and one for the sets containing the inside good (B_3, B_4, B_5, B_6, B_7). The total number of observations (models/markets) is 5,788, where markets refer to the 3 countries and 14 years. NL=Nested Logit.

4.2 Substitution patterns: segment-level price elasticities

The implications of rejecting the Nested Logit in favour of the Ordered Nested Logit model are most clearly illustrated by the implied substitution patterns at segment level. Table 5 presents own- and cross-price elasticities constructed by simulating the effect on demand of a joint 1% price increase of all cars in a given segment.

The own-price elasticities across the four models are similar in terms of magnitude and tend to be higher for the most expensive classes. The monotonic relationship between own-price elasticity and price is the result of the assumption that price enters utility linearly and is clearly mitigated by modelling heterogeneity in consumer preferences for segments (σ_n and ρ_r) and especially income: see Nested Logit II and Ordered Nested Logit II.

The cross-price elasticities are the most interesting results. By construction, the one-level Nested Logit model implies a fully symmetric substitution pattern, namely identical cross-price elasticities in each row. Thus, a 1% price increase to all subcompact cars raises demand in the compact and luxury segments by the same amount, 0.01%. By contrast, the Ordered Nested Logit model delivers more plausible substitution patterns. A 1% price increase in the subcompact segment has a stronger effect on demand of the two proximate segments: compact (+0.22%) and intermediate (+0.09%) compared to luxury (+0.01%). These numbers are comparable to the ones reported by Grigolon and Verboven (2014) in the analysis of the segment-level price elasticities for the random coefficients logit model. The Ordered Nested Logit model I is flexible, but still parsimonious in the number of parameters, so that only the two immediately proximate segments (on the left and on the right) are the neighboring ones. Outside the neighboring segments, the Ordered Nested Logit model still retains the modeling assumptions of the Nested Logit model. Thus, substitution patterns are symmetric outside the neighboring segments.

In the third and fourth models, the property of symmetry outside proximate segments does not hold as both models also incorporate a random coefficient on price. However, cross-price elasticities are still quite symmetric in the Nested Logit II. In the Ordered Nested Logit II, the cross-price elasticities are asymmetric, albeit such asymmetry is slightly less pronounced with respect to the Ordered Nested Logit I.

Table 5: Segment-level Price Elasticities in Germany for Alternative Demand Models

Nested Logit I	Outside	Subcompact	Compact	Intermediate	Standard	Luxury
Subcompact	0.010	-0.593	0.010	0.010	0.010	0.010
Compact	0.015	0.015	-0.872	0.015	0.015	0.015
Intermediate	0.006	0.006	0.006	-1.204	0.006	0.006
Standard	0.009	0.009	0.009	0.009	-1.417	0.009
Luxury	0.011	0.011	0.011	0.011	0.011	-2.092
Ordered Nested Logit I						
Subcompact	0.010	-0.829	0.224	0.093	0.009	0.009
Compact	0.014	0.336	-1.309	0.394	0.164	0.014
Intermediate	0.006	0.059	0.164	-2.514	0.474	0.195
Standard	0.008	0.008	0.095	0.657	-2.607	0.653
Luxury	0.010	0.010	0.010	0.323	0.776	-3.166
Nested Logit II	Outside	Subcompact	Compact	Intermediate	Standard	Luxury
Subcompact	0.009	-0.506	0.008	0.008	0.008	0.008
Compact	0.012	0.012	-0.709	0.012	0.012	0.012
Intermediate	0.005	0.005	0.005	-0.923	0.005	0.005
Standard	0.007	0.007	0.007	0.007	-1.042	0.007
Luxury	0.007	0.008	0.008	0.009	0.009	-1.386
Ordered Nested Logit II						
Subcompact	0.009	-0.716	0.182	0.065	0.007	0.007
Compact	0.012	0.272	-1.043	0.274	0.112	0.011
Intermediate	0.005	0.040	0.114	-1.848	0.357	0.119
Standard	0.006	0.006	0.064	0.487	-1.858	0.402
Luxury	0.006	0.007	0.007	0.193	0.467	-1.974

The table reports the segment-level own- and cross-price elasticities (when the price of all products in the same segment is increased by 1%). The elasticities are based on the parameter estimates in Table 4. They refer to Germany in 2011.

5 Counterfactuals

Entry of premium subcompact Since the early 2000s, luxury brands have entered the lower segments of the car market, such as subcompacts and compacts. The vehicles launched by those brands feature distinctive characteristics with respect to the incumbents: for their power, accessories, image, and, of course, price they resemble a vehicle from a higher segment. This trend has diluted the traditional borders between segments in the automobile market. I consider in particular three premium subcompacts: Audi A1, BMW Mini (both the hatchback and wagon versions) and the Fiat 500 Abarth, an upgraded version of the Fiat 500. Table C.11 in the Appendix presents summary statistics of the characteristics of those three vehicles compared to the average subcompact and compact car. Their price and horsepower are significantly higher, while there is no statistically significant difference in fuel consumption and size with respect to the average subcompact car. In contrast, with respect to the average compact car, only size is significantly lower.

I simulate a counterfactual scenario without those three premium subcompacts. Table 6 summarizes the implied diversion ratios by segment. Those ratios measure the fraction of sales diverted to other products, in the same segment or other segments, when the premium subcompacts are removed. In the simulation I account for the response of other carmakers by solving the differentiated product model for the change in equilibrium prices induced by the removal of the products. The Nested Logit model suggests that, absent the choice of premium subcompacts, 95% of sales would be diverted to other subcompact cars, while sales of upper segments would practically not be affected. The Ordered Nested Logit I, which allows for the possibility of asymmetric correlation between neighboring nests, still predicts that most substitution (93%) happens within the subcompact segment, but now 2% of sales would be diverted to compact cars. In both cases, the diversion ratio to the outside good is around 5%.

The Nested Logit II (with heterogeneity in price sensitivity) yields a slightly higher increase in sales of compact cars with respect to the Nested Logit (0.24 versus 0.10). In contrast, the Ordered Nested Logit II, which incorporates heterogeneity in price sensitivity as well, predicts that 4.7% of sales are diverted to the compact segment, and 11.6% to the outside good. Both the Nested Logit II and the Ordered Nested Logit II predict similar patterns of substitution towards the outside good.

Table 6: Diversion Ratios After the Removal of Premium Subcompact Cars

Diversion ratios (%)	Nested Logit I	Ordered Nested Logit I	Nested Logit II	Ordered Nested Logit II
Outside	5.17	4.63	11.34	11.60
Subcompact	94.60	93.13	88.14	82.85
Compact	0.10	1.94	0.24	4.70
Intermediate	0.03	0.25	0.08	0.64
Standard	0.04	0.03	0.10	0.11
Luxury	0.06	0.03	0.09	0.09

The table reports the diversion ratios (in percent) by segment after removing three premium subcompact car models: Audi A1, BMW Mini (both the hatchback and wagon versions) and the Fiat 500. Diversion ratios: share of fraction of sales diverted to other products in the same segment or other segments. The simulations are based on the parameter estimates in Table 4. They refer to Germany in 2011.

The effects of targeted environmental policies Asymmetric substitution patterns across segments are particularly important when looking at asymmetric policies. An example is a targeted scrapping scheme, which encourages consumers to scrap an old vehicle and purchase a cleaner one. The dataset comprises: (i) the 2009 German scrapping scheme, which was not targeted (it provided an incentive to purchase a new car regardless of its fuel efficiency); (ii) the 2008-2011 French scrapping scheme, which was targeted, and the feebate program (Bonus/Malus); (iii) various Italian scrapping schemes, which are mostly targeted but not sizeable.[16] The French scrapping scheme in combination with the feebate program is the only notably asymmetric policy. In practice, cleaner cars in the dataset

[16] For more information, see Table A1 of Grigolon et al. (2016) and Table 1 of D'Haultfoeuille et al. (2014).

mostly received only a modest rebate (€200), while polluting cars were mostly subject to a modest fee ranging from €200 to 750. Cars emitting more than 160g of CO_2 per kilometer would be subject to the sizeable fee of €2,600, but only a handful of cars in the data meet the requirement, so the asymmetry in the policy is limited.[17]

What would be the effect of a bolder environmental policy? I simulate the impact of a €5,000 subsidy to cars emitting less than 140g of CO_2 per kilometer. The first column of Table 7 illustrates the asymmetry of the policy as it mostly benefits subcompact and compact cars. The other columns simulate the effect of the subsidy. As in the previous counterfactual, I account for the pricing responses of manufacturers. Under the Nested Logit model, subcompact cars gain a significant amount of sales (+ 24%). Also for the compact and intermediate segments sales increase, but by a smaller amount. Most importantly, standard and luxury cars are unaffected by the policy. The Ordered Nested Logit I model tells another story: sales of non-eligible cars, especially in the standard segment, are affected by the policy and decrease by 4.2%. The Ordered Nested Logit II predicts a similar lower decrease (3%).

[17] I tested the predictions of the four models to the French environmental policy. In particular, I compare the market shares observed in 2007 (before the policy) and the simulated market shares of 2008 setting the environmental policy to zero and the fuel prices at the level of 2007. Table C.12 in the Appendix shows that the four models, though suffering from the limitations of a static framework, predict counterfactual shares that are very close to the observed ones. The Ordered Nested Logit models imply counterfactual shares similar to the ones produced by the Nested Logit models because the asymmetry in the policy is actually rather limited.

Table 7: The Effect of a Subsidy to Clean Cars on Market Shares

	Eligible cars %	% Change in Sales Nested Logit I	Ordered Nested Logit I	Nested Logit II	Ordered Nested Logit II
Outside	-	-0.61	-0.60	-0.45	-0.46
Subcompact	93.02	24.28	30.00	18.36	23.37
Compact	39.39	8.57	4.82	6.21	3.01
Intermediate	8.33	5.40	4.43	3.75	2.84
Standard	0.00	-0.61	-4.21	-0.46	-2.97
Luxury	0.00	-0.65	-1.45	-0.47	-1.03

The table reports the effect of €5,000 subsidy to cars emitting less than 140g of CO_2. The simulations are based on the parameter estimates in Table 4. They refer to Germany in 2011.

6 Conclusion

I present a new member of the GEV model family denominated Ordered Nested Logit model. The Ordered Nested GEV model is appealing for three reasons. First, it provides a modeling theory that is more consistent with the particular structure of choices in some segmented markets, such as cars, than a simple Nested Logit model. It creates the potential for neighboring segment effects, or, more precisely, asymmetric substitution patterns across segments. Second, the model relaxes the hierarchical nesting structure imposed by the Nested Logit model while avoiding the simulation techniques of the random coefficients logit model. Third, the Ordered Nested GEV model has the Nested Logit and the Logit as special cases. It can thus serve as a test for the validity of the constraints imposed by the Nested Logit and, a fortiori, the Logit model.

I apply the Ordered Nested Logit model to the car market which is classified into segments that are naturally ordered from subcompact to luxury. Results show that neighboring segment effects are strongly supported in the data. I show that asymmetry in substitution matters when simulating the introduction of vehicles combining features from different segments, such as premium subcompacts, or when studying the consequence of asymmetric policies, such as targeted subsidies.

The model I propose here can be a promising starting point to capture neighboring segment effects. Future research on other industries such as retail brands, lodging and restaurants, could benefit from this modeling strategy: ordering a high number of alternatives can prove impossible, but ordering groups of these alternatives may represent a sensible way to obtain flexible substitution patterns in a tractable setting.

References

Allenby, Greg M. and Peter E. Rossi, "Marketing models of consumer heterogeneity," *Journal of Econometrics*, nov 1998, *89* (1-2), 57–78.

Armstrong, Timothy B., "Large Market Asymptotics for Differentiated Product Demand Estimators With Economic Models of Supply," *Econometrica*, 2016, *84* (5), 1961–1980.

Baltagi, Badi H., *Econometric Analysis of Panel Data*, John Wiley & Sons, 1995.

Ben-Akiva, Moshe and Michel Bierlaire, "Discrete choice methods and their applications to short-term travel decisions," in R. Hall, ed., *Handbook of Transportation Science*, Operations Research and Management Science Kluwer 1999, pp. 5–34. ISBN:0-7923-8587-X.

Berry, Steven, James Levinsohn, and Ariel Pakes, "Voluntary Export Restraints on Automobiles: Evaluating a Trade Policy," *American Economic Review*, jun 1999, *89* (3), 400–431.

Berry, Steven T., "Estimating Discrete-Choice Models of Product Differentiation," *The RAND Journal of Economics*, 1994, *25* (2), 242–262.

_ **and Philip A. Haile**, "Identification in Differentiated Products Markets Using Market Level Data," *Econometrica*, September 2014, *82*, 1749–1797.

_ **, James Levinsohn, and Ariel Pakes**, "Automobile Prices in Market Equilibrium," *Econometrica*, 1995, *63* (4), 841–890.

_ **,** _ **, and** _ , "Differentiated Products Demand Systems from a Combination of Micro and Macro Data: The New Car Market," *Journal of Political Economy*, February 2004, *112* (1), 68–105.

_ **, Oliver B. Linton, and Ariel Pakes**, "Limit Theorems for Estimating the Parameters of Differentiated Product Demand Systems," *Review of Economic Studies*, 07 2004, *71*, 613–654.

Bhat, Chandra R., "An Analysis of Travel Mode and Departure Time Choice for Urban Shopping Trips," *Transportation Research Part B: Methodological*, 1998, *32* (6), 361 – 371.

Bierlaire, Michel, “A theoretical analysis of the cross-nested logit model,” *Annals of Operations Research*, apr 2006, *144* (1), 287–300.

Blundell, Richard and Jean-Marc Robin, “Latent Separability: Grouping Goods without Weak Separability,” *Econometrica*, jan 2000, *68* (1), 53–84.

Brenkers, Randy and Frank Verboven, “Liberalizing A Distribution System: The European Car Market,” *Journal of the European Economic Association*, 03 2006, *4* (1), 216–251.

Bresnahan, Timothy F., Scott Stern, and Manuel Trajtenberg, “Market Segmentation and the Sources of Rents from Innovation: Personal Computers in the Late 1980s,” *The RAND Journal of Economics*, 1997, *28*, S17–S44.

Brunner, Daniel, Florian Heiss, Andre Romahn, and Constantin Weiser, “Reliable estimation of random coefficient logit demand models,” DICE Discussion Papers 267, University of Düsseldorf, Düsseldorf Institute for Competition Economics (DICE) 2017.

Cardell, Scott N., “Variance Components Structures for the Extreme-Value and Logistic Distributions with Application to Models of Heterogeneity,” *Econometric Theory*, April 1997, *13* (02), 185–213.

Chamberlain, Gary, “Asymptotic efficiency in estimation with conditional moment restrictions,” *Journal of Econometrics*, 1987, *34* (3), 305 – 334.

Chu, C., “A Paired Combinational Logit Model for Travel Demand Analysis,” in “Proceedings of Fifth World Conference on Transportation Research,” Vol. 4 Venturea 1989, pp. 295–309.

Daly, Andrew and S. Zachary, “Improved multiple choicemodels,” in Q. Dalvi eds. D.A. Hensher, ed., *Determinants of Travel Choice*, Saxon House Farnborough, 1977, pp. 335–357.

Davis, Peter and Pasquale Schiraldi, “The flexible coefficient multinomial logit (FC-MNL) model of demand for differentiated products,” *RAND Journal of Economics*, March 2014, *45* (1), 32–63.

D’Haultfoeuille, Xavier, Pauline Givord, and Xavier Boutin, “The Environmental Effect of Green Taxation: The Case of the French Bonus/Malus,” *The Economic Journal*, 2014, *124* (578), F444–F480.

Dubé, Jean-Pierre H., Jeremy T. Fox, and Che-Lin Su, "Improving the Numerical Performance of Static and Dynamic Aggregate Discrete Choice Random Coefficients Demand Estimation," *Econometrica*, September 2012, *80* (5), 2231 – 2267.

Eizenberg, Alon, "Upstream Innovation and Product Variety in the U.S. Home PC Market," *The Review of Economic Studies*, 2014, *81* (3), 1003–1045.

Fosgerau, Mogens, Julien Monardo, and André de Palma, "The Inverse Product Differentiation Logit Model," Working Papers hal-02183411, HAL July 2019.

Gandhi, Amit and Jean-François Houde, "Measuring Substitution Patterns in Di?erentiated-Products Industries," Working Paper 26375, National Bureau of Economic Research October 2019.

Goldberg, Pinelopi Koujianou and Frank Verboven, "The Evolution of Price Dispersion in the European Car Market," *Review of Economic Studies*, October 2001, *68* (4), 811–48.

Grigolon, Laura and Frank Verboven, "Nested Logit or Random Coefficients Logit? A Comparison of Alternative Discrete Choice Models of Product Differentiation," *The Review of Economics and Statistics*, December 2014, *96* (5), 916–935.

_ , Nina Leheyda, and Frank Verboven, "Scrapping subsidies during the financial crisis — Evidence from Europe," *International Journal of Industrial Organization*, jan 2016, *44*, 41–59.

Hardy, Michael, "Combinatorics of Partial Derivatives," *The Electronic Journal of Combinatorics*, jan 2006, *13* (1).

Hayashi, Fumio, *Econometrics*, Princeton University Press, December 2000.

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

Knittel, Christopher R. and Konstantinos Metaxoglou, "Estimation of Random Coefficient Demand Models: Challenges, Difficulties and Warnings," Working Paper 14080, National Bureau of Economic Research June 2008.

Lee, Jinhyuk and Kyoungwon Seo, "A computationally fast estimator for random coefficients logit demand models using aggregate data," *The RAND Journal of Economics*, jan 2015, *46* (1), 86–102.

McFadden, Daniel L., "Modelling of Choice of Residential Location," in F. Snickers A. Karlquist, L. Lundquist and J. Weibull, eds., *Spatial Interaction Theory and Residential Location*, 1978.

Nevo, Aviv, "Mergers with Differentiated Products: The Case of the Ready-to-Eat Cereal Industry," *RAND Journal of Economics*, Autumn 2000, *31* (3), 395–421.

Petrin, Amil, "Quantifying the Benefits of New Products: The Case of the Minivan," *Journal of Political Economy*, 2002, *110*, 705–29.

Pinkse, Joris, Margaret E. Slade, and Craig Brett, "Spatial Price Competition: A Semiparametric Approach," *Econometrica*, may 2002, *70* (3), 1111–1153.

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

Rivers, Douglas and Quang Vuong, "Model selection tests for nonlinear dynamic models," *The Econometrics Journal*, jun 2002, *5* (1), 1–39.

Schiraldi, Pasquale, "Automobile replacement: a dynamic structural approach," *The RAND Journal of Economics*, 2011, *42* (2), 266–291.

Small, Kenneth A, "A Discrete Choice Model for Ordered Alternatives," *Econometrica*, March 1987, *55* (2), 409–24.

Small, Kenneth A., "Approximate generalized extreme value models of discrete choice," *Journal of Econometrics*, 1994, *62* (2), 351 – 382.

Train, Kenneth E., *Discrete Choice Methods with Simulation* number 9780521766555. In 'Cambridge Books.', Cambridge University Press, 2009.

Villas-Boas, Sofia Berto, "Vertical Relationships between Manufacturers and Retailers: Inference with Limited Data," *The Review of Economic Studies*, 2007, *74* (2), 625–652.

Vovsha, Peter, "Application of Cross-Nested Logit Model to Mode Choice in Tel Aviv, Israel, Metropolitan Area," *Transportation Research Record: Journal of the Transportation Research Board*, 1997, *1607*, 6–15.

Wen, Chieh-Hua and Frank S Koppelman, "The generalized nested logit model," *Transportation Research Part B: Methodological*, 2001, *35* (7), 627 – 641.

Williams, H., "On the formation of travel demand models and economic evaluation measures of user benefit," *Environment and Planning - Part A*, 1977, *9* (3), 285 – 344.

A Appendix A

A.1 Proof GEV

Proposition 1 *The following conditions are sufficient for Equation (2) to define a GEV generating function:*

(i) M is a positive integer;

(ii) σ_n and ρ_r are constants satisfying $0 \leq \rho_r \leq \sigma_n < 1$;

(iii) $w_m \geq 0$ and $\sum\limits_m w_m > 0$.

Proof The four properties of GEV generating functions are verified as follows. To simplify the notation, let $e^{\delta_j} = Y_j$.

1. G is non-negative since $Y_j \in \mathbb{R}_+ \forall j$, weights are non-negative and at least one weight is strictly positive (condition (iii))

2. G is homogeneous of degree 1, that is $G(\lambda Y_0, ..., \lambda Y_J) = \lambda G(Y_0, ..., Y_J)$

$$\begin{aligned}
G(\lambda Y_0, ..., \lambda Y_J) &= \sum_{r=1}^{N+M} \left(\sum_{n \in B_r} w_{r-n} \left(\sum_{j \in S_n} \exp \left(\lambda^{\frac{1}{1-\sigma_n}} Y_j^{\frac{1}{1-\sigma_n}} \right) \right)^{\frac{1-\sigma_n}{1-\rho_r}} \right)^{1-\rho_r}, \\
&= \sum_{r=1}^{N+M} \left(\sum_{n \in B_r} w_{r-n} \left(\lambda^{\frac{1}{1-\sigma_n}} \sum_{j \in S_n} \exp \left(Y_j^{\frac{1}{1-\sigma_n}} \right) \right)^{\frac{1-\sigma_n}{1-\rho_r}} \right)^{1-\rho_r}, \\
&= \sum_{r=1}^{N+M} \left(\sum_{n \in B_r} w_{r-n} \lambda^{\frac{1}{1-\rho_r}} \left(\sum_{j \in S_n} \exp \left(Y_j^{\frac{1}{1-\sigma_n}} \right) \right)^{\frac{1-\sigma_n}{1-\rho_r}} \right)^{1-\rho_r}, \\
&= \lambda \sum_{r=1}^{N+M} \left(\sum_{n \in B_r} w_{r-n} \left(\sum_{j \in S_n} \exp \left(Y_j^{\frac{1}{1-\sigma_n}} \right) \right)^{\frac{1-\sigma_n}{1-\rho_r}} \right)^{1-\rho_r}, \\
&= \lambda G(Y_0, ..., Y_J).
\end{aligned}$$

3. The limit property holds since weights are non-negative and at least one is strictly positive (condition (iii))

4. The property of the sign of the derivatives holds because $0 \leq \rho_r \leq \sigma_n < 1$ (condition (ii)).

To show that cross-partials exhibit the required property, I begin by simplifying the notation. Set $\alpha_{nr} = (1-\sigma_n)/(1-\rho_r)$ and $\beta_r = 1-\rho_r$, and note that $\alpha_{nr}, \beta_r \in [0,1]$. I can then rewrite G as:

$$G = \sum_{r=1}^{N+M} \left[\sum_{n=1}^{N} w_{r-n} \left(\sum_{j \in S_n} Y_j^{\frac{1}{1-\sigma_n}} \right)^{\alpha_{nr}} \right]^{\beta_r}.$$

Fix some $k \in \{1, \dots, J\}$, and let $K = \{i_1, i_2, \dots, i_k\}$ be a subset of $\{1, \dots, J\}$ containing k elements. I want to show that:

$$\frac{\partial^k G}{\partial Y_{i_1} \dots \partial Y_{i_k}}$$

is non-negative for odd k and non-positive for even k. This is equivalently to showing that:

$$\frac{\partial^k \tilde{G}}{\partial \tilde{Y}_{i_1} \dots \partial \tilde{Y}_{i_k}}$$

is non-negative for odd k and non-positive for even k, where $\tilde{Y}_j \equiv Y_j^{\frac{1}{1-\sigma_n}}$ for every nest n and product $j \in S_n$ and

$$\tilde{G} \equiv \sum_{r=1}^{N+M} \left[\sum_{n=1}^{N} w_{r-n} \left(\sum_{j \in S_n} \tilde{Y}_j \right)^{\alpha_{nr}} \right]^{\beta_r}.$$

I focus on the latter condition, and remove the tildes to ease notation.

For every r, let

$$H^r \equiv \left[\sum_{n=1}^{N} w_{r-n} \left(\sum_{j \in S_n} Y_j \right)^{\alpha_{nr}} \right]^{\beta_r},$$

and note that $H = \sum_r H^r$. Define

$$H_K^r \equiv \frac{\partial^k H^r}{\partial Y_{i_1} \dots \partial Y_{i_k}}.$$

It is enough to show that, for every r, H_K^r is non-negative for odd k and non-positive for even k. In the following, I fix an r, and drop the r index to ease notation. The final object of interest is therefore:

$$H = \left[\sum_{n=1}^{N} v_n \left(\sum_{j \in S_n} Y_j \right)^{\alpha_n} \right]^{\beta} \equiv f\left[g(Y)\right],$$

where $v_n = w_{n-r}$; $f(X) = (X)^\beta$; and $g(Y) = \sum_{n=1}^N v_n \left(\sum_{j\in S_n} Y_j\right)^{\alpha_n}$.

To find the $k-$th derivative of the function composition, I apply the multivariate version of the Faà di Bruno formula (Hardy (2006)), and derive the sign of each term in the Faà di Bruno expression:

$$\frac{\partial^k}{\partial Y_{i_1}\dots\partial Y_{i_k}} H(Y) = \sum_{\pi\in\Pi} f^{(|\pi|)}(Y) \cdot \prod_{B\in\pi} \frac{\partial^{|B|} g(Y)}{\prod\limits_{j\in B} \partial Y_j},$$

where π runs through the set Π of all partitions of the set $\{i_1, ..., i_k\}$; $B \in \pi$ means that the variable B runs through the list of all the blocks of the partition π; $|B|$ denotes the size of the block B and $|\pi|$ is the number of blocks in the partition π.

I now show that each term in the above sum has the desired sign. Fix some $\pi \in \Pi$. I distinguish two cases.

Case 1. Suppose the partition π is such that there exists $B_0 \in \pi$ such that $\forall n$, $B_0 \nsubseteq S_{n(B)}$. Then, clearly, $\prod\limits_{B_0\in\pi} \frac{\partial^{|B_0|} g(Y)}{\prod\limits_{j\in B_0} \partial Y_j} = 0$.

Case 2. Suppose instead that the partition π is such that, for every $B \in \pi$, there exists $n(B)$ such that $B \subseteq S_{n(B)}$. Then:

$$\begin{aligned}
& f^{(|\pi|)}(X) \cdot \prod_{B\in\pi} \frac{\partial^{|B|} g(Y)}{\prod\limits_{j\in B} \partial Y_j} \\
= \; & f^{(|\pi|)}(X) \cdot \prod_{B\in\pi} \frac{\partial^{|B|} v_{n(B)} \left(\sum_{l\in S_{n(B)}} Y_l\right)^{\alpha_{n(B)}}}{\prod\limits_{j\in B} \partial Y_j} \\
= \; & \prod_{i=0}^{|\pi|-1} (\beta - i)(f(X))^{(\beta-|\pi|)} \cdot \prod_{B\in\pi} \prod_{i=0}^{|B|-1} (\alpha_{n(B)} - i) v_{n(B)} \left(\sum_{l\in S_n} Y_l\right)^{(\alpha_{n(B)}-|B|)} \qquad (6)
\end{aligned}$$

The sign of Equation (6) is either zero or:

$$\begin{aligned}
\operatorname{sgn}\left(f^{(|\pi|)}(X)\cdot\prod_{B\in\pi}\frac{\partial^{|B|}g(Y)}{\prod_{j\in B}\partial Y_j}\right) &= (-1)^{|\pi|+1}\cdot(-1)^{|\pi|+\sum_{B\in\pi}|B|}\\
&= (-1)^{2|\pi|+1+\sum_{B\in\pi}|B|}\\
&= (-1)^{\{1+|i_1,i_2,\ldots,i_k|\}}\\
&= (-1)^{1+k}
\end{aligned}$$

Therefore:

$$\frac{\partial^k}{\partial Y_{i_1}\ldots\partial Y_{i_k}}G(Y)\begin{cases}\geq 0 & \text{if} \quad k \text{ is odd}\\ \leq 0 & \text{if} \quad k \text{ is even}\end{cases}$$

A.2 Decomposition into three Logits

According to the GEV postulate, the choice probability of buying product j is:

$$s_j = \frac{e^{\delta_j}\cdot G_j(e^{\delta_0,\ldots,\delta_J})}{G(e^{\delta_0,\ldots,\delta_J})},$$

where $G_j{=}\frac{\partial G}{\partial e^{\delta_j}}$ is the partial derivative of G with respect to e^{δ_j}, as derived above.

As G is defined by Equation (2), choice probabilities are therefore:

$$s_j = \frac{e^{\delta_j}\sum_{r=n}^{n+M}\left(e^{\delta_j}\right)^{\frac{\sigma_n}{1-\sigma_n}}\cdot w_{r-n}Z_n^{\frac{\rho_r-\sigma_n}{1-\rho_r}}\cdot D_r^{-\rho_r}}{\sum_{r=1}^{N+M}\left(D_r\right)^{1-\rho_r}}. \tag{7}$$

I asserted that the product of two conditional and one marginal probabilities in Equation

(3) equals the joint probability in the above Equation (7). I verify the assertion as follows:

$$
\begin{aligned}
s_j &= \frac{e^{\delta_j}\sum_{r=n}^{n+M}\left(e^{\delta_j}\right)^{\frac{\sigma_n}{1-\sigma_n}}\cdot w_{r-n}Z_n^{\frac{\rho_r-\sigma_n}{1-\rho_r}}\cdot D_r^{-\rho_r}}{\sum_{r=1}^{N+M}\left(D_r\right)^{1-\rho_r}}, \\
&= \sum_{r=n}^{n+M}\frac{\exp\left(\frac{\delta_j}{1-\sigma_n}\right)w_{r-n}\left(\sum_{j\in S_n}\exp\left(\frac{\delta_j}{1-\sigma_n}\right)\right)^{\frac{\rho_r-\sigma_n}{1-\rho_r}}\cdot\left(\sum_{n\in B_r}w_{r-n}\left(\sum_{j\in S_n}\exp\left(\frac{\delta_j}{1-\sigma_n}\right)\right)^{\frac{1-\sigma_n}{1-\rho_r}}\right)^{-\rho_r}}{\sum_{r=1}^{N+M}\left(\sum_{n\in B_r}w_{r-n}\left(\sum_{j\in S_n}\exp\left(\frac{\delta_j}{1-\sigma_n}\right)\right)^{\frac{1-\sigma_n}{1-\rho_r}}\right)^{1-\rho_r}}, \\
&= \sum_{r=n}^{n+M}\frac{\exp\left(\frac{\delta_j}{1-\sigma_n}\right)}{\sum_{j\in S_n}\exp\left(\frac{\delta_j}{1-\sigma_n}\right)}\cdot \\
&\quad \frac{w_{r-n}\left(\sum_{j\in S_n}\exp\left(\frac{\delta_j}{1-\sigma_n}\right)\right)^{\frac{1-\sigma_n}{1-\rho_r}}}{\sum_{n\in B_r}w_{r-n}\left(\sum_{j\in S_n}\exp\left(\frac{\delta_j}{1-\sigma_n}\right)\right)^{\frac{1-\sigma_n}{1-\rho_r}}}\cdot \\
&\quad \frac{\left(\left(\sum_{n\in B_r}w_{r-n}\left(\sum_{j\in S_n}\exp\left(\frac{\delta_j}{1-\sigma_n}\right)\right)\right)^{\frac{1-\sigma_n}{1-\rho_r}}\right)^{1-\rho_r}}{\sum_{r=1}^{N+M}\left(\sum_{n\in B_r}w_{r-n}\left(\sum_{j\in S_n}\exp\left(\frac{\delta_j}{1-\sigma_n}\right)\right)^{\frac{1-\sigma_n}{1-\rho_r}}\right)^{1-\rho_r}} \\
&= \sum_{r=n}^{n+M}\frac{\exp\left(\frac{\delta_j}{1-\sigma_n}\right)}{Z_n}\cdot\frac{w_{r-n}Z_n^{\frac{1-\sigma_n}{1-\rho_r}}}{\exp\left(I_r\right)}\cdot\frac{\exp\left((1-\rho_r)I_r\right)}{\sum_{r=1}^{N+M}\exp\left((1-\rho_s)I_s\right)}, \\
&= \sum_{r=n}^{n+M}s(j|n)\cdot s(n|B_r)\cdot s(B_r),
\end{aligned}
$$

where:

$$
\begin{aligned}
Z_n &= \sum_{j\in S_n}\exp\left(\frac{\delta_j}{1-\sigma_n}\right), \\
I_r &= \ln\sum_{n\in B_r}w_{r-n}Z_n^{\frac{1-\sigma_n}{1-\rho_r}}.
\end{aligned}
$$

A.3 First and second-order derivatives of the generating function G

First derivative. For $j \in S_n$, the first cross-derivative $G_j = \frac{\partial G}{\partial Y_j}$ is:

$$G_j = \sum_{r=n}^{n+M} Y_j^{\frac{\sigma_n}{1-\sigma_n}} \cdot w_{r-n} Z_n^{\frac{\rho_r - \sigma_n}{1-\rho_r}} \cdot D_r^{-\rho_r},$$

where Z_n and B_r are defined as follows:

$$\begin{aligned} Z_n &= \sum_{j \in S_n} Y_j^{\frac{1}{1-\sigma_n}}, \\ D_r &= \sum_{n \in B_r} w_{r-n} \left(\sum_{j \in S_n} Y_j^{\frac{1}{1-\sigma_n}} \right)^{\frac{1-\sigma_n}{1-\rho_r}}. \end{aligned}$$

Second derivative. The second cross-derivative $G_{ji} = \frac{\partial G_j}{\partial Y_i}$ is given by:

1. for $i, j \in S_n, i \neq j$

$$\begin{aligned} G_{ji} &= \sum_{r=n}^{n+M} -\frac{\rho_r}{1-\rho_r} Y_i^{\frac{\sigma_n}{1-\sigma_n}} Y_j^{\frac{\sigma_n}{1-\sigma_n}} \cdot w_{r-n}^2 Z_n^{\frac{2(\rho_r - \sigma_n)}{1-\rho_r}} \cdot D_r^{-1-\rho_r} \\ &+ \sum_{r=n}^{n+M} \frac{\rho_r - \sigma_n}{(1-\rho_r) \cdot (1-\sigma_n)} Y_i^{\frac{\sigma_n}{1-\sigma_n}} Y_j^{\frac{\sigma_n}{1-\sigma_n}} \cdot w_{r-n} Z_n^{\frac{1-\sigma_n}{1-\rho_r} - 2} D_r^{-\rho_r}. \end{aligned}$$

2. for $i, j \notin S_n$, and $i, j \in B_r, i \neq j$

$$G_{ji} = \sum_{r=n(i)}^{n(j)+M} -\frac{\rho_r}{1-\rho_r} Y_i^{\frac{\sigma_{n(i)}}{1-\sigma_{n(i)}}} Y_j^{\frac{\sigma_{n(j)}}{1-\sigma_{n(j)}}} \cdot w_{r-n(i)} Z_{n(i)}^{\frac{\rho_r - \sigma_{n(i)}}{1-\rho_r}} \cdot w_{r-n(j)} Z_{n(j)}^{\frac{\rho_r - \sigma_{n(j)}}{1-\rho_r}} \cdot D_r^{-1-\rho_r}.$$

3. for $i, j \notin S_n$ and $i, j \notin B_r, i \neq j$

$$G_{ji} = 0.$$

B Appendix B. Monte Carlo Simulations

B.1 Specification 1: Nested Logit and Ordered Nested Logit

I generate 500 datasets with $T = 10$ independent markets consisting of $J = 100$ products and one outside good. Each product j is described by a constant; one continuous characteristic

x_{jt}; an unobserved product characteristic ξ_{jt} drawn from a normal distribution. The continuous variable x_{jt} intends to mimic the variable price or quality in a non-simulated dataset and is drawn from a triangular distribution truncated at zero. Products are partitioned into five nests. In most markets, nests with cheaper products tend to have a larger number of products than nests grouping expensive products; to mimic this feature, the lowest nest (grouping products with lower values of the continuous characteristic x_{jt}) contains twice as many products with respect to the contiguous nest and so on. I assume that the data is generated according to an Ordered Nested Logit model, where the nesting parameter σ equals 0.5 and the neighboring segment parameter ρ equals 0.2. I use a set of optimal instruments generated within the model, following the approach of Chamberlain (1987) and Reynaert and Verboven (2014). The market shares are computed following the market share equation in (3) in which $M = 2$ and $w_m = 1/(M+1)$. Finally, in the simulation I minimize the GMM objective function using tight convergence criteria for the contraction mapping (1e-12) and the gradient (1e-6).

Table B.1 shows the estimated demand parameters. The correctly specified Ordered Nested Logit produces parameter estimates that are very close to the true parameters, with tight standard deviations. It is most interesting to check the nest-level elasticities, namely the effect of a joint 1% increase in the value of x_{jt} for all products in a given nest. Table B.2 shows the effect of a 1% increase in the price of all goods in nest 5, the 'luxury' nest (with products with the highest value of the continuous variable x_{jt}). Under the correctly specified Ordered Nested Logit model, if the price of all goods in nest 5 increases by 1%, consumers will be more likely to substitute to the neighboring segment (sales in nests 4 increase by 0.078%) with respect to the more distant ones (sales in nest 1 increase by 0.003%). By construction, the Nested Logit model implies fully symmetric substitution patterns, namely identical cross-elasticities: the Nested Logit model misses the asymmetry and tends to underestimate substitution outside the nest. As expected, the correctly specified Ordered Nested Logit model approximates the true elasticities well.

Table B.1: Results with Simulated Data; Set up 1: Parameter Estimates

	True	Nested Logit	Ordered Nested Logit
Constant	-5.00	-5.48	-5.07
		(0.07)	(0.10)
x_j	-1.00	-0.85	-1.00
		(0.02)	(0.02)
σ	0.50	0.51	0.50
		(0.02)	(0.02)
ρ	0.20	n/a	0.20
			(0.03)

The table reports the coefficient estimates and standard deviations (in parentheses) of the model parameters: the constant, the continuous characteristics x_{jt}, the nesting parameter (σ) and the neighboring nesting parameter (ρ). The estimates are based on 500 random samples of 10 markets and 100 products per market. The true model is the Ordered Nested Logit model.

Product misallocation I test the flexibility of the Ordered Nested Logit in handling misclassifications of products into nests, which may sometimes prove difficult in these models because alternatives need to be partitioned into non-overlapping groups. I generate data according to a Nested Logit model. I then fit a misspecified Nested Logit and an Ordered Nested Logit in which I vary the threshold of assignment to a nest: in particular, I assign the product with the highest value in nest 1 to nest 2. Table B.3 reports the extent of the bias in the elasticities of the misclassified product (product A). The bias in the own- and cross-price elasticities resulting from the misspecified Ordered Nested Logit is always smaller than the one resulting from the misspecified Nested Logit model.

Table B.2: Segment Elasticities: Ordered Nested Logit vs Nested Logit

	Nest 1	Nest 2	Nest 3	Nest 4	Nest 5
			Nested Logit		
Nest 5	0.0026	0.0026	0.0026	0.0026	-2.0353
			Ordered Nested Logit		
Nest 5	0.0030	0.0030	0.0224	0.0783	-2.6663
			True		
Nest 5	0.0030	0.0030	0.0227	0.0784	-2.6738

The table reports the nest-level own- and cross-price elasticities, when the price of *all* products in nest 5 is increased by 1%. The segment-level elasticities are based on the parameter estimates reported in Table B.1.

Table B.3: Nested Logit vs Ordered Nested Logit: Handling Misclassifications of Products into Nests

	Nested Logit (misclassified)			Ordered Nested Logit (misclassified)			Nested Logit (correctly classified)	
Bias	A	B	Bias	A	B	True	A	B
A	-0.1731	0.0064	A	0.0011	-0.0003	A	-0.9752	0.0138
B	0.0050	-0.2267	B	-0.0007	0.0017	B	0.0154	-1.1152

The table reports, on the right-hand side, product A and B own- and cross-price elasticities from simulated data generated according to a Nested Logit in which product A is classified in Nest 1 (True) and product B in Nest 2. On the left-hand side, the table reports the bias of a misspecified Nested Logit and Ordered Nested Logit in which product A is misclassified in nest 2. The estimates are based on 500 random samples of 10 markets and 100 products per market.

B.2 Specification 2: Ordered Nested Logit and Random Coefficient Logit

The second specification is similar to the first one. Again, I generate 500 synthetic datasets for $T = 10$ independent markets consisting of $J = 100$ products and one outside good for each market. Each product j is described by a constant; one continuous characteristic x_{jt} drawn from a triangular distribution truncated at zero; an unobserved product characteristic ξ_{jt} drawn from a normal distribution. Products are partitioned into five nests on the basis of the continuous characteristic x_{jt}: such partition is irrelevant for the DGP and will only be used in the estimation of the Ordered Nested Logit. Now, I specify the random coefficients vector β_i as a 2×1 vector of mean valuations for the constant and the continuous characteristic x_{jt} and Σ as a 2×2 matrix of parameters:

$$\beta_i = \beta + \Sigma \nu_i,$$

where ν_i is a vector of standard normal variables. The mean valuations for the constant and the continuous characteristic are set at $\beta = (-5, -1)$.

The matrix of parameters governing the heterogeneity in taste preferences is set at

$$\Sigma = \begin{bmatrix} 6 & 0.5 \\ 0.5 & 0.5 \end{bmatrix}.$$

Rather than estimating the variance-covariance matrix directly, I estimate the Choleski decomposition: $\Sigma = LL'$ where L is a lower diagonal matrix with positive diagonal elements.

These parameters are important to obtain realistic substitution patterns, but are typically hard to precisely identify: with market share data, we can only use the mean choice probabilities (the market shares) as moments that identify the heterogeneity parameters. Good instruments would mimic the ideal experiment of random variation in the characteristics of products, but such variation cannot be exploited, for example, in the case of a random coefficient on the constant. Hence, estimates of the standard deviation on the constant tend to be rather imprecise: see for example Berry et al. (1999); Nevo (2000); Petrin (2002) (the specification using only macro moments); Eizenberg (2014). Also, the majority of the literature that estimates random coefficient logit models does not allow consumer valuations to be correlated across characteristics, again because of the difficulties in the identification of those parameters.[18] The Ordered Nested Logit relies on the same variation in the data to identify the nesting and neighboring nesting parameters; by assuming and estimating a correlation structure based on the proximity of product groups, the model can be a parsimonious alternative to the Random Coefficient Logit model. In the simulations, for example, I will estimate three random coefficients for the correctly specified Random Coefficient Logit

[18] Nevo (2000), Villas-Boas (2007) obtain significant coefficient estimates by interacting the characteristics with demographics; Allenby and Rossi (1998) use Bayesian procedures to estimate a full covariance matrix of random coefficients for each brand.

model and two random coefficients for the misspecified Ordered Nested Logit.

I assume that data is generated by a Random Coefficient Logit process, so the market share equation is given by the logit choice probability integrated over the individual-specific valuations. I use the simulated data to estimate a Random Coefficient Logit model, and an Ordered Nested Logit with $M = 2$, in which there are no random coefficients.

Table B.4 shows the estimated demand parameters. The parameter of the correctly specified model, the Random Coefficient Logit, are estimated within the correct range. As before, the implications of the parameter estimates are illustrated by looking at the nest-level price elasticities. Table B.5 represents the effect of a 1% increase in price (the continuous variable x_{jt}) of all products in nest 5 on the market shares of the other nests. The true values of the elasticities show the asymmetry in substitution driven by the presence of the random coefficients; if the price of goods in nest 5 increases by 1%, consumers will be more likely to buy a product from a contiguous nest (sales in nests 4 increase by 0.11%) rather than buying a 'cheap' product (sales in nest 1 increase by 0.03%).[19] As expected, such a pattern is well captured by the correctly specified Random Coefficient Logit. The Ordered Nested Logit approximates such asymmetric substitution pattern even if the model is misspecified, with a slight overestimation of substitution toward the most immediate neighbor and underestimation toward the distant ones. In contrast, the substitution patterns to neighboring segments produced by the Nested Logit model are not only symmetric, but also underestimated by an order of magnitude. I consider variations in the degree of heterogeneity by varying the values in the matrix Σ. Intuitively, lower heterogeneity implies lower values of the nesting and neighboring nesting parameters σ and ρ.

[19] I experimented by adding more random coefficients on continuous variables; asymmetry becomes more pronounced, and the conclusions on the comparison between models hold.

Table B.4: Results with Simulated Data; Set up 2: Parameter Estimates

	True	Random Coefficient Logit	Ordered Nested Logit
Constant	-5.00	-5.11	-1.65
		(0.20)	(0.19)
x_{jt}	-1.00	-0.97	-0.93
		(0.14)	(0.07)
L_{11}	2.45	2.77	n/a
		(0.57)	
L_{21}	0.20	0.22	n/a
		(0.11)	
L_{22}	0.67	0.64	n/a
		(0.03)	
σ	n/a	n/a	0.74
			(0.10)
ρ	n/a	n/a	0.32
			(0.27)

The table reports the coefficient estimates and standard error (in parentheses) of the model parameters: the constant, the continuous characteristics x_{jt}, and the elements of L, the Choleski decomposition of the matrix of standard deviations. For the Ordered Nested Logit: the nesting parameter (σ) and the neighboring nesting parameter (ρ) and $M = 2$. The true model is the Random Coefficient Logit model.

Table B.5: Segment Elasticities: Ordered Nested vs Random Coefficients Logit

Panel A	Nest 1	Nest 2	Nest 3	Nest 4	Nest 5
			Random Coefficient Logit		
Nest 5	0.0452	0.0621	0.0832	0.1068	-2.2430
			Ordered Nested Logit		
Nest 5	0.0311	0.0311	0.0623	0.1921	-2.8551
			True		
Nest 5	0.0455	0.0634	0.0863	0.1127	-2.2985

The table reports the nest-level own- and cross-price elasticities (when the price of all products in one nest is increased by 1%). The segment-level elasticities are based on the parameter estimates reported in Table B.4.

Designing the nesting structure I use simulated data to give guidance on the nesting structure, with a focus on (i) the choice of the number of nests (N); (ii) the choice of the number of neighboring nests (M); (iii) the nesting weights. I start from set-up 1, in which the Ordered Nested Logit is correctly specified in terms of number of nests ($N = 5$). First, I estimate a model with a misspecified number of nests ($N = 15$). In the empirical application, the choice mimics more a detailed segment classification sometimes adopted by the industry and the European Commission where, for example, subcompact cars are split into city/mini cars and small cars. Table B.6 presents the parameter estimates of the misspecified Ordered Nested Logit (specification 1) along with the correctly specified one (specification 2, which reproduces the results in Table B.1). Results show that the neighboring nesting parameter tends to be overestimated: in 30% of the simulated datasets, the neighboring nesting parameter is greater than nesting parameters ($\rho > \sigma$), which is inconsistent with random utility maximization, while in the correctly specified case it happens only in 0.8% of the cases. After dropping the simulations for which $\rho > \sigma$, we see that both the segment and the neighboring nesting parameters are still overestimated and the standard deviation tends to be an order of magnitude larger with respect to the parameter estimates of the correct specification. Intuitively, the neighboring nest parameter is upward biased as it tries to capture the close substitution of products that should belong to the same nest by overestimating the neighboring nest parameter. I verify that the same intuitive upward bias holds when correctly specified DGP is the Random Coefficient Logit (Specification 2) and the number of nests is 15 instead of 5.

Second, I estimate a specification in which the number of neighboring nests M is misspecified. Table B.7 reports the parameter estimates. In column (2), the Ordered Nested Logit incorrectly assumes $M = 1$, while the true value in the DGP is $M = 2$. Such misspecification leads to a downward bias of the neighboring nest parameter. Also the substitution patterns are downward biased, especially for the neighboring products, but they are still closer to the true ones with respect to the Nested Logit model. When the true number of neighbors is $M = 1$, while the estimated Ordered Nested Logit model incorrectly specifies $M = 2$, the pattern is reversed: the nesting parameter σ presents a slight downward bias, and the neighboring nest parameter ρ an upward bias. I run the same exercise when the correctly specified DGP is the Random Coefficient Logit model in Specification 2: Table B.8 reports the parameter estimates and Table B.9 reports the substitution patterns associated to those estimates. In that case, when we look at the true elasticities it is evident that using $M = 2$ should give more flexibility and better approximation (as in the parameter estimates reported above). If I instead use $M = 1$, I find that the estimated nesting parameter is higher ($\sigma = 0.84$ versus $\sigma = 0.74$) and the neighboring nesting parameter lower ($\rho = 0.28$ versus $\rho = 0.32$). The substitution patterns to the most proximate neighbor are closer to the true value (0.1285 versus 0.1921) but present a larger underestimation toward the distant ones. In sum, using $M = 1$ instead of $M = 2$ yield overestimation of σ and underestimation of ρ as above. The pattern is reversed when $M = 2$.

Third, I examine to the role of weights in the Ordered Nested Logit. I experiment with a DGP in which weights are estimated rather than fixed. Estimation of weight coefficients requires the use of additional instruments to disentangle those parameters from the neighboring nesting parameter ρ and the nesting parameter σ. Table B.10 reports the parameter estimates of a specification in which weights are estimated rather than calibrated. The nesting parameters are correctly estimated, albeit their standard deviation is larger, especially for the neighboring nesting parameters ρ, which also presents a slight upward bias. Weights are not precisely estimated. The substitution patterns closely approximate the true ones. I also assess the role of the weight choice by estimating a model in which fixed weights are intentionally misspecified (but not estimated). I find that the demand parameters are hardly impacted by the misspecification; the substitutional patters are, again, close to the true ones. In conclusion, possible misspecifications in weight specification do not seem to affect the parameter estimates of interest to a large extent.

Table B.6: Results with Simulated Data; Incorrect number of nests: Parameter Estimates

	True	Ordered Nested Logit	Ordered Nested Logit
		(1)	(2)
		N misspecified	N correctly specified
Constant	-5.00	-4.80	-5.07
		(0.49)	(0.10)
x_j	-1.00	-0.78	-1.00
		(0.41)	(0.02)
σ	0.50	0.62	0.50
		(0.21)	(0.02)
ρ	0.20	0.42	0.20
		(0.30)	(0.03)

The table reports the coefficient estimates and standard deviations (in parentheses) of the model parameters: the constant, the continuous characteristics x_{jt}, the nesting parameter (σ) and the neighboring nesting parameter (ρ). The estimates are based on 500 random samples of 10 markets and 100 products per market. Specification (1) reports the parameter estimates of the model in which the number of nests N is misspecified. Specification (2) reports the parameter estimates of the correctly specified Ordered Nested Logit model.

Table B.7: Results with Simulated Data; Incorrect number of neighboring nests: Parameter Estimates

	True	Ordered Nested Logit	True	Ordered Nested Logit
		(1)		(2)
		M misspecified		M misspecified
Constant	-5.00	-5.21	-5.00	-4.68
		(0.08)		(0.12)
x_j	-1.00	-0.95	-1.00	-1.06
		(0.02)		(0.02)
σ	0.50	0.51	0.50	0.48
		(0.02)		(0.02)
ρ	0.20	0.15	0.20	0.30
		(0.03)		(0.05)
M	2	1	1	2

The table reports the coefficient estimates and standard deviations (in parentheses) of the model parameters: the constant, the continuous characteristics x_{jt}, the nesting parameter (σ) and the neighboring nesting parameter (ρ). The estimates are based on 500 random samples of 10 markets and 100 products per market. Specification (1) reports the parameter estimates of the model in which the number of neighboring nests is $M = 1$ instead of $M = 2$. Specification (2) reports the parameter estimates of the model in which the number of neighboring nests is $M = 2$ instead of $M = 1$.

Table B.8: Results with Simulated Data; Set up 2: Parameter Estimates, Ordered Nested Logit M=1

	True	Random Coefficient Logit	Ordered Nested Logit
Constant	-5.00	-5.11	-1.38
		(0.20)	(0.30)
x_{jt}	-1.00	-0.97	-1.00
		(0.14)	(0.04)
L_{11}	2.45	2.77	n/a
		(0.57)	
L_{21}	0.20	0.22	n/a
		(0.11)	
L_{22}	0.67	0.64	n/a
		(0.03)	
σ	n/a	n/a	0.84
			(0.08)
ρ	n/a	n/a	0.28
			(0.34)

The table reports the coefficient estimates and standard error (in parentheses) of the model parameters: the constant, the continuous characteristics x_{jt}, and the elements of L, the Choleski decomposition of the matrix of standard deviations. For the Ordered Nested Logit: the nesting parameter (σ) and the neighboring nesting parameter (ρ) and $M = 1$. The true model is the Random Coefficient Logit model.

Table B.9: Segment Elasticities: Ordered Nested vs Random Coefficients Logit

Panel A	Nest 1	Nest 2	Nest 3	Nest 4	Nest 5
			Random Coefficient Logit		
Nest 5	0.0452	0.0621	0.0832	0.1068	-2.2430
			Ordered Nested Logit		
Nest 5	0.0306	0.0306	0.0306	0.1285	-2.6041
			True		
Nest 5	0.0455	0.0634	0.0863	0.1127	-2.2985

The table reports the nest-level own- and cross-price elasticities (when the price of all products in one nest is increased by 1%). The segment-level elasticities are based on the parameter estimates reported in Table B.8.

Table B.10: Results with Simulated Data; Set up 1: Parameter Estimates with Weights

	True	Ordered Nested Logit	Ordered Nested Logit	Ordered Nested Logit
		(1)	(2)	(3)
		w estimated	w misspecified	w correctly specified
Constant	-5.00	-4.90	-5.00	-5.07
		(0.16)	(0.10)	(0.10)
x_j	-1.00	-0.99	-1.00	-1.00
		(0.03)	(0.03)	(0.02)
σ	0.50	0.50	0.50	0.50
		(0.02)	(0.02)	(0.02)
ρ	0.20	0.30	0.20	0.20
		(0.12)	(0.03)	(0.03)
w_1	0.33	0.42	fixed	fixed
		(0.17)		
w_2	0.33	0.35	fixed	fixed
		(0.13)		
w_3	0.33	0.51	fixed	fixed
		(0.19)		

The table reports the coefficient estimates and standard deviations (in parentheses) of the model parameters: the constant, the continuous characteristics x_{jt}, the nesting parameter (σ), the neighboring nesting parameter (ρ), and the weights. The estimates are based on 500 random samples of 10 markets and 100 products per market. Specification (1) reports the parameter estimates of the model in which weights are estimated. Specification (2) reports the parameter estimates in which weights are misspecified. The true model is the Ordered Nested Logit model in Specification (3).

C Appendix C. Additional Tables

Table C.11: Summary Statistics Premium Subcompact vs Subcompact and Compact

	Premium Sub	Subcompact	p-value	Premium Sub	Compact	p-value
Price	19,038	13,039	0.000	19,038	19,468	0.771
Power (in kW)	87.75	53.62	0.000	87.75	80.18	0.138
Fuel efficiency (€/100 km)	5.68	5.23	0.131	5.68	6.22	0.054
Size (m^2)	6.44	6.24	0.612	6.44	7.87	0.000

The table reports the summary statistics of Premium Subcompact cars vs. Subcompact cars and Premium Subcompact vs. Compact cars. It reports the means of four characteristic and the p-value of the difference of the means.

Table C.12: The Effect of Removing the French Feebate and Scrapping Subsidy

	2007 Observed	Nested Logit I	Ordered Nested Logit I	Nested Logit II	Ordered Nested Logit II
Subcompact	57.32	58.69	58.58	58.63	58.47
Compact	25.30	25.26	25.40	25.37	25.54
Intermediate	10.50	10.00	10.04	10.00	10.04
Standard	4.13	4.09	4.03	4.08	4.04
Luxury	2.75	1.96	1.95	1.92	1.92

The table reports: (i) the 2007 observed market shares by segment (first column); (ii) the simulated market shares obtained from the 2008 market shares after setting the French feebate program and the scrapping subsidy to zero and using the fuel price of 2007. The simulations are based on the parameter estimates in Table 4.