Supplementary Material

Supplement to “Blurred boundaries: A flexible approach for segmentation applied to the car market”

(*Quantitative Economics*, Vol. 12, No. 4, November 2021, 1273–1305)

Laura Grigolon
Department of Economics, University of Mannheim, MaCCI, and CEPR

Appendix B: Monte Carlo simulations

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

I generate 500 data sets 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 nonsimulated data set 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

Laura Grigolon: laura.grigolon@uni-mannheim.de

Available at http://qeconomics.org. https://doi.org/10.3982/QE1434

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)

Note: 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.

the asymmetry and tends to underestimate substitution outside the nest. As expected, the correctly specified Ordered Nested Logit model approximates the true elasticities well.

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 nonoverlapping 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

Note: 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

Note: 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 Coefficients Logit*

The second specification is similar to the first one. Again, I generate 500 synthetic data sets 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, Levinsohn, and Pakes (1999); Nevo (2000); Petrin (2002) (the specification using only macro moments); Eizenberg (2014). Also, the majority of the literature that estimates Random Coefficients

Logit models does not allow consumer valuations to be correlated across characteristics, again because of the difficulties in the identification of those parameters.[1] 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 Coefficients Logit model. In the simulations, for example, I will estimate three random coefficients for the correctly specified Random Coefficients Logit model and two random coefficients (the nesting parameters) for the misspecified Ordered Nested Logit.

I assume that data is generated by a Random Coefficients 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 Coefficients Logit model, and an Ordered Nested Logit with $M = 2$.

Table B.4 shows the estimated demand parameters. The parameter of the correctly specified model, the Random Coefficients 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

TABLE B.4. Results with simulated data; Set-up 2: parameter estimates.

	True	Random Coefficients 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)

Note: 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 Coefficients Logit model.

[1] Nevo (2000) and 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. Gandhi and Houde (2019) provide a very helpful discussion on the identification of correlated random coefficients. They use interactions between characteristics pairs to identify the correlation in taste heterogeneity and show the strength of their instrumenting strategy in a controlled environment.

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

Panel A	Nest 1	Nest 2	Nest 3	Nest 4	Nest 5
		Random Coefficients 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

Note: 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.

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%).[2] As expected, such a pattern is well captured by the correctly specified Random Coefficients 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.

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 data sets, 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

[2] 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.6. Results with simulated data; incorrect number of nests: parameter estimates.

	True	Ordered Nested Logit (1) N misspecified	Ordered Nested Logit (2) 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)

Note: 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.

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 Coefficients 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

Table B.7. Results with simulated data; incorrect number of neighboring nests: parameter estimates.

	True	Ordered Nested Logit (1) M misspecified	True	Ordered Nested Logit (2) 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

Note: 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 Coefficients 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)

Note: 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 Coefficients Logit model.

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 Coefficients 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

TABLE B.9. Segment elasticities: Ordered Nested vs. Random Coefficients Logit.

Panel A	Nest 1	Nest 2	Nest 3	Nest 4	Nest 5
		Random Coefficients 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

Note: 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.

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.10. Results with simulated data; Set-up 1: parameter estimates with weights.

	True	Ordered Nested Logit (1) w estimated	Ordered Nested Logit (2) w misspecified	Ordered Nested Logit (3) 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)		

Note: 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).

Appendix C: Additional tables

Table C.1. 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

Note: 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.2. 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

Note: 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.

References

Allenby, G. M. and P. E. Rossi (1998), "Marketing models of consumer heterogeneity." *Journal of Econometrics*, 89, 57–78. [4]

Berry, S., J. Levinsohn, and A. Pakes (1999), "Voluntary export restraints on automobiles: Evaluating a trade policy." *American Economic Review*, 89, 400–431. [3]

Chamberlain, G. (1987), "Asymptotic efficiency in estimation with conditional moment restrictions." *Journal of Econometrics*, 34, 305–334. [1]

Eizenberg, A. (2014), "Upstream innovation and product variety in the U.S. home PC market." *The Review of Economic Studies*, 81, 1003–1045. [3]

Gandhi, A. and J.-F. Houde (2019), "Measuring substitution patterns in differentiated-products industries." Working Paper 26375, National Bureau of Economic Research. [4]

Nevo, A. (2000), "Mergers with differentiated products: The case of the ready-to-eat cereal industry." *RAND Journal of Economics*, 31, 395–421. [3, 4]

Petrin, A. (2002), "Quantifying the benefits of new products: The case of the minivan." *Journal of Political Economy*, 110, 705–729. [3]

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

Villas-Boas, S. B. (2007), "Vertical relationships between manufacturers and retailers: Inference with limited data." *The Review of Economic Studies*, 74, 625–652. [4]

Co-editor Peter Arcidiacono handled this manuscript.

Manuscript received 4 September, 2019; final version accepted 20 April, 2021; available online 13 May, 2021.