A regression model of product differentiation

Mogens, Fosgerau

Danish Technical University

1 July 2016

A regression model of product differentiation

Mogens Fosgerau

July 29, 2016

Abstract

This paper develops a model of product differentiation that can be estimated using standard regression techniques and applies it to a panel data set of new car sales. The model allows for complex substitution patterns according to an overlapping nest structure that makes cars closer substitutes if the share brand, body type, and/or quality level. A nest comprising all the car alternatives ensure that they are closer substitutes with each other than with the outside good. In addition, the model comprises fixed effects by car model, controlling for unobserved car quality.

1 Introduction

This note provides a model to estimate the demand for a differentiated product that allows for complex substitution patterns and numerous fixed effects. The models is estimated using standard regression techniques without any convergence issues. The model is a specific instance of the generalized entropy model proposed by Fosgerau and de Palma (2016). It is here applied to a publicly available market level panel data set of new car sales covering 250 different kinds of cars, 5 countries and 30 years.

Modeling new car sales entails two fundamental issues that arise in many similar situations. The first issue is that the quality of cars is correlated with price but imperfectly observed; this creates an endogeneity issue. The second issue is the presence of complex substitution patterns that are not well described by a simple model such as the logit model.

The state of the art in the current empirical literature is the well-known BLP model (Berry et al., 1995). It addresses the two issues by using a random coefficients logit with fixed effects. However, estimation of the BLP model is complicated. It requires a nested fixed point algorithm where a fixed point iteration determining the fixed effects is nested within a numerical maximization routine. Moreover, the presence of random coefficients necessitates numerical integration to compute the likelihood. Knittel and Metaxoglou (2014) discuss the numerical

stability of BLP and finds that results can be critically unstable, even when the procedure is well carried out.

It is then desirable to have a way to deal with endogeneity and complex substitution patterns that does not suffer from the drawbacks of the BLP model. This note presents such a model that is applicable to panel data of market shares.

Generalized entropy models were proposed by Fosgerau and de Palma (2016) (FdP, henceforth). It is a general class of models that comprises dual representations of all ARUM as well as more general models. This note applies a particular instance of generalized entropy models that generalizes the nested logit model by allowing arbitrarily overlapping nests.

2 Model formulation

A representative consumer with income y faces goods $j = 0, ...J$, where $j = 1, ..., J$ are different car models and 0 is an outside good. Demand for cars is $q = (q_0, q_1, ..., q_J)$, which is non-negative and sums to 1, i.e., $q \in \Delta$, where Δ is the unit simplex. Each car has an associated price p_j and quality v_j, while $p_0 = v_0 = 0$. Income is sufficiently large that $y > \max_j \{p_j\}$. The representative consumer chooses demand to maximize utility

$$u(q) = \tau y + q \cdot (v - \tau p) + \Omega(q),$$

where $\tau > 0$ is a constant marginal utility of income and Ω is a generalized entropy with the following properties (from FdP): it is a concave function $\Omega : [0, \infty)^{J+1} \to \mathbb{R} \cup \{-\infty\}$ given by

$$\Omega(q) = \begin{cases} -q \cdot \ln S(q), & q \in \Delta \\ -\infty, & q \notin \Delta \end{cases}, \tag{2.1}$$

where $S: [0, \infty)^{J+1} \to [0, \infty)^{J+1}$ is continuous, homogenous of degree 1, and globally invertible. Furthermore, S is differentiable at any $q \in \operatorname{relint}(\Delta)$ with

$$\sum_{j=1}^{J} q_j \frac{\partial \ln S^{(j)}(q)}{\partial q_k} = \kappa, k \in \{1, ..., J\},$$

where $\kappa > 0$.

Utility maximization leads to demand

$$q(v, p) = \left(\frac{H^{(1)}\left(e^{v-\tau p}\right)}{\sum_{j=1}^{J} H^{(j)}\left(e^{v-\tau p}\right)}, ..., \frac{H^{(J)}\left(e^{v-\tau p}\right)}{\sum_{j=1}^{J} H^{(j)}\left(e^{v-\tau p}\right)} \right), \tag{2.2}$$

where $H = S^{-1}$ is the inverse of S. Moreover,

$$\ln S(q) = (v - \tau p) + c, \tag{2.3}$$

where $c \in \mathbb{R}$ is a constant that depends on $(v - \tau p)$.

As shown by FdP, this class of models comprises dual representations of all ARUM discrete choice models. For example, when $S(q) = q$ is the identity, then also H is the identity and demand is just logit demand.

Another illustrative example is the nested logit model. Partition the set of alternatives $\{0, 1, ..., J\}$ into nests $g \in \mathcal{G}$, denote by g_j the nest that contains alternative j, and let $q_g = \sum_{j \in g} q_j$. Then define

$$S^{(j)}(q) = q_j^{\mu_{g_j}} q_{g_j}^{1-\mu_{g_j}}, j = 0, ...J \tag{2.4}$$

where $\mu_g \in]0, 1]$ are parameters. Then S satisfies the conditions given. With this specification of S, the generalized entropy Ω contains terms $q_j \left(1 - \mu_{g_j}\right) \ln q_{g_j}$ that makes alternatives in the same nests closer substitutes (perfect substitutes if $\mu_{g_j} = 1$). The resulting demand is

$$q_j = \frac{e^{\frac{v_j - \tau p_j}{\mu_{g_j}}}}{\sum_{i \in g_j} e^{\frac{v_i - \tau p_i}{\mu_{g_j}}}} \frac{e^{\mu_{g_j} \ln\left(\sum_{i \in g_j} e^{\frac{v_i - \tau p_i}{\mu_{g_j}}}\right)}}{\sum_{g \in \mathcal{G}} e^{\mu_g \ln\left(\sum_{i \in g} e^{\frac{v_i - \tau p_i}{\mu_g}}\right)}},$$

which is the nested logit model.

Here we shall employ a more general version of this that allows for overlapping nests. The set of alternatives $\{0, 1, ..., J\}$ is grouped according to different criteria c. A criterion c assigns a value to each element of $\{0, ..., J\}$. Alternatives are grouped together on criterion c if they are assigned the same value by c. Then defining $\sigma_c(j) = \{k | c(k) = c(j)\}$, $\sigma_c(j)$ is the set of alternatives grouped with alternative J on criterion c. Define nesting parameters $\mu_c > 0$ with $\sum_c \mu_c < 1$ and let $\mu_0 = 1 - \sum_c \mu_c$. Then define for all j

$$S^{(j)}(q) = q_j^{\mu_0} \prod_c q_{\sigma_c(j)}^{\mu_c}.$$

As shown in FdP, this satisfies the conditions set out above. As before, each term $q_{\sigma_c(j)}^{\mu_c}$ makes alternatives closer substitutes if they belong to the same nest on criterion c, where the degree of substitutability is controlled by the parameter μ_c.

From (2.3) we obtain that

$$\mu_0 \ln q_j + \sum_c \mu_c \ln q_{\sigma_c(j)} = v_j - \tau p_j + c.$$

Since $\mu_0 > 0$ we may equivalently write

$$\ln q_j = -\sum_c \frac{\mu_c}{\mu_0} \ln q_{\sigma_c(j)} + \frac{1}{\mu_0} v_j - \frac{\tau}{\mu_0} p_j + \frac{c}{\mu_0}, \tag{2.5}$$

which suggests that this model may be estimated using regression.

3 Empirical model formulation

We employ panel data giving new car sales in countries indexed by m, and years indexed by t. Alternative $j = 0$ is the outside alternative. A nesting structure is defined for the inside alternatives. Equation (2.5) is then be elaborated into

$$\ln q_{jmt} = -\sum_c \gamma_c \ln q_{\sigma_c(j),mt} + \beta x_{jmt} + \eta_j + \eta_{mt} + \xi_{jmt}, j > 0 \tag{3.1a}$$

$$\ln q_{0mt} = \eta_0 + \eta_{mt} + \xi_{0mt}, \tag{3.1b}$$

where the structural parameters μ_c may be recovered from $\gamma_c = \mu_c/\mu_0$ using $\mu_0 = \frac{1}{1+\sum_c \gamma_c}$, $\mu_c = \gamma_c \mu_0$, since $\mu_0 + \sum_c \mu_c = 1$.

Going through the terms in (3.1) one by one, this means that the log market share for car j depends first on the log market share for cars that belong to the same category as car j on each of the criteria c. Second, the qualities v_j are parametrized by $\beta x_{jmt} + \eta_j$. Explanatory variables in x_{jmt} include observable car characteristics that change across countries and years as well as price. All other aspects of car quality are captured by the fixed effect η_j. Third, country and year specific fixed effects η_{mt} allow the constant c in (2.5) to be omitted. Finally, random shocks ξ_{jmt} are included and assumed to be mean independent of (x, η).

The model has thus been translated into a regression model with panel data. One issue remains, namely that terms $\ln q_{\sigma_c(j),mt}$ are endogenous, since they depend on the random shocks ξ_{jmt}. There are instruments available within the model, that are constructed by averaging independent variables across nests:

$$\bar{x}_{\sigma_c(j),mt} = \frac{1}{|\sigma_c(j)|} \sum_{i \in \sigma_c(j)} x_{imt}.$$

Variable	p5	mean	p95
$\ln q_j$	-11.12	-8.410	-6.031
$price$	.3833	.8287	1.580
$\ln weight$	6.430	6.771	7.223
$\ln fuel$	1.752	2.054	2.442
home	0	.1842	1

Table 1: Summary statistics

4 Data and estimation results

The data are from Frank Verboven's website (https://sites.google.com/site/frankverbo/), downloaded June 2014. The dataset covers five countries (Belgium, France, Germany, Italy, and the UK) over the 30 year period 1970-1999. Car models are grouped into 262 representative cars that cover most of sales during the period. The share for the outside good is taken to be the size of the population minus the total car sales in each country and year. The data comprises 11447 observations of annual car sales by country, year and car. Table 1 provides some summary statistics.

The models presented below use the following categorizations of car alternatives to form nests: The first categorization variable distinguishes inside goods from the outside good, such that all inside goods can be closer substitutes with each other than with the outside good. The second categorization used is the car class, which divides cars into subcompact, compact, intermediate, standard, and luxury. The third categorization is defined according to a perceived quality of different car brands.[1] There are four quality categories: one for high quality cars, primarily German and Swedish; one intermediate and one low quality category. A fourth category distinguishes the smaller Asian brands from the rest. The fourth categorization distinguishes between car producing firms.

The explanatory variables included in x are price, log weight, log fuel consumption, and a dummy for whether a car is produced domestically. The price variable is price relative to per capita GDP. To avoid overidentification, only the price variable is used to form instruments.

Models are estimated using the Stata module reghdfe (Correia, 2014), which allows for multiway fixed effects. This takes essentially zero time. Estimation results are provided in Table 2, which shows four models. The first, M1, includes four layers of nesting, as well as car price, weight, fuel consumption and a dummy indicating whether the car is produced domestically. All parameter estimates have the expected sign, but the nesting parameter for the inside alternatives $\ln q_{inside}$ is not significantly different from zero. Model M2 omits $\ln q_{inside}$; this has only small effect on the other parameter estimates. Model M3 also omits $\ln q_{firm}$, and this leads to some change in the other parameter estimates. Model M4 adds back $\ln q_{inside}$ with a corresponding parameter that is insignificantly different from zero while the other parameters

[1] Federica Liberini suggested this variable.

$\ln q_j$	M1	M2	M3	M4
$\ln q_{inside}$	-.9234			-1.210
$\ln q_{class}$	-.2798**	-.2867**	-.2221**	-.2144**
$\ln q_{quality}$	-.6747***	-.6978***	-.6665***	-.6368***
$\ln q_{firm}$	-.4394**	-.4456**		
$price$	-2.100***	-2.110***	-1.778***	-1.770***
$\ln weight$	1.354***	1.369***	1.309***	1.290***
$\ln fuel$	-1.111***	-1.106***	-1.128***	-1.133***
$home$	3.006***	3.030***	2.267***	2.250***
F-stat	538.77	607.52	959.36	833.67
dof	(8,11040)	(7,11041)	(6, 11042)	(7, 11041)
R^2	0.6192	0.6141	0.7155	0.7194
RMSE	1.115	1.122	.9633	.9568

Table 2: Summary statistics

Parameter significance indicated using *:p<10%,**:p<1%,***:p<0.1%

are not much affected. All models pass tests for underidentification.

5 Generating counterfactuals

This section illustrates how the generalized entropy model may be used to create counterfactual scenarios. Let q^0 be these market shares and use (3.1) to compute $v^0 = \ln S(q^1)$. Say a counterfactual scenario changes βx^0 to βx^1 and define $v^1 = v^0 + \beta(x^1 - x^0)$. Then, from (2.3), counterfactual demand q^1 may be found by solving

$$\ln S\left(q^1\right) = v^1 + c, \tag{5.1}$$

where c is a normalizing constant ensuring that demand sums to 1. FdP shows that this equation has a unique solution and that the following iteration always converges to this solution. Given a current candidate solution $q^{(n)}$, the next candidate is found as

$$q_j^{(n+1)} = \frac{q_j^{(n)} e^{v_j^1} / S^{(j)}\left(q^{(n)}\right)}{\sum_k q_k^{(n)} e^{v_k^1} / S^{(k)}\left(q^{(n)}\right)}, \tag{5.2}$$

and the iteration stops when the change from $q^{(n)}$ to $q^{(n+1)}$ is small. The intuition for (5.2) is straightforward: The numerator adjusts $q_j^{(n)}$ in the direction required by (5.1), while the denominator ensures that demand sums to 1.

As an example, a counterfactual scenario was generated using the parameter estimates from model M1, which reasonably allows inside goods to be closer substitutes to each other than to

the outside good. The example uses the market shares for the UK in the year 1999. The iteration (5.2) was implemented in a spreadsheet.

The price of Ford cars was raised by 20% in the counterfactual scenario. This kind of change allows the substitution patterns to be understood quite intuitively. Overall, the model is able to describe a rich pattern of cross-elasticities as shown in Figure 1, which shows a scatter plot of the change in demand, $\ln q^1 - \ln q^0$, against the base demand $\ln q^0$.

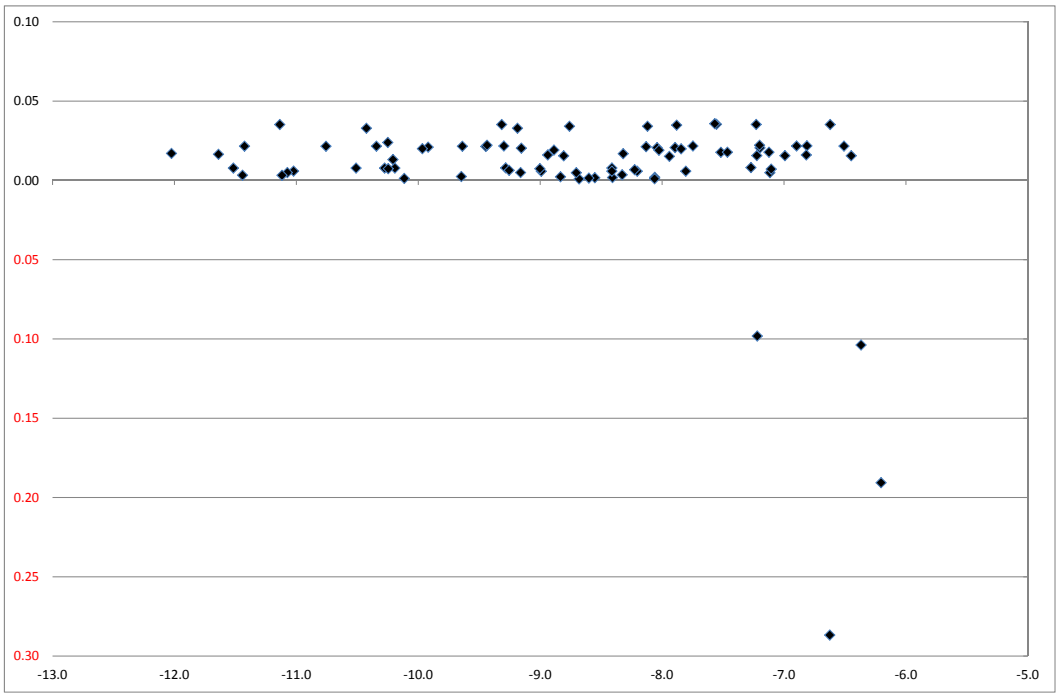

Figure 1: The change in log demand (vertical axis) against observed log demand (horizontal axis) for the 85 car models present on the UK market in 19999

The effect of the price increase on Ford cars is that the sales of Ford drop by 16%. The response in the sales of other car brands falls in two groups. The sales of Citroen, Honda, Mazda, Mitsubishi, Opel, Peugeot, Renault, Toyota, and VW increase between 1.8% and 2.9%, with sales of Mitsubishi increasing the most; these car brands have all been assigned the same quality category as Ford. The sales of Alfa Romeo, Audi, BMW, Daihatsu, Fiat, Honda, Hyundai, Mercedes, Nissan, Saab, Seat, Skoda, Suzuki, Volvo, Kia and Daewoo increase between 0.5% and 1.3%, with sales of Alfa Romeo increasing the least; these car brands have all been assigned to quality categories different from that of Ford. Overall, the sale of cars drops by 1.4%.

References

Aguirregabiria, V. and Mira, P. (2010) Dynamic discrete choice structural models: A survey *Journal of Econometrics* **156**(1), 38–67.

Berry, S., Levinsohn, J. and Pakes, A. (1995) Automobile Prices in Market Equilibrium *Econometrica* **63**(4), 841–890.

Correia, S. (2014) *REGHDFE: Stata module to perform linear or instrumental-variable regression absorbing any number of high-dimensional fixed effects*. Published: Statistical Software Components, Boston College Department of Economics.

Fosgerau, M. and de Palma, A. (2016) Generalized entropy models.

Fosgerau, M., Frejinger, E. and Karlstrom, A. (2013) A link based network route choice model with unrestricted choice set *Transportation Research Part B: Methodological* **56**, 70–80.

Knittel, C. R. and Metaxoglou, K. (2014) Estimation of Random-Coefficient Demand Models: Two Empiricists' Perspective *Review of Economics and Statistics* **96**(1), 34–59.

Kuminoff, N. V., Smith, V. K. and Timmins, C. (2013) The New Economics of Equilibrium Sorting and Policy Evaluation Using Housing Markets *Journal of Economic Literature* **51**(4), 1007–62.