Adam Woźniak,
Theodore Davies

Project 2

Advanced Microeconometrics

Submitted:
2023-12-13

1. Note to peer reviewers!

Hi! We're fully aware that this is a work in progress; we're having some issues with our code and thought it would be good to still submit something, so that we can see how other groups have approached the assignment. We would really really value your feedback!

2. Introduction

In this paper we investigate whether there is evidence for a home-bias in the demand for cars, and whether this home-bias affects the own-price elasticity of demand. To do so, we construct a Conditional Logit estimator, using characteristics of the cars as controls. We conclude that there is little evidence to support the existence of a home-bias in the present data.

3. Data

We employ a panel dataset covering a range of variables describing car demand in 30 countries over 5 years. This gives N=150 total markets, where a 'market' is defined as a country-year pair. For each market, there is data on J=40 cars (alternatives), including market share, a number of price variables, technical measures such as weight and horsepower, and manufacturing information such as nationality and brand. There are thus 6000 total observations (40 cars times 150 markets).

4. Econometric Theory

a) Utility and Conditional Choice Probability

We use three indices to structure our analysis, namely:

Markets $i = 1, ..., N$. $(N = 150.)$

Alternatives $j = 1, ..., J$. $(J = 40.)$

Households $h = 1, ..., H$. (Assumed that $H \to \infty$.)

Underpinning our demand model for cars is a household utility function, wherein the utility of household h from car j in market i is modelled as the product of $\mathbf{x}_{ij}$, a 1xK vector of characteristics of the car itself, and β_o, an unobserved Kx1 vector of parameters, plus ϵ_{ijh}, a scalar error term that is observable to the household but not the researcher. Formally:

$$u_{ijh} = \mathbf{x}_{ij}\beta_o + \epsilon_{ijh}, j = 1, ..., J. \tag{1}$$

Because we cannot observe the error term directly, we treat it as random, and therefore must make assumptions about its distribution. To facilitate Conditional Logit estimation, here we assume that the error term has a Type-I extreme value (Gumbel) distribution:

$$\textbf{Error distribution: } \epsilon_{ijh} \sim \text{Gumbel}(0, 1). \tag{2}$$

This is a relatively strong assumption that is challenging to test. The conditional choice probability for each alternative is the probability that a household chooses that particular car, given the characteristics of that car and the other 39 alternatives. Because of the assumption placed on the error term, it can be shown that this conditional choice probability is given by

$$\text{Pr(household } h \text{ chooses } j|\mathbf{x}_i) = \frac{\exp(\mathbf{x}_{ij}\beta_o)}{\sum_{k=1}^{J}\exp(\mathbf{x}_{ik}\beta_o)} = s_j(\mathbf{x_i}, \beta_o). \tag{3}$$

As households are assumed to be identical (β is household-invariant), and as the number of households in each market is assumed to be large, the conditional choice probability for a car will be equal to its market share. In other words, if each household has a 5% probability of choosing a car, then across the whole market it will be chosen 5% of the time. Equation 3 describes the conditional choice probability/market share function $s_j(\mathbf{x_i}, \beta_o)$ given the true (unknown) parameters; we can also describe the general case, where β is allowed to vary:

$$s_j(\mathbf{x}, \beta) = \frac{\exp(\mathbf{x}_j\beta)}{\sum_{k=1}^{J}\exp(\mathbf{x}_k\beta)} = \frac{\exp(\mathbf{x}_j\beta)}{\sum_{k=1}^{40}\exp(\mathbf{x}_k\beta)}, \beta \in \mathbb{R}^K. \tag{4}$$

From this, we can derive one of the key characteristics of a Conditional Logit estimator: the independence of irrelevant alternatives (IIA). The ratio of conditional choice probabilities for two distinct alternatives will simplify to

$$\frac{s_j(\mathbf{x}, \beta)}{s_k(\mathbf{x}, \beta)} = \frac{\exp(\mathbf{x}_j\beta)}{\exp(\mathbf{x}_k\beta)}, \tag{5}$$

which implies that the characteristics of all other alternatives are irrelevant to the preference between any two. In this model, my preference between a Ford Escort and a Skoda Octavia must be uninfluenced by whether a BMW M3 or Ferrari 458 is also available. In some contexts IIA fails; it is unclear whether it holds here.

b) Log-Likelihood and Estimation

We can assess candidate parameters by calculating the log-likelihood of obtaining the observed market shares y_{ij}, given the observed car characteristics and the candidate parameters. The log-likelihood contribution of each market (across all alternatives) is given by

$$l_i(\beta) = \sum_{j=1}^{J} y_{ij} \ln s_j(\mathbf{x}_i, \beta) = \sum_{j=1}^{40} y_{ij} \ln s_j(\mathbf{x}_i, \beta). \tag{6}$$

The final Conditional Logit estimator, $\hat{\beta}$, is the set of parameters that maximises the mean of the individual market log-likelihood contributions. To identify the parameters in our case, we will thus numerically minimize the negation of this function:

$$\hat{\beta} \in \underset{\beta \in \mathbb{R}^K}{\text{argmax}} \frac{1}{N} \sum_{i=1}^{N} l_i(\beta), \tag{7}$$

or in our case:

$$\hat{\beta} \in \underset{\beta \in \mathbb{R}^K}{\text{argmax}} \frac{1}{150} \sum_{i=1}^{150} l_i(\beta). \tag{8}$$

c) Characteristic Selection

For a Conditional Logit estimator to be identified, regressors must describe characteristics of the alternative (here the cars), rather than characteristics of the selector (here the markets), as parameters do not vary across selectors. We can see the extreme case from Equation 5 – if $\mathbf{x}$ contains only characteristics that are equal across $\mathbf{x}_j$ and $\mathbf{x}_k$, then the conditional choice probability of each alternative will always be equal. Because of this, the range of variables available to use in analysis is limited. For example, the parameter on income per capita (which is included in our dataset) would not be identified as it varies across markets but not alternatives. As a result, we here limit ourselves to the following characteristics:

Table 1: Car Characteristics

Variable	k	Description
logp	1	Log of price relative to per capita income.
home	2	Domestic car dummy.
cy	3	Cylinder volume or displacement (in cc).
hp	4	Horsepower (in kW).
we	5	Weight (in kg).
li	6	Average of three fuel efficiency measures (in liters per km).
brand	7-38	Brand name dummies (excluding ref. category for consistency).

d) Elasticity and Home-Bias

To estimate the home-bias, we will calculate the Average Partial Effect (APE). For each alternative in each market, the conditional choice probability will be calculated with home set to 1 and with it set to 0, with the other characteristics held constant at their initial values. The mean of the difference between these two probabilities gives us the APE. We can express this more formally:

$$\text{APE} = \frac{1}{40 \cdot 150} \sum_{j=1}^{40} \sum_{n=1}^{150} s_j(\mathbf{x}_{home}, \beta) - s_j(\mathbf{x}_{foreign}, \beta). \tag{9}$$

We are also interested in whether the home-bias affects own-price elasticity of demand. The own-price elasticity of demand is given by:

$$e_{ij}(x) = \frac{\delta s_j((x, \hat{\beta}))}{\delta p_{ij}} \frac{p_{ij}}{s_j((x, \hat{\beta}))}, \tag{10}$$

where p_{ij} denotes the price variable, here logp. To find the elasticity across the sample, we calculate the change in conditional choice probability (for each alternative in each market) produced by an incremental change in price, holding other variables constant, and then find the mean of these elasticities. We then estimate the elasticity in this way across two subsamples, cars in domestic markets and cars in foreign markets, to evaluate the impact of home-bias on demand elasticity.

5. Empirical Analysis

The coefficients estimated using Logit are set out in Table 2. None of the coefficients were found to be statistically significant (or close to being so). The estimated coefficient on home was positive, indicating an increased probability of buying a domestic car, but the standard error was high and thus the t-statistic very small. Indeed, the estimate of the APE, as shown in Table 3 was negligible. There was, however, a slight difference in the own-price demand elasticities between home and foreign cars, giving a very small amount of credence to the hypothesis that buyers are less sensitive to price if the car is produced in their own country.

6. Discussion and Conclusion

There are limitations to our findings; foremost among them methodological. The optimization algorithm used was highly sensitive to initial values. While the initial values used here allowed convergence, values nearby did not converge, while other (drastically different) values converged to (drastically) different final estimates. In future, a more robust approach would be appropriate. Moreover, ideally a multinomial choice model would allow for their to be information included about both the choice (cars) and the selector (markets), as there may be relevant information about the latter that feeds into the likelihood of selecting a certain model of car.

Through constructing a Logit estimator, we have not been able to reach a strong conclusion concerning the existence of a home-bias in car markets. Although our methodology should have enabled the use of car characteristics in estimating the conditional choice probability, it did not identify any statistically significant relationships here.

7. Results

Table 2: Parameter Estimates

Variable	$\hat{\beta}_k$	SE	t
logp	3.27e+07	9.70e+10	0.0003
home	5.48e+06	4.58e+11	0.0001
cy	3.95e+07	3.83e+11	0.0001
hp	3.99e+07	3.37e+11	0.0001
we	3.66e+07	2.55e+10	0.0014
li	2.89e+07	4.24e+10	0.0007

Note: *p <0.1,**p<0.05,***p<0.01. The 32 brand dummy estimates have been excluded for clarity.

Table 3: Average Partial Effect

	APE
home	0.000

Table 4: Own-Price Demand Elasticities

Group	Elasticity
All Cars	-97.5
Home Cars	-96.3
Foreign Cars	-98.1
(Home-Foreign)	*(1.8)*