UNIVERSITY OF COPENHAGEN
FACULTY OF SOCIAL SCIENCES

Advanced Microeconometrics

Project 3

Car Demand and Home Market Bias

Mikala Ann de Wet Bræmhøj (vrz848), Caroline Bergholdt Hansen and Julie Cathrine Krabek Sørensen

13 December 2023

Contents

1 Introduction **1**

2 Data **1**

3 Theory **1**

3.1 Model Specification and Conditional Logit 1

3.2 Maximum Likelihood 2

3.3 Parameter Interpretation 3

3.3.1 The Partial Effect at the Average (PEA) 3

3.3.2 The Marginal Willingness to Pay (MWP) 3

3.3.3 The Own-Price and Cross-Price Elasticities (OPE and CPE) 3

3.4 Inference 4

4 Analysis **4**

5 Discussion and Conclusion **5**

References **6**

Appendix **7**

1 Introduction

Home bias reflects consumers' inclination to favor vehicles from their region, indicating a preference for locally produced brands. This has implications for understanding consumer behavior and international trade dynamics in the car industry. This paper aims to investigate home bias in car demand in the European markets. This is done by estimating a conditional logit (CL) model. We find significant partial effects on average, which leads us to accept the hypothesis of home bias.

2 Data

Our data set spans 30 years (1970-1999) across the car market in five countries (Belgium, France, Germany, Italy, and UK), treating each country-year as a market indexed by i. It includes 6,000 entries, representing the market share and attributes of the top 40 selling cars. The dependent variable is the market share of each car in market i. Explanatory variables are log-transformed GDP per capita-adjusted car prices, a domestic car indicator (home), cylinder volume, weight, horsepower, fuel efficiency, and brand dummies.

3 Theory

The CL model is particularly suitable for problems where agents choices are based on observable features of each alternative. Since the data includes a range of variables capturing various attributes of cars, the CL is arguably well-suited to model this problem.

3.1 Model Specification and Conditional Logit

We consider a random utility model (RUM), typically used to model the choice among mutually exclusive alternatives. RUM involves the agent choosing j to maximize utility:

$$y_{ijh} = \underset{j \in \{1,\ldots,J\}}{\operatorname{argmax}} u_{ijh} \tag{1}$$

$$u_{ijh} = \mathbf{x}_{ij}' \boldsymbol{\beta}_o + \varepsilon_{ijh} \tag{2}$$

where u_{ijh} denotes the utility of household h in market i from buying car j. Together, equations (1) and (2) are an additive RUM. The errors are known to the households but are unobservable to the researcher, and assumed to be IID Extreme value (EV) distributed. By this assumption, conditional on x_i, the probability of j maximizing utility is:

$$\Pr(\text{household } h \text{ chooses car } j|\mathbf{x}_i) = \frac{\exp(\mathbf{x}_{ij}\boldsymbol{\beta}_o)}{\sum_{k=1}^{J} \exp(\mathbf{x}_{ik}\boldsymbol{\beta}_o)} := s_j(\mathbf{x}_i, \boldsymbol{\beta}_o) \tag{3}$$

This can be seen as the individual household's choice probability function and the market share function. s_j can further be extended to take any candidate parameter $\boldsymbol{\beta} \in \mathbb{R}^K$. The actual choices y_{ij} are indicated by the market share observed in the data. An important restriction for CL is the independence from irrelevant alternatives (IIA) assumption:

$$\frac{p_j(\mathbf{x}_j)}{p_h(\mathbf{x}_h)} = \frac{\exp(\mathbf{x}_j\boldsymbol{\beta})}{\exp(x_h\boldsymbol{\beta})} = \exp\left((\mathbf{x}_j - \mathbf{x}_h)\boldsymbol{\beta}\right) \tag{4}$$

This implies that the ratio between the probability of two choices is not affected by adding another alternative or changing a third alternative.

3.2 Maximum Likelihood

We estimate $\hat{\boldsymbol{\beta}}$ by maximum likelihood estimation (MLE). The log-likelihood for i is:

$$\ell_i(\boldsymbol{\beta}) = \sum_{j=1}^{N} y_{ij} \ln s_j(\mathbf{x}_i, \boldsymbol{\beta}) \tag{5}$$

We derive an estimator by minimizing the average negative log-likelihood contributions across N markets for the candidate parameter vectors $\boldsymbol{\beta}$:

$$\hat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta} \in \mathbb{R}^K}{\operatorname{argmin}} -\frac{1}{N} \sum_{i=1}^{N} \ell_i(\boldsymbol{\beta}) \tag{6}$$

As the model is nonlinear, an analytical solution to (6) is unavailable. Consequently, estimates for $\hat{\beta}$ are obtained through numerical optimization. The BFGS optimizer, which depends on numerical gradients and approximations of Hessians, is employed for this purpose. Table 1 shows the most important assumptions for $\hat{\boldsymbol{\beta}}$[1]. 12.2 (1) holds if $\boldsymbol{\beta}_o$ is identified such that any other parameterization yields a different density than the true one. 12.2 (2) is taken as given. As the assumption of an IID EV error term implies continuity of the error term and, thus, of $\ell_i(\boldsymbol{\beta})$ as well, 12.2 (3) holds given the assumption about the distribution of the error term holds. Thus the MLE is consistent. 12.3 (1) holds as $\boldsymbol{\beta}_o$ is identified and therefore known to be an interior solution. The functional form of $\ell(\boldsymbol{\beta})$ lives up to 12.3 (2). The M-estimator is then asymptotic normally distributed:

$$\sqrt{N}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_o) \xrightarrow{d} N(\mathbf{0}, \mathbf{A}_o^{-1}\mathbf{B_o}\mathbf{A}_o^{-1}), \mathbf{A}_o = E[H(w_i, \boldsymbol{\beta}_o)], \mathbf{B}_o = E[s(w_i, \boldsymbol{\beta}_o)s(w_i, \boldsymbol{\beta}_o)']$$

where $\boldsymbol{\beta}_o$ are the true parameters and $w = (\mathbf{y}, \mathbf{x})$. By the analogy principle, we substitute expectations with averages and true parameters with corresponding estimates, such that:

$$\hat{\mathbf{A}} = \frac{1}{N} \sum_{i=1}^{N} \hat{\mathbf{H}}_i = \mathbf{H}(\mathbf{w}, \hat{\boldsymbol{\beta}}), \qquad \hat{\mathbf{B}} = \frac{1}{N_i} \sum_{i=1}^{N_i} \hat{s}_i \hat{s}_i{}', \quad \hat{s}_i := \mathbf{s}(\mathbf{y}_i, \mathbf{x}_i, \hat{\boldsymbol{\beta}})$$

[1] Further technical assumptions necessary are available in Wooldridge (2010)

For MLE, the Information Matrix Equality holds, which means that $A_o = B_o$, and the variance formula simplifies to $Avar(\hat{\boldsymbol{\beta}}) = \mathbf{A}_o^{-1}/N$. Nonetheless, we compute standard errors using the sandwich variance as it is arguably more robust.

3.3 Parameter Interpretation

As the coefficients of a RUM say something about the utility of the agents whose behavior we are modeling, they are scale-indeterminate, implying that their magnitude alone does not directly provide any useful information. Thus, we must compute various measures that allow us to interpret their magnitude. We focus on three such measures: the partial effect at the average, the marginal willingness to pay, and the own-price/cross-price elasticities.

3.3.1 The Partial Effect at the Average (PEA)

The PEA is defined as the effect of an explanatory variable on the response probability in the hypothetical case that the remaining regressors are equal to their sample averages. In this context, we are particularly interested in the binary regressor $Home_{ij}$, whose PEA tells us something about the effect of a car (hypothetically) changing from being produced non-domestically to being produced domestically on the market share of the average car:

$$PEA(Home_{ij}) = \frac{1}{N}\sum_{i=1}^{N}\{\mathbb{P}(y = j|\overline{\mathbf{x}}_i, Home_{ij} = 1) - \mathbb{P}(y = j|\overline{\mathbf{x}}_i, Home_{ij} = 0)\} \tag{7}$$

where $\overline{\mathbf{x}}_i$ is sample averages of the remaining regressors, and j is the alternative in question.

3.3.2 The Marginal Willingness to Pay (MWP)

Alternatively, we can compute the MWP. This is possible because there is a price variable in the model, which enables us to quantify the effect of a non-monetary variable, such as $Home_{ij}$, in monetary terms. More specifically, we calculate how much the price must change to keep the observed part of the utility of a car produced domestically constant:

$$MWP = \left|\frac{\hat{\boldsymbol{\beta}}^{price}}{\hat{\boldsymbol{\beta}}^{Home}}\right| \tag{8}$$

The MPW, calculated as the marginal rate of substitution between domestically and foreign produced cars, captures household's WTP for a domestically produced car.

3.3.3 The Own-Price and Cross-Price Elasticities (OPE and CPE)

Finally, a third option is to compute the OPE and the CPE. Both of these measures are elasticites, generally computed as: elasticity $= \frac{dy}{y}\frac{x}{dx}$. The OPE and CPE are defined as:

$$OPE = \frac{\text{\% change in demand}}{\text{\% change in own price}}, CPE = \frac{\text{\% change in demand}}{\text{\% change in price of another good}} \tag{9}$$

Because we do not have access to a variable that directly captures the quantity demanded, we instead use the dependent variable, the market share of car j in market i, as a measure of this and calculate both the OPE and CPE for domestic and foreign cars, respectively.

3.4 Inference

Hypothesis testing for CL models involves the size of measures such as the PEA, MWP, or OPE/CPE rather than the 'underlying' coefficients. When the corresponding test statistics, we have to account for them being transformations of the estimated coefficients. For this, we can use the delta method. Given the partial effects $PE_j(\mathbf{x}) := h(\boldsymbol{\beta}_o)$ with the corresponding estimated partial effects $\widehat{PE}_j(\mathbf{x}) := h(\hat{\boldsymbol{\beta}})$, it can be shown that if the estimated coefficients are $\sqrt{N}$-asymptotically normally distributed with mean $\mathbf{0}$ and variance $\mathbf{V}$, the estimated partial effects are also $\sqrt{N}$-asymptotically normally distributed with mean $\mathbf{0}$ and variance $\nabla_{\boldsymbol{\beta}} h(\boldsymbol{\beta}_o)\mathbf{V}\nabla_{\boldsymbol{\beta}} h(\boldsymbol{\beta}_o)'$ (Wooldridge (2010)). Using the consistent estimator of the variance of the PE, $\hat{\mathbf{V}} = \nabla_{\boldsymbol{\beta}} h(\hat{\boldsymbol{\beta}})\hat{\boldsymbol{\Sigma}}\nabla_{\boldsymbol{\beta}} h(\hat{\boldsymbol{\beta}})'$, where $\hat{\boldsymbol{\Sigma}}$ is an estimator of the covariance matrix for $\hat{\boldsymbol{\beta}}$, enables us to test the hypothesis of home bias by:

$$\mathrm{H}_0 : h(\boldsymbol{\beta}_o) = 0, \qquad \mathrm{H}_A : h(\boldsymbol{\beta}_o) \neq 0 \tag{10}$$

That is, we test the null hypothesis that there is no home bias against the alternative that there is home bias. This is a two-sided test, which we test using a 5% significance level.

4 Analysis

To investigate the hypothesis of home bias, we specify two versions of our model: a basic model (1) and one where we add an interaction term between price and home (2) so as to capture heterogeneous effects of prices across domestic and foreign cars. The result of our analysis is in Table 2. The PEA for the home variable is positive and statistically significant in both models, implying that higher prices lead to lower demand and that consumers are positively biased toward domestically-produced cars. More specifically, the PEA for home implies that a domestically-produced car, on average, will have a market share that is approximately 3.7%-4.7% higher than that of an identical, but foreign-produced, car. Hence, the findings seemingly confirm our hypothesis of home bias, as we reject the null hypothesis in (10).

In the basic model specification (without an interaction term), neither the MWP nor the price elasticities OPE and CPE are significant, and so, it does not make much sense to conclude upon their signs/magnitudes. In the extended model, however, the OPE and CPE are both statistically significant. In this model, the OPE is found to be ~ -0.1660, implying that a 1% increase in the price of a car leads to a decrease in the market share of that car of $\sim 16.60\%$. In addition, the CPE is ~ 0.0043, which means that an increase in the price of a car has a positive effect on the market share of other cars. Both these findings correspond to our economic intuition.

5 Discussion and Conclusion

Several concerns are worth highlighting. First of all, the IIA assumption introduced previously is unlikely to hold when the alternatives are similar, as in our analysis. E.g., considering the introduction of a new car model l, we would expect this to affect the probability ratio $\frac{p_j(x_j)}{p_h(x_h)}$; for the ratio to remain unchanged requires that the probability that a household chooses each of the two existing car models is affected equally by the introduction of car l. Say that car j is an electric car, car h is a non-electric car, and car l is an electric car, we would expect $p_j(x_j)$ to decrease by more than $p_h(x_h)$, since car j is a closer substitute to car l. This issue could have been resolved, or at least mitigated, by using a nested logit model instead.

As evident by the model in equation (2), CL also assumes that household preferences $\boldsymbol{\beta}_o$ are constant across both time and markets. This would require that consumers have not changed tastes in terms of cars within the 30-year time frame in question and that tastes are identical across the five countries. Both of these assumptions are unlikely to hold. We would, e.g., expect consumers in the late 1990s to prefer environment-friendly cars to a higher degree than consumers in the early 1970s. Moreover, we suspect that preferences related to specific car features differ across countries, due to differences in geographic and urban factors such as weather conditions and infrastructure. An estimation method that might be better at capturing such heterogeneity is mixed logit estimation.

A final concern regards the standard errors. We have used the delta method to compute standard errors for hypothesis testing, but we could have instead used bootstrap standard errors, which are robust to misspecification of the covariance matrix as well as finite sample bias. A limitation of these standard errors, in turn, is the computational cost associated with obtaining them, which is why we have chosen not to compute them. Despite these concerns, the findings of this paper support the hypothesis of home bias.

References

Wooldridge, Jeffrey M (2010). *Econometric analysis of cross section and panel data*. MIT press.

Appendix

Table 1: Assumptions

Theorem	Assumption	Cont.	A.N.
12.2 (1)	$\boldsymbol{\beta}_o$ uniquely minimizes $E[-\ell(\boldsymbol{\beta})]$	*	
12.2 (2)	$\boldsymbol{\beta} \subseteq \mathbb{R}^P$ compact (i.e. closed and bounded)	*	
12.2 (3)	$\ell(\boldsymbol{\beta})$ is a continuous function in $\boldsymbol{\beta}$	*	
12.3 (1)	$\boldsymbol{\beta}_o$ is an interior solution to $\mathbb{R}^P$	*	*
12.3 (2)	$\ell(\boldsymbol{\beta})$ is continuous and twice differentiable on the interior of the compact parameter space	*	*

Table 2: Informative Estimates

	(1)	(2)
PEA	0.0372***	0.0470***
	(0.0011)	(0.0047)
MWP	8.3696	11.7134
	(6.0052)	(10.0252)
OPE	-0.1585	−0.1660**
	(0.1134)	(0.0897)
CPE	0.0041	0.0043**
	(0.0029)	(0.0027)
Interaction	No	Yes

Note: Standard-errors are in parenthesis.
Significance level: *** 0.01 ** 0.05 * 0.1.