Quebec Energy Case

In order to estimate Binary Logit models, we consider a subset of the complete dataset with two alternatives, instead of nine, for a total of 2695 households. The two alternatives are:

1. Dual energy/Electricity (dee)
2. Electricity/Electricity (ee)

Some general statistics on the dataset are reported in table 1.

Binary Logit with Generic Attributes

Files to use with BIOGEME:
Model file: *BL_generic_4-7.mod*
Data file: *QE_data_bin.dat*

We assume that households choose the type of space-water heating systems as a function of the fixed and operating cost associated with the available alternatives. We wish to translate this simple theory of economic choice behavior into a concrete model suitable for the mathematical analysis of energy choice. In order to be consistent with the theory of random utility models, we will estimate a Binary Logit model in which the utility of the alternatives is the sum of a deterministic part and of a random term. We define the deterministic part of the utility functions including the alternative specific constants (ASC), the attribute *coutm* (annual operating cost in thousand dollars) and the attribute *coutfix* (annual fixed cost in thousand dollars). These two attributes are weighted with their respective generic coefficients, namely β_{coutm} and β_{coutfix}. More formally, the utility functions are expressed as follows:

$$\begin{aligned} V_{\text{dee}} &= \text{ASC4} + \beta_{\text{coutm}}\text{coutm4} + \beta_{\text{coutfix}}\text{coutfix4} \\ V_{\text{ee}} &= \text{ASC7} + \beta_{\text{coutm}}\text{coutm7} + \beta_{\text{coutfix}}\text{coutfix7} \end{aligned}$$

Variable	Obs	Mean	Std. Dev.	Minimum	Maximum
secteur	2695	1.998706	1.316036	1	4
hddm	2695	4.844169	0.504153	4.359	7
choice	2695	6.69644	1.263997	1	9
edanconv	2695	1.941748	2.898271	0	8
emhnhabi	2695	2.946602	2.888014	2	13
emdndaco	2695	8.158252	3.14563	1	12
eminpiec	2695	6.178964	1.941691	1	24
empnpers	2695	3.086408	1.226122	1	9
emrnprlo	2695	1.102913	0.3038939	1	2
surf	2695	1.656537	0.7409359	0.3	3.25
age	2695	4.211974	1.391425	2.1	8.4
rev	2695	4.253641	2.097062	0.5	7.5
coutm4	3090	1.496247	0.3656796	0.460287	4.009572
coutm7	3090	1.43987	0.3478127	0.453367	3.671097
coutfix4	3090	0.9915023	0.1098705	0.138219	1.385848
coutfix7	3090	0.2796577	0.2119302	0.065063	0.912808
coutfiy4	3090	4.247216	2.199511	0.0790607	10.39386
coutfiy7	3090	1.266823	1.358479	0.0328052	6.846063
aval4	3090	1	0	1	1
aval7	3090	1	0	1	1

Table 1: General Statistics

```
[Choice]
choice

[Beta]
// Name   Value     LowerBound UpperBound  status (0=variable, 1=fixed)
ASC4         0.0          -100.0     100.0         1
ASC7         0.0          -100.0     100.0         0
b1coutm      0.0          -100.0     100.0         0
b1coutfix    0.0          -100.0     100.0         0

[Utilities]
// Id  Name  Avail  linear-in-parameter expression
4   Alt4    aval4   ASC4 * one +  b1coutm * coutm4 + b1coutfix * coutfix4
7   Alt7    aval7   ASC7 * one +  b1coutm * coutm7 + b1coutfix * coutfix7
```

Figure 1: Snapshot from the BIOGEME code

One of the alternative specific constants (arbitrarily chosen, for example the one associated with *dual energy/electricity*) is normalized to zero for identification purpose.

A snapshot from the BIOGEME code is shown in figure 1. It illustrates how we translate the specification of the deterministic part of utilities into the syntax of BIOGEME. The section `[Choice]` indicates the dependent variable in our dataset, that is the variable related to the mutually exclusive set of alternatives. The section `[Beta]` lists the parameters which we intend to use in the utilities. For example, ASC4 is the alternative specific constant for the alternative number 4. Its status is set to 1 which means that the parameter is kept fixed at its starting value which is zero. This is how we normalize one of the ASC. In the section `[Utilities]`, we define the expression of utilities. See BIOGEME user manual for further details on BIOGEME.

The estimation results for this first model are presented in table 2. Given our specification, the positive sign of the alternative specific constant indicates that, everything else being equal, the alternative *electricity/electricity* has a higher utility than the alternative *dual energy/electricity*

Estimation results				
Variable number	Variable name	Coefficient estimate	Robust standard error	Robust *t statistic*
1	*ASC7*	2.166	0.717	3.020
2	$\beta_{1coutfix}$	-0.270	0.970	-0.278^*
3	β_{1coutm}	-24.311	1.175	-20.675

Summary statistics
Number of observations $= 2695$
$\mathcal{L}(0) = -1868.03$
$\mathcal{L}(\hat{\beta}) = -218.682$
Likelihood ratio test $= 3298.7$
$\bar{\rho}^2 = 0.882935$

Table 2: Estimation results

option. In other words, there is an intrinsic preference towards the alternative *electricity/electricity* that is not captured by the other explanatory variables.

The parameter related to the fixed cost (*coutfix*) is not significantly different from zero. BIOGEME indicates that by putting an asterisk (*) on the right of the column which contains robust t-tests. (The robust t-test account for non severe misspecification errors, see Biogeme manual as starting reference) The parameter associated with operating cost (*coutm*) is significant and has a negative sign sign (as expected), implying that high operating cost decrease the utility of the corresponding alternative.

Binary Logit with Alternative-Specific Coefficients

Files to use with BIOGEME:
Model file: *BL_specific_4-7.mod*
Data file: *QE_data_bin.dat*

In order to analyze if the coefficient of the cost attributes is alternative-specific, we estimate one cost parameter for each alternative. The deterministic part of the utilities are then given by

$$\begin{aligned} V_{dee} &= ASC_{dee} + \beta_{1_coutm}\text{coutm4} + \beta_{1_coutfix}\text{coutfix4} \\ V_{ee} &= ASC_{ee} + \beta_{2_coutm}\text{coutm7} + \beta_{2_coutfix}\text{coutfix7}. \end{aligned}$$

The estimation results for this second model are presented in Table 3. This specification relaxes the assumption that the impact of fixed and operating cost are the same for each alternative. The *coutfix* parameters are still not significantly different from zero. The coefficients associated with the *coutm* variable are estimated with the expected signs and are significantly different from zero. The reader will learn in the specification testing case study how to the whether β_{1coutm} is significantly different than β_{2coutm}.

Estimation results				
Variable number	Variable name	Coefficient estimate	Robust standard error	Robust *t statistic*
1	ASC7	11.183	4.121	2.713
2	$\beta_{1coutfix}$	5.724	4.330	1.321*
3	β_{1coutm}	-35.730	7.880	-4.534
3	$\beta_{2coutfix}$	0.0775	0.294	0.263*
3	β_{2coutm}	-38.473	8.026	-4.793
Summary statistics				
Number of observations = 2695				
$\mathcal{L}(0) = -1868.03$				
$\mathcal{L}(\hat{\beta}) = -193.758$				
Likelihood ratio test = 3348.55				
$\bar{\rho}^2 = 0.896277$				

Table 3: Estimation results

Binary Logit with Socio-Economic Characteristics

Files to use with BIOGEME:
Model file: *BL_socio_econ_4-7.mod*
Data file: *QE_data_bin.dat*

It is reasonable to assume that people make choices not only in relation to the attributes that characterize the alternatives but also depending on some personal characteristics or socio-economic indicators. The availability of individual-specific information gives us the opportunity to model in part the heterogeneity present in the population we are interested in. We modify the previous model by adding in the utilities the income (*rev*) of the households, and the *age* as extra explanatory variables:

$$\begin{aligned} V_{\text{dee}} &= \text{ASC}_{\text{dee}} + \beta_{\text{coutm}}\text{coutm4} + \beta_{\text{coutfix}}\text{coutfix4} \\ V_{\text{ee}} &= \text{ASC}_{\text{ee}} + \beta_{\text{coutm}}\text{coutm7} + \beta_{\text{coutfix}}\text{coutfix7} + \\ &\quad \beta_{7_\text{rev}}\text{rev} + \beta_{7_\text{age}}\text{age} \end{aligned}$$

Please note that, as these variables do not vary on the alternatives, we need to normalize one alternative to zero and interpret the estimated coefficients for the remaining alternatives with respect to the reference alternative, which is arbitrarily *dual electricity/ electricity*.

Since income and age do not vary on the alternatives, and as only difference in utility matter, we have normalized the alternative number 4 to zero and therefore only two extra parameters need to be estimated. Note that in BIOGEME the parameters $\beta_{7\text{age}}$ and $\beta_{7\text{rev}}$ have only been included in the second utility function (alternative number 7). We could have also specified the first utility function (alternative number 4) with $\beta_{4\text{rev}}$ and $\beta_{4\text{age}}$, and fix them to zero in `[Beta]` section, as we did for ASC4.

We report in table 4 the corresponding estimation results. The β parameters of the *rev* and *age* variables are interpreted as the impact of income and age on the likelihood that a decision maker will choose *electricity/-electricity*, with respect to the reference alternative *dual energy/electricity*.

Estimation results				
Variable number	Variable name	Coefficient estimate	Robust standard error	Robust *t statistic*
1	*ASC7*	5.757	1.124	5.119
2	$\beta_{1\mathrm{coutfix}}$	0.159	0.944	0.168^*
3	$\beta_{1\mathrm{coutm}}$	-23.799	1.313	-18.093
3	$\beta_{7\mathrm{age}}$	-0.552	0.084	-6.528
3	$\beta_{7\mathrm{rev}}$	-0.164	0.084	-1.940^*
Summary statistics				
Number of observations $= 2695$				
$\mathcal{L}(0) = -1868.03$				
$\mathcal{L}(\hat{\beta}) = -202.199$				
Likelihood ratio test $= 3331.67$				
$\bar{\rho}^2 = 0.891758$				

Table 4: Estimation results

A negative sign of the *rev* parameter can be interpreted as the higher the income of the household the lower the likelihood for choosing the alternative *electricity/electricity* compared to the *dual energy/electricity* heating system. For the *age* variable the coefficient is negative and statistically significant. Older households are less likely to choose a system of *electricity/electricity* compared to the *dual energy/electricity* heating system.