Michel Bierlaire.
TAs: Cloe Cortes, Pavel Ilinov, Evangelos Paschalidis, Anne-Valérie Preto, Negar Rezvany

Introduction (correction) — September 10, 2024

Question 1: Electric vehicles

1. Estimate the parameters π_1, π_2 and π_3 using maximum likelihood estimation, where:

$$\begin{aligned}\Pr(\text{EV} = \text{yes} \mid \text{income} = \text{low}) &= \pi_1,\\ \Pr(\text{EV} = \text{yes} \mid \text{income} = \text{medium}) &= \pi_2,\\ \Pr(\text{EV} = \text{yes} \mid \text{income} = \text{high}) &= \pi_3.\end{aligned} \tag{1}$$

The likelihood function is

$$\mathcal{L}^*(\pi) = \pi_1^{15}(1-\pi_1)^{200}\pi_2^{50}(1-\pi_2)^{450}\pi_3^{135}(1-\pi_3)^{150}.$$

The log likelihood function is

$$\begin{aligned}\mathcal{L}(\pi) =& \ln(\mathcal{L}^*(\pi))\\ =& 15\ln(\pi_1) + 200\ln(1-\pi_1)\\ &+ 50\ln(\pi_2) + 450\ln(1-\pi_2)\\ &+ 135\ln(\pi_3) + 150\ln(1-\pi_3).\end{aligned}$$

In order to find the maximum, we calculate the derivatives, and find the value of $\widehat{\pi}$ such that the derivatives are zero[1]:

$$\frac{\partial \mathcal{L}}{\partial \pi_1} = 0 \Longleftrightarrow \frac{15}{\widehat{\pi}_1} - \frac{200}{1-\widehat{\pi}_1} = 0 \Longleftrightarrow \widehat{\pi}_1 = \frac{15}{215} = 0.0698.$$

Similarly, we obtain

$$\begin{aligned}\widehat{\pi}_2 &= \frac{50}{500} = 0.1,\\ \widehat{\pi}_3 &= \frac{135}{285} = 0.474.\end{aligned}$$

Note that these values are the ratio, for each income category, of the number of individuals owning an electrical vehicle, and the total number of individuals in the income category.

[1] In principle, we should also check that the second derivatives matrix is negative definite, to guarantee that it corresponds to a maximum

2. The fact that the value of the parameters increases with income is consistent with the intuition.

3. The final log likelihood of the model is

$$\begin{aligned}\mathcal{L} =& 15\ln(0.0698) + 200\ln(1-0.0698)\\ &+ 50\ln(0.1) + 450\ln(1-0.1)\\ &+ 135\ln(0.474) + 150\ln(1-0.474)\\ =& -414.097.\end{aligned}$$

4. Under the H_0 hypothesis that $\pi_1 = \pi_2$, the log likelihood function of the restricted model is

$$\begin{aligned}\mathcal{L}_1(\pi) =& 15\ln(\pi_1) + 200\ln(1-\pi_1)\\ &+ 50\ln(\pi_1) + 450\ln(1-\pi_1)\\ &+ 135\ln(\pi_3) + 150\ln(1-\pi_3).\end{aligned}$$

   The maximum likelihood estimates of the parameters are

$$\begin{aligned}\widehat{\pi}_1 &= 0.0909,\\ \widehat{\pi}_3 &= 0.474.\end{aligned}$$

   The value of the log likelihood function at the maximum is $\mathcal{L}_1(\widehat{\pi}) = -414.967$.

   In order to test the hypothesis, we perform a likelihood ratio test. Under H_0, the quantity

$$-2(\mathcal{L}_1 - \mathcal{L}) = 1.74$$

   follows a χ^2 distribution with $3-2=1$ degree of freedom. As the 95% quantile of the distribution is 3.84, we cannot reject the H_0 hypothesis at that level. As the 90% quantile is 2.71, the hypothesis cannot be rejected at this level either. Therefore, we decide to keep the restricted model. Indeed, the hypothesis that they are both equivalent is accepted, and we prefer the most parsimonious one, with the least number of parameters.

5. Under the hypothesis that $\pi_1 = \pi_2 = \pi_3$, the log likelihood function of the restricted model is

$$\begin{aligned}\mathcal{L}_2(\pi) =& 15 \ln(\pi_1) + 200 \ln(1-\pi_1) \\ &+ 50 \ln(\pi_1) + 450 \ln(1-\pi_1) \\ &+ 135 \ln(\pi_1) + 150 \ln(1-\pi_1).\end{aligned}$$

The maximum likelihood estimate of the parameter is

$$\widehat{\pi}_1 = 0.2.$$

The value of the log likelihood function at the maximum is $\mathcal{L}_2(\widehat{\pi}) = -500.402$.

In order to test the hypothesis, we perform a likelihood ratio test against the restricted model that we have decided to accept. Under H_0, the quantity

$$-2(\mathcal{L}_2 - \mathcal{L}_1) = 170.87$$

follows a χ^2 distribution with $2-1=1$ degree of freedom. As the 95% quantile of the distribution is 3.84, we can reject the H_0 hypothesis at that level. As the 99% quantile is 6.63, the hypothesis can also be rejected at this level. Therefore, we decide to keep the model with two parameters. Indeed, the hypothesis that they are both equivalent is rejected, and we prefer the model that fits the data best.

6. As we have decided to keep the model such that $\pi_1 = \pi_2$, the market share of EV is predicted by our model as

$$W_{\text{EV}} = W_{\text{low}}\widehat{\pi}_1 + W_{\text{medium}}\widehat{\pi}_1 + W_{\text{high}}\widehat{\pi}_3,$$

where W_k is the share of the population in income category k. Therefore, the market share in this scenario is 29.2%.

7. No, we could not have used linear regression in this case. The main motivation behind using discrete choice models is the fact that the dependent variable is discrete (to have or not to have an electric vehicle). In linear regression, the dependent variable must be continuous.