Assignment Advanced Marketing Models (FEM21024) ©Econometric Institute, Erasmus University Rotterdam

Dennis Fok

2019-2020

1 General information

This assignment corresponds to 10% of the final grade, the exam corresponds to the remaining 90%. The assignment should be done individually. If you do not hand in the assignment on time, you will receive a 1. Note that in this case you can still take the exam.

Any programming language can be used. You are not allowed to use pre-programmed packages that implement (variants of) a mixture model. It is allowed to use packages that help you process data, do basic linear algebra, calculate density functions, etc.

Deadline for this assignment is **Wednesday, December 11, 23.59**. Use Canvas to hand in the assignment (pdf with answers and separate file with programming code). Questions about the assignment can only be asked through the discussion board on Canvas.

2 The model

We consider a model for the sales of a particular product, measured at the store level at a weekly frequency. Denote the sales of the product in store $i = 1, \dots, N$ in week $t = 1, \dots, T$ by S_{it}. The sales are explained by the price of the product at time t, that is, p_t. All stores have the same price in any given week. We specify the following heterogeneous model

$$\log S_{it} = \alpha_i + \beta_i \log p_t + \varepsilon_{it}, \qquad \varepsilon_{it} \sim N(0, 1).$$

The store specific parameters are assumed to follow a mixture distribution with K segments/clusters.[1] We use $C_i \in \{1, \dots, K\}$ to denote the (unobserved) cluster to which store i belongs. Denote $\Pr[C_i = c] = \pi_c,\ c = 1, \dots, K$ with $\pi_c \geq 0$ and $\sum_{c=1}^{K} \pi_c = 1$. For

[1] The variance of the error term is set to 1 in order to simplify this assignment. In general this is of course not a reasonable assumption.

stores in segment c the intercept and the price elasticity equal $\alpha^{[c]}$ and $\beta^{[c]}$, respectively. Therefore

$$\alpha_i = \alpha^{[c]} \text{ and } \beta_i = \beta^{[c]} \text{ if } C_i = c.$$

Finally, denote $\theta = (\alpha^{[1]}, \beta^{[1]}, \ldots, \alpha^{[K]}, \beta^{[K]})'$ and $\pi = (\pi_1, \ldots, \pi_K)'$.

3 Assignments/questions

In this assignment you need to implement the EM algorithm to estimate the parameters of the above model. Your code should (at least) contain the functions specified in the questions below. In the end you need to apply your code to a specific data set (see Canvas). The exact data set that you need to use depends on your student number. For example, if your student number is 123456, you will need to use the data `data123456.csv`.

You need to hand in your code and a short pdf file containing the answers to the questions 1 and 4 and the estimates/standard errors/log likelihood values and insights of question 9. It is wise to test each function that you create with the provided data and with test cases you create yourself. Do not hand in these tests.

Questions

1. Give the log-likelihood function of the above model as well as the complete-data log-likelihood function in terms of θ and π.

2. Program a function called `LogL` that evaluates the log-likelihood function of the above model. This should be a function that allows you to input θ, π, y and X and outputs the (scalar) log-likelihood value, where

   - y ($N \times T$ matrix of log sales)
   - X ($T \times 2$ matrix of a constant and the log prices)

3. Program a function called `EStep` to do the E-step of the EM algorithm, that is, make a function that takes as input θ, π, y and X and returns the $N \times K$ matrix of conditional cluster probabilities, where the (i, k) entry equals $\mathrm{E}[C_i = k | y, X, \theta, \gamma]$. Note: be careful to avoid under- or overflow in the calculations (eg. taking exponents of very negative or very large numbers).

4. Argue why the M-step of the EM algorithm comes down to (1) solving K Weighted Least Squares [WLS] problems and (2) calculating a (closed-form) solution for π. What are the weights here?

5. Program a function called `MStep` that performs the above mentioned estimations. This function should have as input

   - W ($N \times K$ matrix of conditional cluster probabilities)

- y ($N \times T$ matrix of log sales)
- X ($T \times 2$ matrix of a constant and the log prices)

and as output new estimates of θ and π.

6. Program a function called `EM` that iterates the E and M step until the weights do not change anymore given a chosen tolerance. Think about how to set starting values (you can initialize the weights or the parameters).
Input: K, y and X.
Output: estimates of θ and π

7. Write a function called `Estimate` to perform the estimation of the model parameters for a given value of K. This function calls `EM` to run the actual estimation. To avoid ending up in a bad local optimum perform 10 different calls to `EM` and select the solution with the highest log likelihood value.

8. Finally, extend the function `Estimate` with code to estimate the variance matrix of the parameter estimates. In general, for maximum likelihood estimation the variance can be estimated by inverting -1 times the hessian matrix of the log-likelihood function at the parameter estimates. The non-negativity and sum to 1 restrictions on the π-parameters complicate this general statement. These complications can be circumvented by reparametrizing the model such that no restrictions are needed. To this end we define the $(K-1) \times 1$ vector γ as a transformation of π, that is,

$$\gamma_c = \log(\pi_c) - \log(\pi_K), \text{ for } c = 1, \ldots, K-1,$$

and the other way around

$$\pi_c = \frac{\exp(\gamma_c)}{1 + \sum_{k=1}^{K-1} \exp(\gamma_k)}, \text{ for } c = 1, \ldots, K-1; \qquad \pi_K = \frac{1}{1 + \sum_{k=1}^{K-1} \exp(\gamma_k)}.$$

We next see the log-likelihood as a function of θ and γ.

   (a) Program a new log likelihood function, where now θ and γ are the input parameters. Avoid duplicating work by calling `LogL` from this function.
   (b) Use a pre-programmed routine in your programming language to numerically approximate the hessian of this function at the estimated parameter values for θ and γ.
   (c) Estimate the covariance matrix by inverting $-1\times$ this hessian.
   (d) Extract standard errors for θ from this matrix.

9. Apply all your code to the given data set in `dataXX.csv`. Run your code for $K = 2, 3$ and 4 segments. For each value of K report the estimated parameters (θ and π), standard errors (only θ) and log-likelihood value. Also comment on any convergence problems if you encounter them. On the basis of these results, which value of K do you prefer?