LAB REPORT
AH2304: Individual Choice Modeling

Instructors:
shiva.habibi@abe.kth.se
oskar.blom@abe.kth.se

INTRODUCTION

The lab sessions and final report are intended to increase your knowledge and familiarity with discrete choice models by estimating MNL and mixed logit and developing your skills in the interpretation and analysis of related results. The goals are to:

- Implement maximum likelihood estimation
- Experiment with model specification, comparison among different models and try to come up with a MNL model that is as good as possible
- Implement mixed logit and estimate using Maximum Simulated Likelihood. Discuss relevant results, and compare different mixed logit models with different specifications and also with MNL model

The results from the labs are documented in a report that is graded.

Matlab is available for all students at KTH via
http://www.kth.se/en/student/support/itsc/progdist/welcome-to-kth-program-distribution-1.322700

THE DATA

The data provided in the lab are under taken from a travel survey in the Stockholm region in September and October 2004 and 2006, before and after congestion charging was introduced in Stockholm. The choice set contains the most common travel modes for journey to work: walk, bike, car and public transport. It is worthwhile to note that only work trips on weekdays are selected for this dataset. The data file is stored in “rvu_data_workNH.csv” and is loaded and processed for you with the provided Matlab files. In Table 1 you find the descriptions of the available variables.

Variable Name	Parameter	Description
mode	Travel mode	1: walk, 2: bike, 3: car, 4: public transport, 5: other
dist	Distance	trip distance (in mil = 10km)
indiv_id	Individual ID	individual id
age	Age	1: 12-18 years, 2: 19-24, 3: 25-39, 5: 40-64, 6: 65+ in 2004
no_child	Number of children	number of children 0-12 years
nohhdl	Number of DL	number of driving licenses in householsd, 4 means "4 or more"
income	Income	monthly income in SEK, 1: 0-7500, 2: 7501-10000, 3: 10001-15000, 4: 15001-25000, 5: 25001-40000, 6: 40001-55000, 7: 55001-70000, 8: 70001+
cons_cap	Consumption capacity	consumption capacity according to the social norm, 1: low, 2: low to average, 3: average, 4: average to high, 5: high
occup	Employment	1: gainfully employed, 2: student, 3: other
cartime	Car time	car driving time (in hours)
carcost	Car cost	car cost based on per distance cost, e.g. fuel (16 SEK/mil)
PTtime	Public Transport time	total travel time by PT (PTaux + PTinvt + PTtotwt)
PTcost	Public Transport cost	cost of public transport trip
PTauxt	Public Transport time	public transport: walk time to station/bus stop
PTfwt	Public Transport time	public transport: first waiting time
PTinvt	Public Transport time	public transport: in vehicle-time (on board-time)
PTnoboard	Public Transport time	public transport: number of boardings
Pttotwt	Public Transport time	public transport: total wait time
Boolean Parameters		
male	Male	1 if male
inncity	Innercity HH location	1 if residence located in the innercity
single_househ	Single HH	1 if household with single adult
househ_w_kids	Children	1 if household with kids 0-12 years
own_drv_lic	Driving license	1 if respondent has driving licence
caracc	Car accessibility	1 if accessibility to car in household
self_empl	Self-employed	1 if self-employed
flext	Flextime working	1 if flexible working hours
parkposs	Work place parking	1 if parking possibilities at workplace
cheappark	Work place parking	1 if free or cheap parking at workplace
compcar	Company car	1 if access to company car
PTsubs	Public transit card	1 if employer subsidises public transport (travel card or other)
envcar	Environmental Car	1 if car is in the "environmental" category (no toll and 2 h free parking)
car_md	Car accessibility daily	1 if car accessible on measurement day
Ptcard	Public transit card	1 if possessing travel card for public transport on measurement day

Table 1: Variable descriptions for data set

Lab 1a. Implementing a Discrete Choice Model – 2013-09-10

Goal for Lab 1: Your task in this lab is to: 1) complete the matlab-code by specifying the logit-probabilites, t-statistics and log-likelihood value; 2) experiment with alternative MNL model specifications; and 3) use the model to predict the change in mode share due to some different policy changes.

1)

To get the program running, the following lines needs to be changed:

Logit.m

Line 15-22: Specify the choice probability for respectively mode according to the logit-model. The utility for each person and choice is given in the vectors V_walk, V_bike etc. The variable

sum_exp can be used to calculate the sum of the exponentials for all modes only once, and thus save computational time.

LogLikelihood.m:

Line 23: Change the output LL to the sum of the log-likelihood of the observed choices. P_choosen is a vector containing the likelihood of the observed choices for each individual.

You should now be able to run the code and get some result, with parameters that we have pre-specified for you. Look at the predicted shares for respectively mode, are they reasonable? Do you think you have the correct estimates? Have the program really converged to a solution?

ICM2013main.m

Line 25: Change the 'MaxFunEvals' so that the optimization algorithm converges.

After completing the above matlab-code, you should be able to estimate the model. On lines 32-end in ICM2013main.m, some results that might be interesting to analyze will now be calculated. You might want to add some more result, and especially need to specify the Value of Time of car and public transport.

ICM2013main.m

Line 76-77: Calculate the Value of Time (VoT) for car and public transport. If you change the names of the parameters for 'cost' and 'PT_time' and 'Car_time' in SpecifyVariables.m, you should change them accordingly on line 65-69.

2)

The model specification provided in the code is very simplistic and the next task is to "play" with the model specification.

The variable specification is done in the file SpecifyVariables.m. You should change variables by adding or subtracting parameters to/from relevant variables vectors. If you think that some of the parameters need to be generic and not mode specific you should give them the same name. Observe that 'cost' is a generic parameter in the original specification.

SpecifyVariables.m

Line 67-77: Specify the model by determining which variable that should influence which mode. This is the only lines that you need to change when specifying the model. If a variable has the same name for multiple alternatives, the associated parameter will be the same.

Line 43-60: These lines include some examples of possible variable. Some more examples of variables can be found in the bottom of the file.

3)

Once the model is working and you have found a good specification you should use the model to predict the change in mode share in response to possible policy. Two alternatives are considered: a) an increase in car cost by 10% (for example through an increase in fuel taxes); and b) a decrease in public transport cost by 10% (for example through subsidies). Use the mode shares from the original scenario and compare with the two new. Which is the best policy if the goal is to decrease the mode share of car?

Reporting

I. AIMS AND CONTENT

The aim of the lab is to allow you to work with discrete choice analysis in practice using a transport application. The results from the labs should be documented in reports, and these report forms the most important base for your final grade in the course. Therefore, the reports should be written individually (although you should collaborate on the results and coding) and the reports should contain certain results and analysis (described in this document) to fulfill the requirements of the course.

The main goal is to estimate the "best" mode choice model you can and use it to forecast the change in demand in response to: a) an increase in car cost by 10% (for example through an increase in fuel taxes), or b) a decrease in public transport cost by 10% (for example through subsidies). Use the mode shares of the original scenario and compare with the two other. Which is the best policy if the goal is to decrease the share of car?

You should motivate your choice of "best" model by comparing it to other models you have estimated. You should therefore include estimation results of at least four models, including two logit and two mixed logit specifications. Consequently you can and should compare and critically assess the results of two models of the same type and two models of different types.

II. OUTLINE OF THE REPORT

The report should be divided into the sections as described below and within each section the text should be well structured in a clear and concise language. Your analysis and conclusions should be well-founded and explained in a clear way (section III contains the detailed assessment criteria).

Abstract

A concise and factual presentation of the purpose and objective of the work, the principal results and major conclusions. An abstract is often presented separately from the report and should be able to stand-alone.

Introduction

State the purpose and objective of the work and some background on the data. A review of relevant methods should also be provided.

Model specifications and estimation results

Present the different model specifications and the corresponding estimation results. Analyze the models using informal and formal statistical tests (t-tests LL-ratio tests etc). Each model should be summarized

in a table with estimates, t-statistics etc. Also, compute the value-of-time for car and public transport and discuss if you find these values reasonable and why.

Choose your best model and motivate your choice based on an analysis of the results. You should include the results and statistics you find relevant for your analysis.

Prediction results

Present and analyze the forecasting results (the output "model prediction tables") for each of the scenarios and compare to the base case. Note that you should only to this for the one model you have chosen to be the "best".

Compare forecasted results with computed elasticity and discuss if and in this case why they are different and motivate which policy you believe is the best one.

Conclusion

The purpose, objective, method and the main conclusions are summarized in this section.

References

Provide the details of all the references used in the report.

Check list

Include the check list as an appendix to your report. The check list can be found on Bilda.

Note that the report should be **maximum 10 pages** long so choose your tables and figures carefully. The report can of course be shorter. *It is the quality of the report that will be assessed* against the criteria described in this document*, not the length*. However, if you exceed the 10 pages we will not read or assess page 11 and onwards.

Also note that there is not one right answer here, the content of your report is assessed based on the motivations and analysis you do to justify your statements.

III. ASSESSMENT OF THE REPORT

The report will be assessed and graded based on the table below (see separate page). Please read it carefully to take all of the assessment criteria into account. The intended learning outcomes (ILOs) that will be assessed are:

1. choose between different discrete choice models for a given problem and explain their advantages and disadvantages;

2. analyze and statistically compare different model types and model specifications for a given problem;
3. critically assess the results of a given model formulation;
4. explain maximum likelihood and simulated maximum likelihood estimation;
5. formulate and estimate logit based models using standard mathematical software such as MATLAB;
6. Choose one of your models as the best one and argue for its advantages;
7. use an estimated model for prediction/policy analysis. If you were to report your results to authorities, what would be the final model that you would choose and why?
8. Include a discuss for, and against, defining mode-specific coefficients for time and price and the effects of such mode-specific parameters on the value of time.

If you aim for a higher grade you should, besides doing the above points very well, also:

9. use a mixed logit model for prediction/policy analysis
10. introduce correlation among parameters in the mixed logit model.

Both 7 and 8 needs to be done for an A and either 7 or 8 for a B.

III. Deadline

Hand in your report no later than 8 Oct 2013, 23:59.