Maastricht University Leading in Learning!

COURSE MANUAL

Second-year Project II

EBS2003

Academic Year 2017/2018

Coordinator: Dr. André Berger, Dr. Alain Hecq

Contents

Chapter 1

Course Information

1.1 Learning Goals

In this course students will learn how to apply the methods and skills from their previous courses in Econometrics and Operations Research to a real-life case. The case will be provided and supported by representatives from DHL LLP and involves the analysis of shipment data of one of the biggest airplane manufacturers in the world.

1.2 Course Structure

1.2.1 Opening Lecture

During the opening lecture (Monday, June 18th 2018) a representative from DHL LLP will introduce you to the company and the case, which will be the basis of the assignments for the course. You have to work on the case in teams of two students. On the day of the opening lecture you will have to submit the names of the team mates, so that we can schedule all teams for the tutorial meetings.

1.2.2 Tutorials week 1

On the Thursday of the first week (June 21st 2018) there will be one tutorial meeting. It is essential that you carefully read the case description, get familiar with the data, and start working on the assignments before this tutorial meeting. During the tutorial meeting you can discuss with the other students questions about the case and the data and possible approaches. The tutor will be there to assist you and answer any questions.

1.2.3 Tutorials week 2

On the Tuesday of the second week (June 26th 2018) there will be tutorial meeting in a computer room in the library. During that tutorial you can work on your implementation and ask the tutor for help with any implementation issues.

1.2.4 Presentation

On the Friday of the second week (June 29th) each group will have to present the main findings of their work in a 15 minute presentation. This presentation will be graded and both team members have to present in order to pass the course.

1.2.5 Questions & the discussion board

If you have any questions about the case, your approach, or about the implementation of your ideas, please ask the tutors during either of the two tutorial sessions or via the discussion board. The discussion board will be set up on the student portal so that you can ask questions about the case, the data, or the assignments. Usually answers to these questions are useful for others as well and everyone should benefit from the same information. Therefore please refrain from sending such questions to the tutors or coordinators via email. On the student portal you can subscribe to the discussion board. Then you will receive email notifications whenever someone made a post on the discussion board.

1.3 Grading & Deliverables

In the case description you will find several assignments. You have to submit a report of at most 20 pages, containing a short introduction to the case as well as answers to Questions 1–6. For each question, describe your approach, including any formulas or algorithms that you use to answer the question. Moreover, provide numeric results in tables and/or charts. Each question is worth 10 points.

At the end of the report, you have to provide a two page business summary that gives a concise summary of your results. This should be an easy to understand text for the non-expert manager from which managerial conclusions can be drawn and should include some figures and/or charts representing the main results. This part of the report is also worth up to 10 points.

Finally, on the last day of the course period, your team has to give a 15 minute presentation on your approaches and results. The presentation will be evaluated on clarity, creativity of the ideas and on how convincing your conclusions are. You can obtain up to 10 points for the presentation. The total number of points achievable for the report and the presentation is 80. You pass this course if you achieve at least 44 points in total and if you score at least 5 points for the presentation. In that case, your final grade is equal to your total number of points divided by eight, rounded to the nearest half-integer. Every student is expected to participate actively in the team work. If it becomes apparent that a student contributed insufficiently to the team work, different grades can be assigned to the members of the team.

Please compress all your files in a single .zip folder (**no tar, rar or other compressing formats!**). Organize them in a clear way, i.e. use sub-folders for your data files and implementation files and only save the final report and the final presentation in the top-level folder.

Deadline:

All files have to be submitted via the student portal before **Thursday, June 28th, 2018, 23:59.**

Resit

If you fail to submit the report by the deadline, if you score less than 5 points for the presentation or less than 44 points in total, you will have to do a resit assignment as determined by the course coordinator.

No partial results or points can be transferred if you enroll in this course in another academic year.

1.4 Cheating and plagiarism

The work for this course has to be done in teams of two students. It is okay to ask another student/team about a specific implementation/compilation issue that you are stuck with or to exchange some general ideas. However, it is not allowed to share any processed data files, source code etc. You have to develop your own models, data structures, algorithms and write your own code. All work that you prepare for this course has to be your own. The school's policies on cheating and plagiarism apply. More details on these policies can be found on the student portal.

Chapter 2

Case Description

2.1 Introduction

DHL LLP is a group within the DHL Global Forwarding division belonging to the Deutsche Post DHL. LLP stands for Lead Logistics Partner and acts as a neutral partner. DHL LLP instigates and manages change across the customer's entire supply chain, to meet changing business and customer demands. LLP does this by bringing continuous improvement and cost reduction, introducing lean logistics processes and optimizing logistics networks and transportation flows. More will be explained about DHL LLP during the guest lecture.

For this case you will consider shipment data of one of one of the biggest airplane manufacturers in the world, for which DHL LLP manages the transportation. You will be provided with a data set containing data on origins, destinations, volumes, weights, etc. for these shipments.

It will be your task to use and analyze that data in order to build time series, test for seasonality, make forecasts, and analyze and improve the given clustering. You should achieve this using the techniques and the software that you know from your Econometrics and Operations Research courses.

2.2 The data set

The file `TRsFullDataUM.xlsx` contains data of 81520 shipments that were carried out during the time frame from January to October 2017. Below you find a table with the explanation of the different columns. You can save this file as a CSV-file in order to read the data into a Java program.

Column	Column Title	Content
A	TR Code	Unique code of transportation request
B	TR Creation Date/Time	Creation date of transportation request
C	TR Source Location Code	Unique code of source location
D	TR Dest Location Code	Unique code of destination location
E	TR Gross Weight (KG)	Gross weight (in kg)
F	Nb of Ship Units	Number of shipped units
G	Origin Country	Country of origin
H	TR Source Location Postal Code	Postal code of origin
I	Origin City	City of origin
J	OriginFull	Full address of origin
K	OriginCluster	Cluster ID of origin
L	OriginClusterLat	Latitude of origin cluster
M	OriginClusterLong	Longitude of origin cluster
N	OriginLat	Latitude of origin location
O	OriginLong	Longitude of origin location
P	Dest Country	Country of destination
Q	TR Dest Location Postal Code	Postal code of destination
R	Dest City	City of destination
S	DestinationFull	Full address of destination
T	DestinationCluster	Cluster ID of destination
U	DestinationClusterLat	Latitude of destination cluster
V	DestinationClusterLong	Longitude of destination cluster
W	DestLat	Latitude of destination location
X	DestLong	Longitude of destination location
Y	TR Pickup - Event Day	Weekday of actual pick up date (0=Sunday, 1=Monday, etc.)
Z	TR Gross Volume (M3)	Gross volume of transportation request
AA	PUDate	actual pick up date
AB	Distance	Distance between origin and destination (in km)
AC	ClusterDistance	Distance between respective clusters (in km)

2.3 Assignments

There are in total six questions that you have to work on, the first half focuses on Econometrics techniques, the second half on Operations Research techniques.

1. Data (pre)processing [10 points]

   An important task in applied econometrics is the collection of data. Sometimes it 's easy. You click and download data from OECD data base for instance. Sometime it's more messy due to the use of different sources, to breaks in the definition of series, to different frequencies. The data set proposed by DHL has already been preprocessed. However, a bit of manipulation should be done to obtain the time series we want you to work with.

   - You first task is to find the most frequent origin cluster. For that cluster, find the two most frequent destination clusters (column T). Let 's call the link between the origin and destination clusters a lane, such as you have to develop models for lane 1 and lane 2.
   - For each lane we are interested in 3 series (You consequently have 6 series to investigate):
     - Column E: Gross Weight (KG), W_t
     - Column F: Nb of Ship Units, U_t
     - Column Z: Gross Volume (M3), V_t.
   - Build time series for those lanes/series. You have daily data for six months, sometimes with more than one shipment per day. Aggregate them to have only one observation per day. Choose probably a 5 days a week frequency.

2. Daily seasonality [10 points]

   DHL has observed that shipments are dependent of a particular day in a week. There might consequently exist a seasonal pattern in the series that is worth exploiting.

   Then,

   - Determine/test whether you prefer to take the variables in levels, in log-levels or in growth rates. Maybe just by looking at series and not necessarily using formal (e.g. unit roots) tests. Indeed the span of data (six months) is likely too short to really trust those tests.
   - Carry out regressions for the six series to determine whether there is a significant daily effect. For instance for the gross weight for lane 1, the regression will be

   $$W_t^{L1} = \alpha_0 + \alpha_1 Tuesd + \alpha_2 Wed + \alpha_3 Thur + \alpha_4 Frid + \varepsilon_t$$

   - Test the null of no seasonality, i.e.

   $$H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0$$

   - Look at residuals and detect (test, plot ACF) whether you have some autocorrelation. Add lags of the dependent variable if needed or identify the ARMA model or add exogenous variables, etc.
   - Provide final specifications with robust standard errors if you have heteroskedasticity (test for that).

3. Forecasting [10 points]

   Once you are happy with your specifications, forecast the next week shipments, i.e. the week after the last calendar point available in the data set.

Information for questions 4–6

For questions 4–6 you should implement classes and algorithms in Java that can be used to answer the given questions. Carefully think about which and how many classes you are going to use before you start implementing. This part considers the assignment of origin and destination locations to the respective origin and destination cluster. In order to reduce the number of shipments for testing your algorithms, pick up dates will be neglected. Therefore you should, in a pre-processing step, aggregate the shipment data (weight, volume, number of shipments) for a given origin-destination pair over the given time period. This can be done either using Excel or using the Java data structures that you will develop for question 4–6.

4. Clustering I [10 points]

   Develop an appropriate data structure that will allow you to analyze the clustering of the shipments that is given in the data set. Your data structure should be flexible enough so

that it can be extended to change the origin or destination cluster assigned to a shipment or similar other changes. Read the remaining questions before finalizing your data structure.

Using your data structure, implement methods that can compute the following performance measures of the current clustering:

- the total distance of all shipments from the origin location to the origin cluster,
- the total weight that is being transported from the origin locations to the origin clusters,
- the total volume that is being transported from the origin locations to the origin clusters,
- the total number of shipments that is being transported from the origin locations to the origin clusters,
- the same four measures as above, but for the destination clusters.

Summarize the above measures for the current clustering by listing the values for these measures for the average cluster as well as the three clusters having the lowest (highest) such values.

5. Clustering II [10 points]

   Develop and implement an algorithm that can find a different assignment of the locations to clusters, so that the total distance of all shipments from the location to the assigned cluster (origin and destination) is small (not necessarily minimum). The algorithm should take as input the number k of clusters that can be used.

   Run your algorithm for different values of k and evaluate the measures from above for your obtained solutions.

6. Clustering III [10 points]

   Suggest alternative ways of clustering the shipments and implement solutions to find such clusterings. This could for example be based on the direction of the flow, destination country, the length of the shipments, etc. Using your approach you should be able to create distinct groups of locations that follow a similar shipping profile.

   Argue why you chose for your proposed alternative and discuss the outcomes.

7. Business Summary [10 points]

   Write a two page business summary that gives a concise summary of your results. This should be an easy to understand text for the non–expert manager from which managerial conclusions can be drawn and should include some figures and/or charts representing the main results.