Maastricht University

Case Description

Second-year Project II

EBS2003

Academic Year 2020/2021

Coordinator: Prof. dr. Tjark Vredeveld, Prof. dr. Alain Hecq

Contents

1 Introduction

In this course, you are provided with shipment data of a manufacturer of machinery for landscaping companies. This manufacturer is outsourcing its logistics process to a Logistics Service Provider (LSP). You will be provided with a data set containing data on origins, destinations, volumes, weights, etc. for these shipments.

To organize the shipments, the LSP has clustered the origins and destination locations in several clusters. Then, on a certain day, the LSP collects all shipments within one origin cluster and brings it to a cross dock (central location). In this cross dock, the shipments are sorted on destination cluster and then the shipments for each destination cluster are transported by trucks. Finally, once the shipments arrive at the cross dock of the destination cluster, these shipments are then delivered to their final destinations.

To be able to organize their transportation better, the LSP wants to know an estimate on the volume/weight that is transported between the clusters as well as on how to efficiently organize the collection to and distribution from the cross dock within one cluster. It will be your task to use and analyze that data in order to build time series, test for seasonality, make forecasts, and analyze the given clustering in terms of collection and distribution costs. You should achieve this based on the techniques and the software that you know from your Econometrics and Operations Research courses (some additional techniques will be needed).

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)

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]**

   As you might know now, an important task in applied econometrics is the collection of data. Sometimes it's easy. You just click and download data. Sometime it's more messy due to the use of different data sources, to breaks in the definition of series, to different data frequencies, typos, data conversions, etc. The data set has already been preprocessed although it is maybe not perfect. However, a bit of manipulation should be done to obtain the time series we want you to work with.

   - You must first take 1 of the 4 most frequent origin cluster. This is to avoid to listening 25 times the same story.
   - To know what origin cluster you have to investigate, add the last digit of the team members and substract a multiple of 4 (modulo operator) to get your origin cluster. Examples: (1+0)mod(4)=1, (3+4)mod(4)=3, (9+6)mod(4)=3, (2+6)mod(4)=4.
   - Then for that cluster, take your most frequent destination clusters (column T). You should consider a distance between the origin and the destination of more than 20 kilometers[1] and less then 400 kilometers. Let's call the link between the origin and destination clusters a lane, such as you have to develop models for one of the 4 lanes.
   - For your lane, we are interested in 3 series :
     - Column E: Gross Weight (KG), W_t
     - Column F: Nb of Ship Units, U_t
     - Column Z: Gross Volume (M3), V_t.
   - Build the three time series for your lanes. 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 a 5 days a week frequency, include a zero when there isn't any shipment on that day and create an empty 0 line when the day is off. Indeed, keep for convenience the time series balanced, namely a week has always 5 days. Make plots to detect obvious extreme outliers that you should/could control for in the next parts of the case.

2. **Daily seasonality [10 points]**

   The LSP 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,

[1] The most frequent lane consists in shiping goods to the other side of the street. Starting with lanes for at least 20 kilometers we try to avoid to consider those trivial lanes. Under 400km, the shipment is still considered as a short distance shipment.

- Determine/test whether you prefer to take the variables in levels, in log-levels. Do you think that the series are stationary? Maybe just by looking at graphs of 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 each of the 3 series to determine whether there is a significant daily effect. For instance for the gross weight variable for lane 1, the regression will be

$$W_t^{L1} = \alpha_0^{L1} + \alpha_1^{L1} Tuesd + \alpha_2^{L1} Wed + \alpha_3^{L1} Thur + \alpha_4^{L1} Frid + dummies + \varepsilon_t \quad (1)$$

  where dummies are 0/1 variables with 1 for the additional observations you have created to account for a bank holiday and/or extreme ourliers.
- Look at residuals and detect (test) whether you have some autocorrelation. Look at the correlogram and significant ACF and PACF, particular at order 1 and 5/6.
- Identify the SARMA(p,q)(P,Q) model and reestimate the whole model with dynamics and seasonality. Those identification schemes will be explained in a lecture.
- Once you are confident that you have correctly controlled for the presence of dynamic, test the null that each day is non significant, for instance for the first lane with $i = 1$ to 4

$$H_0 : \alpha_i^{L1} = 0 \quad (2)$$

  as well whether there is a difference in daily effects, namely for equation (1)

$$H_0 : \alpha_1^{L1} = \alpha_2^{L1} = \alpha_3^{L1} = \alpha_4^{L1} = 0 \quad (3)$$

- Provide final specifications with information about mispecification tests.

3. **Forecasting [10 points]**

When you are happy with your specifications, forecast, using both (i) static and (ii) dynamic forecasts two weeks of data.

- First reestimate your 3 equations without the last 10 observations and forecast the last 10 points. Compute RMSE (given in EViews).
- Next, reestimate the model for the whole sample and forecast the next 10 future observations. Can you compute static and dynamic forecasts and the RMSE as well?

Information for questions 4–6

For questions 4–6 you need to determine efficient methods for collecting the shipments for the clients location to the cross dock of the cluster (cluster center); the location of the cross dock is given by OriginClusterLat and OriginClusterLong.

All shipments to be collected should be picked up by a truck that can handle a total weight of at most 22 tons and a total volume of at most $82m^3$. The fixed cost of a truck is €450 and there is a variable cost of €1.45 per kilometer.

For Questions 5 and 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.

To test your algorithm(s), you need to select at least 10 (OriginCluster, PUDate)-combinations. Out of these 10 combinations, 5 need to have at least 30 shipments that need to be transported to the cross doc of the cluster and the other 5 need to have at least 20 shipments that need to be shipped.

4. Data cleansing and base case [10 points]

   This part can also be done in Excel, if you want.

   As mentioned, you need to test your methods on some pick up dates for some clusters. When having a good look at the data, you will see that several shipments are already at the location of the cluster center and other shipments are too large to be handled by a regular truck. These shipments do not need to be considered for the collecting part and you can filter these out. You may restrict yourself for this case to shipments to be collected from Cluster 2, as this cluster has the most days with more than 20 shipments on one day. However, if you want, you can also search some days in another cluster.

   - Find 5 days that have at least 20 shipments to be collected in Cluster 2 (or any other cluster of your liking) and 5 days that have at least 30 shipments to be collected in Cluster 2.
   - For each day that you selected in the previous part, you can save the (important) data of the relevant(!) shipments to a text file that you can read into a java program.
   - The base case for the transportation problem, is that each shipment is collected by its own truck. Determine for each of the 10 days the total distance that needs to be traveled.

     Hint: How can you use the longitude and latitude of the shipment location as well as the cluster center to find a good approximation for the distance?

5. Find a good routing [10 points]

   A solution to the problem of collecting problem, for a given day, consists of a number of trucks that need to be used, an assignment of the shipments to the trucks and for each truck a tour along its assigned shipments starting and ending at the cross dock.

   Propose an easy method that finds a good initial solution and implement it to get solutions for the 10 selected days. How much do they improve on the base case of the previous question.

6. Finding a better routing [10 points]

   Once you have obtained a feasible solution, you can improve this solution by making local changes. One widely used method for making these local changes is *local search*. Local search methods iteratively search through the set of feasible solutions by moving from one solution to a neighboring solution until some stopping criteria are met. A *neighboring solution* is obtained by making some local changes.

Figure 1: Example of a 2-opt exchange

For example, one could move a shipment from one truck to another truck, or swap the trucks of two shipments. In this case, you should also take care of where to insert the changed shipment into the route of the truck.

Another way of obtaining a neighboring solution is to change the order in which the shipments are collected for one truck, without changing the shipments assigned to the truck. Standard methods for such a change are the 2-opt neighborhood, see Figure 1.

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 a non–expert manager from which managerial conclusions can be drawn and should include some figures and/or charts representing the main results.