Third Assignment

Christopher Gandrud, Hertie School of Governance, Spring 2016

Emilia Sicari & Rafael Lopez V.

March 25, 2016

Contents

1 Research question and project description

In our collaborative research project, we seeks to provide an answer to the following question: **How the rise in inequality and economic growth influences the purchase of cars as an example of luxury good, compared to the usage of public transportation systems in Singapore, from 1995 to 2014**. Therefore, we collected data on economic growth, income inequality, usage of public transports and purchase of cars covering the time span of 29 years, from 1995 until 2014. As suggested by our research question, economic growth, income inequality are the explanatory variables, while purchase of cars and usage of public transports are the dependent variables. The reason why we chose cars as example of luxury goods showing social status, is that in Singapore purchasing of cars is particularly expensive, due to high taxation and a certificate of car entitlement, whose cost can be higher than 70.000 dollars. For more details about the research proposal and case justification see ResearchProposal.

2 Processing data

2.1 Data sources and data gathering

The data that we need for our empirical analysis are to be retrieved from different sources:

- IMF Cross Country Macroeconomic Statistics open data available on Quandl, a website providing high-quality financial and economic data in different formats to facilitate data analysis. From this source we downloaded data showing the trend in Singopore's GDP per capita imported on R in csv format.
- Knoema, a knowledge platform connecting data with analytical and presentation tools, in order to allow users to access, present and share data-driven content. From The World Top Incomes Database - providing access to data on the distribution of top incomes in more than twenty five countries across the globe - we downloaded data on the top 10% average income and bottom 90% average income in Singapore. The reason why we did not gather data from the database Clio Infra available on Quandl as we had defined in our ResearchProposal , is that it did not provide sufficient data for the time span we are considering.

2.2 Cleaning, processing and merging of datasets

After importing data on R and cleaned them, using the "year" as common demonitator for every dataframe, in order to merge them afterwards.

- Singapore's GDP per capita is measured in national currency and at current prices. We selected only the time span of interest and and removed the unnecessary columns, to avoid later merging conflicts.
- As for data on bottom 90% and top 10% average income, data were considered starting from 1995. Since there were no data availble after 2009, we forecasted the value of the average income of the two categories performing a linear regression. Results were available in a new dataframe, which was then bound with the orginal one, in order to have the entire time series. After that, we used these data to create a new dataframe showing the trend in inequality from 1995 until 2014; these data were obtained dividing the values of the top 10% average by the values of the bottom 90% average income for each year, to show how many times the top 10% average income is higher than the bottom 90% throughout the considered period. After cleaning the datasets, we proceeded with merging them. However, the various datasets were merged one by one, since more cleaning was sometimes needed in order to avoid merging problems. The final dataframe has 19 observations and 14 variables, observed throughout the years 1995-2014. "'

3 Descriptive and inferential statistics

3.1 Creating a summary table

The table below shows the basic decriptive statistics for our variable included in the complete datasets.

% Table created by stargazer v.5.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: Ven, Apr 15, 2016 - 16:34:24

Table 1: General data summary

Statistic	N	Mean	St. Dev.	Min	Max
date	20	2,004.5	5.9	1,995	2,014
gdp.per.capita	20	50,294.8	12,748.0	35,345.5	71,317.9
inequality	20	6.1	1.3	3.9	7.8
cars	20	466,148.3	97,290.5	342,245	607,292
rentalcars	20	10,097.0	3,881.3	5,144	18,847
taxis	20	21,958.7	4,201.0	16,517	28,736
buses	20	13,993.9	2,302.3	10,723	17,554
motorbikes	20	138,985.9	6,435.2	129,587	148,160
other	20	146,548.7	10,180.8	134,756	161,698
bus.u	20	3,159.9	254.5	2,779	3,751
mrt.u	20	1,504.2	635.1	740	2,762
top	20	174,881.0	38,176.0	113,402.5	235,450.0
bottom	20	29,022.2	2,851.4	22,602.4	34,043.3
lrt.u	20	62.8	45.0	0	137

% Table created by stargazer v.5.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: Ven, Apr 15, 2016 - 16:34:24

Table 2: General data summary

labels1

3.2 Trends in gdp per capita, bottom 90% and top 10% average income

The following graph shows the trend of the three explanatory variables, throughout the period of time taken into consideration in Singapore. As we can see,

Figure 1 − Gdp per capita, top 10% and bottom 90% average income
in Singapore measured in national currency at current prices (1995−2014)

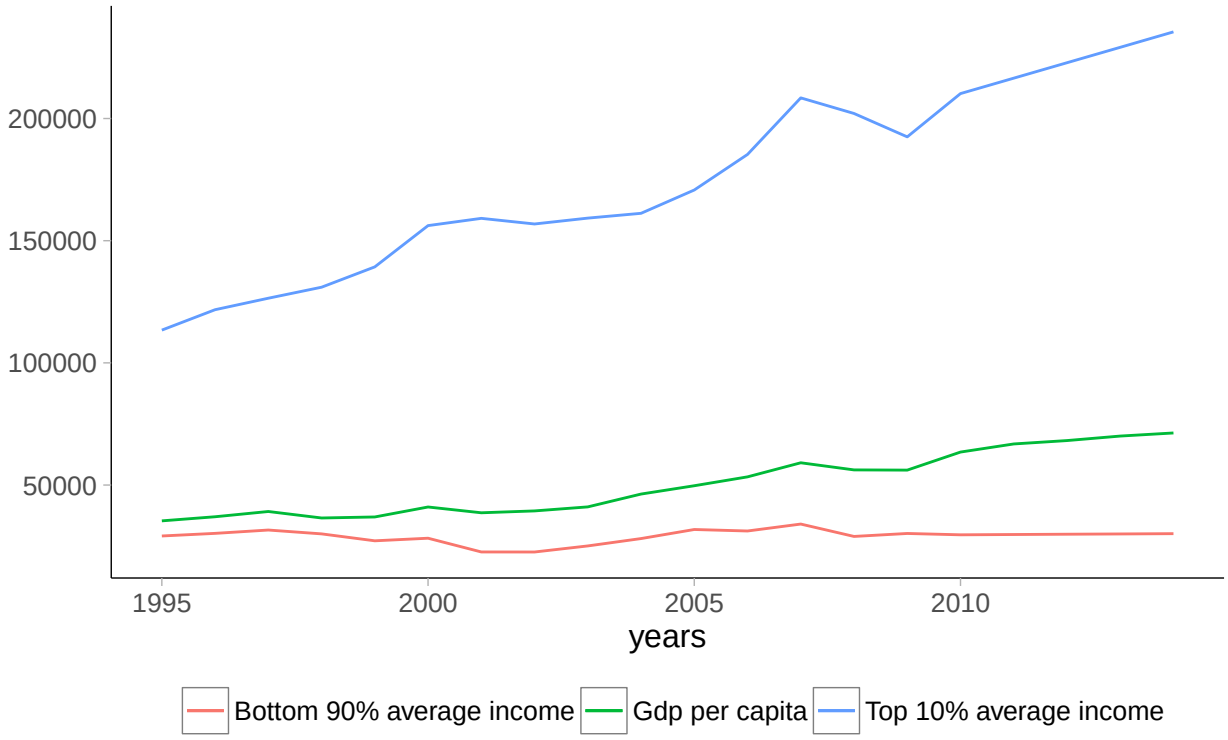


Figure 2 − Inequality in Singapore measured by the difference between
the top 10% and bottom 90% average income in number of times (1995−2014)

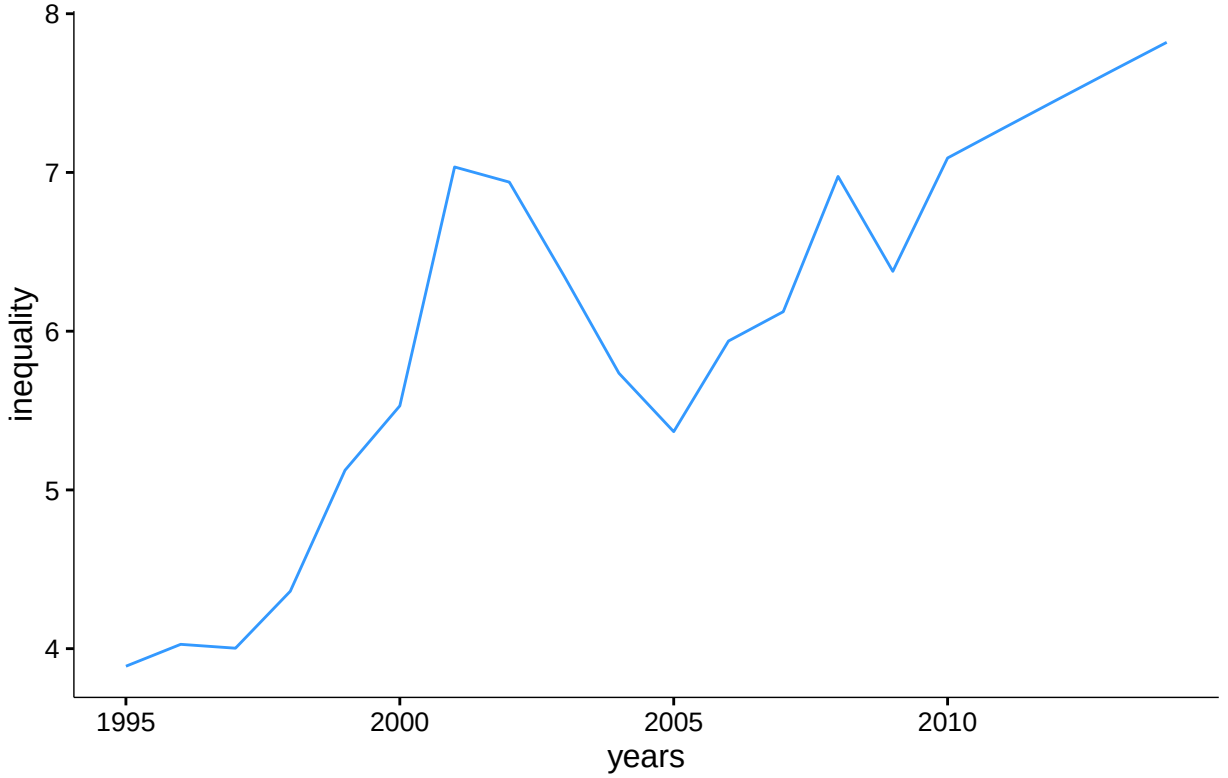


Figure 3 − Number of public and private vehicles in Singapore (1995−2014)

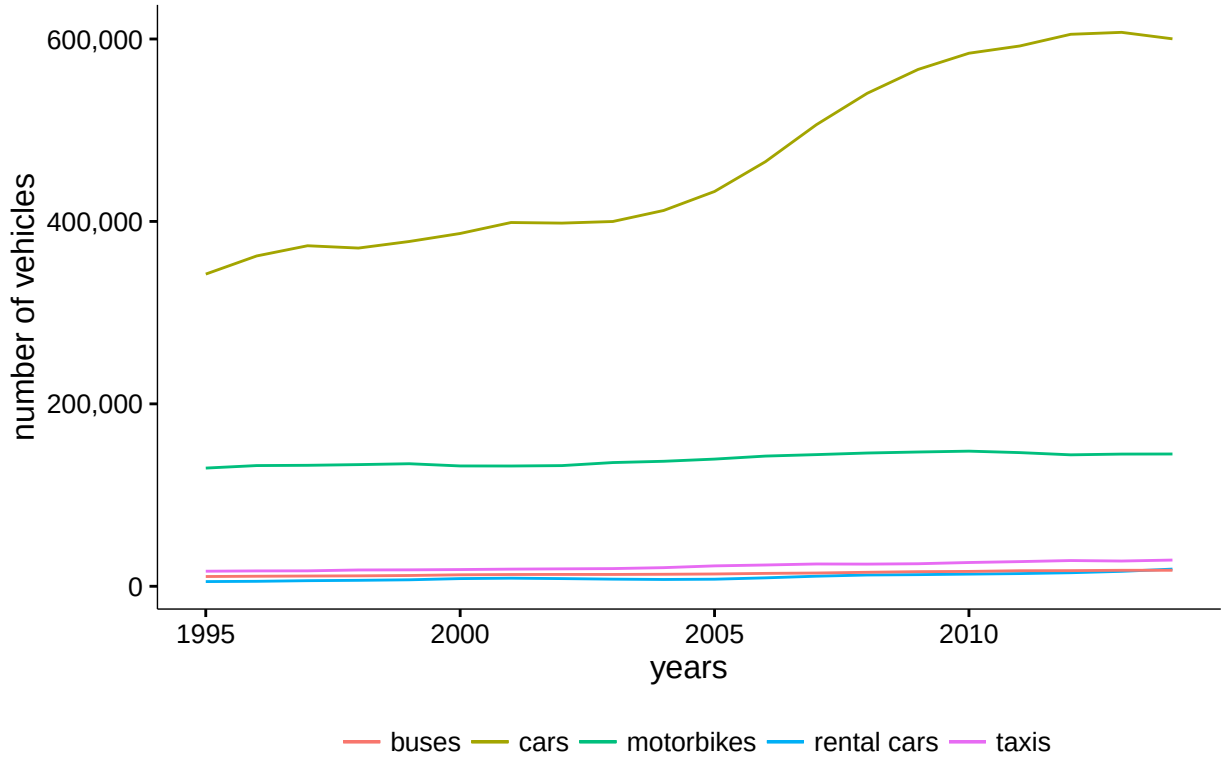
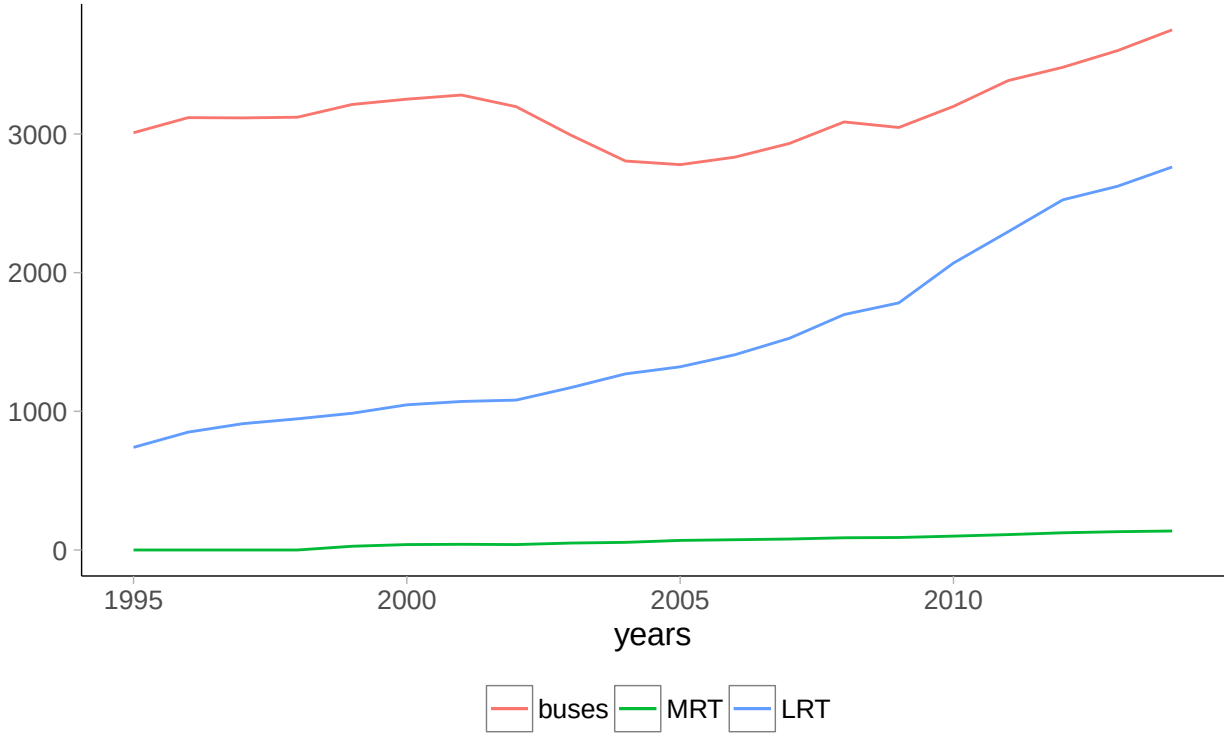


Figure 4 − Average daily passengers on public transports in Singapore in thousands (1995−2014)



3.3 Multiple regression analysis

```
##
## \begin{table}[!htbp] \centering
##   \caption{Multiple regeression models}
##   \label{}
## \begin{tabular}{@{\extracolsep{5pt}}lcccc}
## \\[-1.8ex]\hline
## \hline \\[-1.8ex]
##  & \multicolumn{4}{c}{\textit{Dependent variable:}} \\
## \cline{2-5}
## \\[-1.8ex] & cars & bus.u & mrt.u & buses \\
## \\[-1.8ex] & (1) & (2) & (3) & (4)\\
## \hline \\[-1.8ex]
##  gdp.per.capita & 5.62$^{*}$ & 0.004 & 0.04$^{***}$ & 0.05 \\
##   & (2.68) & (0.01) & (0.004) & (0.04) \\
##   & & & & \\
##  inequality & 16,917.56 & 82.29 & 82.35$^{*}$ & 436.15$^{**}$ \\
##   & (12,232.97) & (62.70) & (40.97) & (171.16) \\
##   & & & & \\
##  bus.u & $-$39.70 &  &  & $-$0.72 \\
##   & (49.70) &  &  & (0.70) \\
##   & & & & \\
##  mrt.u & 71.41 &  &  & 1.68 \\
##   & (71.02) &  &  & (0.99) \\
##   & & & & \\
##  lrt.u & $-$745.93 &  &  & 5.87 \\
##   & (869.85) &  &  & (12.17) \\
##   & & & & \\
##  Constant & 145,753.90 & 2,477.90$^{***}$ & $-$1,110.31$^{***}$ & 8,336.84$^{***}$ \\
##   & (162,408.10) & (249.55) & (163.05) & (2,272.35) \\
##   & & & & \\
## \hline \\[-1.8ex]
## Observations & 20 & 20 & 20 & 20 \\
## R$^{2}$ & 0.96 & 0.32 & 0.95 & 0.99 \\
## Adjusted R$^{2}$ & 0.95 & 0.24 & 0.95 & 0.98 \\
## Residual Std. Error & 22,249.39 (df = 14) & 222.36 (df = 17) & 145.28 (df = 17) & 311.31 (df = 14) \\
## F Statistic & 69.86$^{***}$ (df = 5; 14) & 3.94$^{**}$ (df = 2; 17) & 173.04$^{***}$ (df = 2; 17) & 1
## \hline
## \hline \\[-1.8ex]
## \textit{Note:}  & \multicolumn{4}{r}{$^{*}$p$<$0.1; $^{**}$p$<$0.05; $^{***}$p$<$0.01} \\
## \end{tabular}
## \end{table}
```

4 References