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 goods, 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 19 years, from 1995 until 2014. As suggested by our research question, economic growth and 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 the need to own 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 for our empirical analysis were retrieved from:

- IMF Cross Country Macroeconomic Statistics open data available on Quandl. From this source we downloaded data showing the trend in Singopore's GDP per capita meassured in singaporean dollars from 1995 to 2014. The data was provided in csv format and imported on R using the URL of the website.
- World Top Incomes Database available on Knoema, providing access to data on the distribution of top incomes in more than twenty five countries across the globe. From this source we downloaded data on the top 10% average income and bottom 90% average income in Singapore from 1947 until 2009, meassured in singaporean dollars. Since it was not possible to directly import the database to R, we requested and received the data via e-mail in csv format. Such data are available in the repository.
  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.
- Singapore's open data portal offered two databases:
  - The Annual Motor Vehicle Population, providing provides the number of public and private vehicles from 1995 to 2014, including: mortorbikes, rental cars, buses, taxis and other type of vehicles. While mortorbikes, rental cars and cars are private means of transportation, buses and taxis are to be considered public since in Singapore even the taxis are provided by the state.
  - Public transport utilization. This data is expressed as the daily average of thounsand commuters using public trasport by year. It covers the span from 1995 to 2014 and includes the following modes of transportation: MRT (underground), LRT (a localised rail systems acting as feeder services to the Mass Rapid Transit network), taxis (publicly run) and buses.

The following table summarizes the variables taken into consideration for the analysis.

Variable	Description	Time.frame
gdp per capita	measured in singaporean dollars at current prices. From 1995 to 2014	1980-2014
inequality	top 10% and bottom 90%	1947-2009
anual motor vehicle	singaporean's average income measured in singaporean dollars number of: cars, rental cars, buses, taxis, buses, motorbikes	1960-2015
public transport utilization	average commuters using daily: MRT, LRT, Buses, Taxis	1995-2014

2.2 Cleaning, processing and merging of datasets

- After importing data we used the "date" variable (year) as a unique identifier for all four datasets, in order to merge them afterwards.
- Since time frames of the data were different, we selected a common span of time: 1995-2014. In the case of bottom 90% and top 10% average income, we had to make a linear regression to forecast missing values (from 2009 until 2014). The results, available in a new dataframe, were later on bounded with the orginal one, in order to have the entire time series. As for LRT, values from 1995 until 1998 were missing since the service started to be provided from 1999; therefore, we completed the dataframe giving the value "0" for the first 4 years of the time span taken into consideration.
- Cleaning the data was limited to changing column names, eliminating the unnecesary ones and organizing the various data frames so to merge them more easily afterwards, using the year as common denominator. Only in the case of the dataframe cointaining the number of private cars in Singapore from 1995 until 2014 (car.pop.1) we had to change the format of the data from characters to integers, due to an incorrect import.
- In order to have an indicator showing the trend in inequality in Singapore between 1995 and 2014, we created a new variable - named "inequality" - by divididing the top 10% avereage income by the bottom 90% average income for each year: the coefficient of the division shows how many times Singaporeans earning the top 10% average income are reacher than the bottom 90% earners of the population.
- As for the number of cars, we simply divided them into the categories provided in the data original set: cars, buses, etc. Originally, they were in one column so we separate them in several ones to have the year as a unique identifier.
- Finally, we merged all the single dataframes into the new one, containing all the variables that we used to perfom descriptive and inferential statistical analyses.
  #Descriptive and inferential statistics

2.3 Descriptive statistics and central tendency

The table below shows the basic decriptive statistics for our variables.

% Table created by stargazer v.5.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: Lun, Apr 18, 2016 - 10:12:32

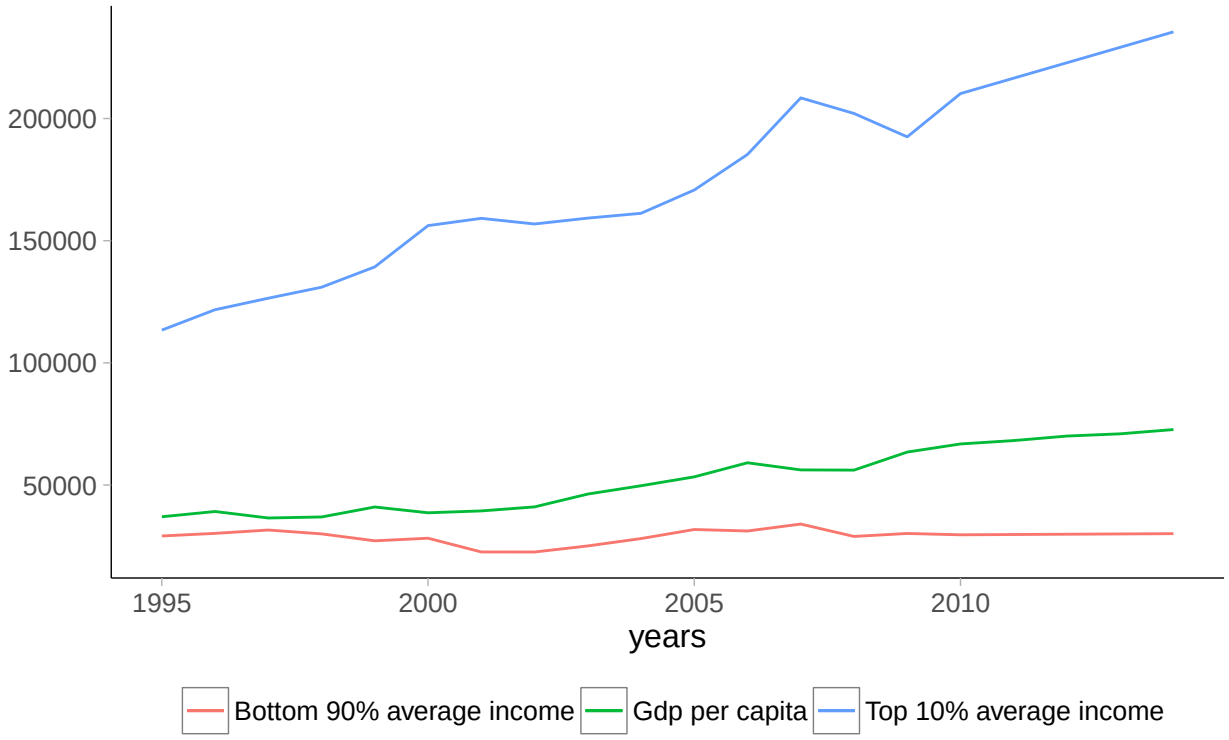
Table 2: General data summary

Statistic	N	Mean	St. Dev.	Min	Max
gdp.per.capita	20	52,145.5	13,146.2	36,524.7	72,711.4
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

2.4 Trends in gdp per capita and inequality

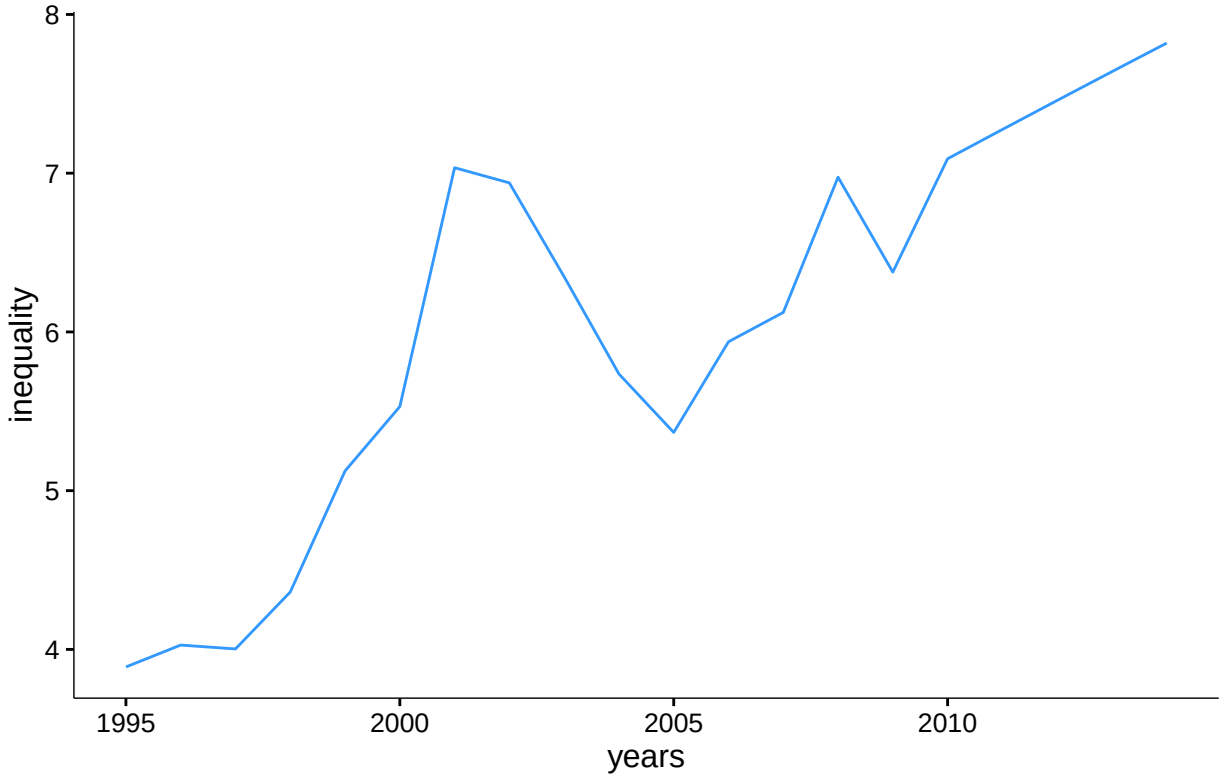
The following graph shows the trend of the three explanatory variables in Singapore between 1995 and 2014. As we can see, although slowly, the gdp per capita has risen thourghout the whole period, despite a slight decline between 2002 and 2005 and a more serious reduction in the years of the financial crisis, between 2008 and 2010. The top 10% average income shows the same trend: a steady increase throughout the whole period (in 2014 its value was more than 100% higher than the initial one), with a slight decline between 2002 and 2005, and a more serios reduction in the years of the financial crisis. However, the value of the bottom 90% average income has barely changed, enlarging the difference between the richest and the poorest and showing that the increase in the gdp per capita has benefitted only the top earners at the expensed of the poorest.

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



The growing difference between the top and bottom earners is clarified by the following graph, showing trend in inequality in Singapore, measured in number of times by which the top 10% earners are richer than those earning the bottom 90% average income. The graph confirms what already highlighted above: the difference between the rich and the poor has been increasing all the time, and the trend only reversed between 2002 and 2005 and beteen 2008 and 2010.

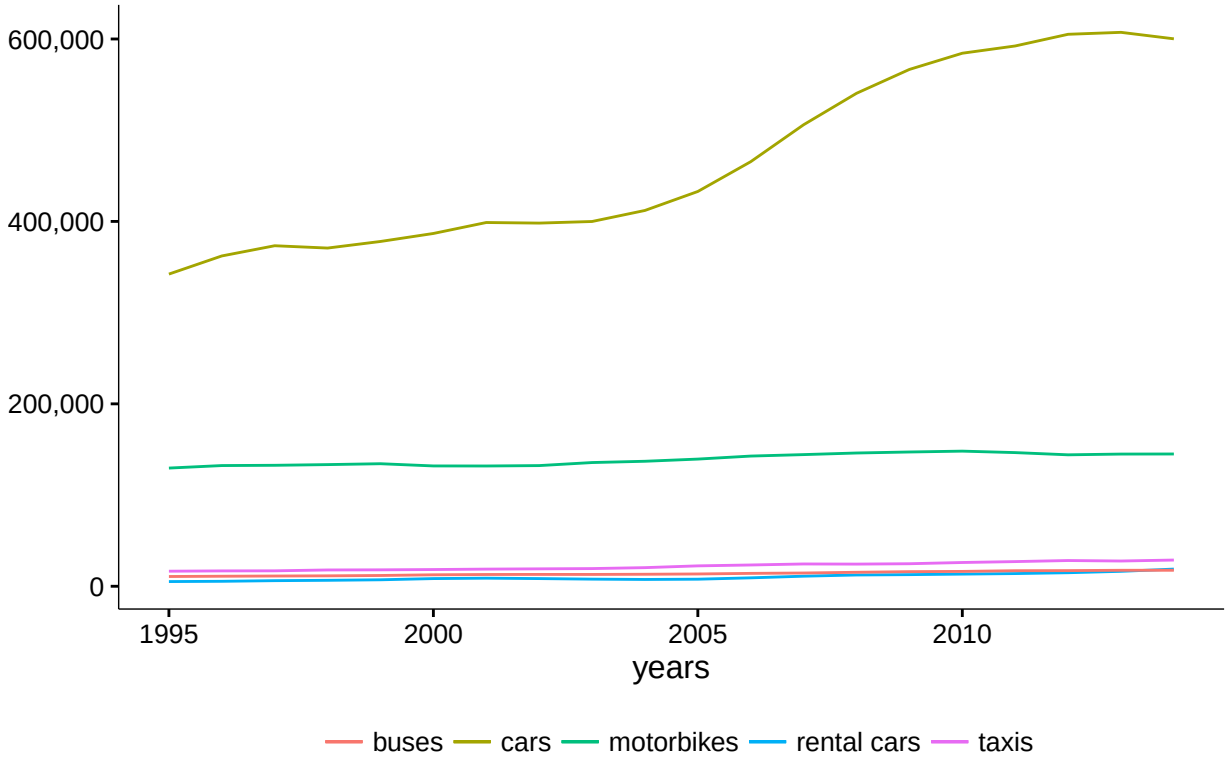
Figure 2 − Inequality in Singapore measured by the difference between
the top 10% and bottom 90% average income in number of times (1995−2014)



2.5 Trends in ownership and usage of vehicles

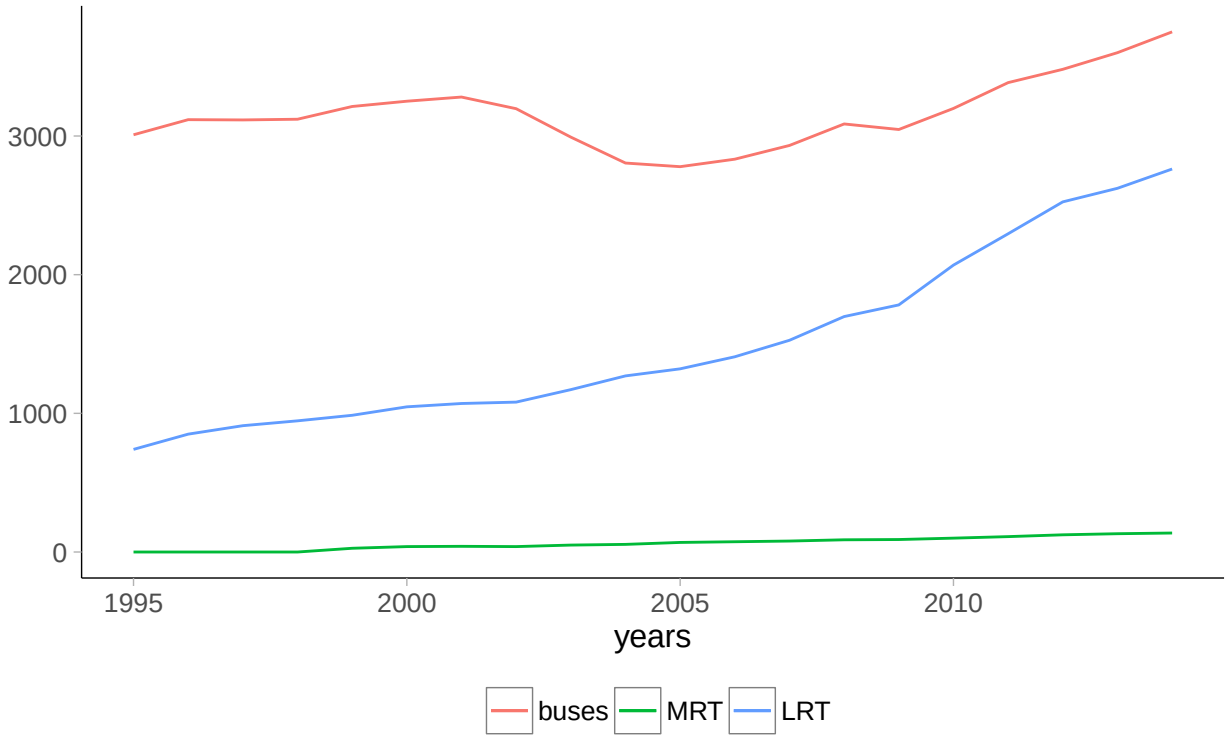
The following graph shows the number of public and private vehicles present in Singapore between 1995 and 2015. What is striking, is the continuous and rapid increase in tge number of privately owned cars (especially since 2006): in fact, in Singapore owing a car is particularly expensive, since not only cars are subjected to high taxation, but also the government requires car owners to purchase certificate of car entitlement, whose cost can be higher than 70.000 dollars. Therefore our first two hypotheses are verified: the higher the economic growth - measued by gdp per capita - and the higher the inequality, the higher the purchase of cars (for more details about the hypotheses see https://github.com/EmiliaSicari/ResearchProposal). A further assumption to be investigated is that such increase might be linked to the likewise rise in the top 10% average income: as the rich become richer, the purchase of luxury goods, such as cars, increases as well.

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



At the same time, even the number of passengers in the main public transportation (MRT and buses) has increased consistenlty over time. Despite that, the publicly owned buses have not chaged in number, as shown in the previous graph. Consequently, the third hypotheses - the lower the usage of public transports,the the purchase of luxury cars - is falsifies: in fact, usage of public transports has either increased (in the case of MRT and buses) or stayed the same (in the case of LRT), while the number of private cars has grown consistently. Even in this case a further assumption to be investigated is that those usign public transports are the lower earners.

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



3 Coorelation analysis

A first step to check out our new assumptions is to see the correlation between the variables included in the model, shown by the plot below: blue indicates negative correlation between variables, while red highlights positive correlations; color intensity and the size of the circles are proportional to the correlation coefficients, and the correlation coefficients corresponding to the color intensity are shown in the legend below the diagram. What clearly emerges from the plot, is that the variables are in almost all of the cases highly and positively correlated to each other. Bottom 90% average income and buse utilization are less correlated to the other variables, and bottom 90% average income is also negatively correlated with both inequality and buses utilization (which weakens the assumption that the poorest are those who use more public transportations). However, high correlation among explanatory variables might create problems in multiple regression analysis (multicollinearity). Therefore, it is necessary to explore collinearity before proceeding to the regression analysis.

3.1 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 & 4.61 & 0.0004 & 0.04$^{***}$ & 0.06 \\
##   & (3.19) & (0.01) & (0.005) & (0.04) \\
##   & & & & \\
##  inequality & 12,907.00 & 107.68 & 95.23$^{*}$ & 420.84$^{**}$ \\
##   & (12,733.87) & (63.37) & (50.22) & (164.85) \\
##   & & & & \\
##  bus.u & $-$18.93 &  &  & $-$0.29 \\
##   & (65.64) &  &  & (0.85) \\
##   & & & & \\
##  mrt.u & 80.79 &  &  & 1.40 \\
##   & (84.09) &  &  & (1.09) \\
##   & & & & \\
##  lrt.u & $-$597.73 &  &  & 5.57 \\
##   & (929.23) &  &  & (12.03) \\
##   & & & & \\
##  Constant & 123,631.40 & 2,489.72$^{***}$ & $-$1,114.29$^{***}$ & 6,891.02$^{**}$ \\
##   & (234,806.60) & (252.19) & (199.86) & (3,039.80) \\
##   & & & & \\
## \hline \\[-1.8ex]
## Observations & 20 & 20 & 20 & 20 \\
## R$^{2}$ & 0.96 & 0.30 & 0.93 & 0.99 \\
## Adjusted R$^{2}$ & 0.94 & 0.22 & 0.92 & 0.98 \\
## Residual Std. Error & 23,801.75 (df = 14) & 224.57 (df = 17) & 177.98 (df = 17) & 308.14 (df = 14) \\
## F Statistic & 60.69$^{***}$ (df = 5; 14) & 3.70$^{**}$ (df = 2; 17) & 112.47$^{***}$ (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}
```

3.2 Further steps

- investigating collinearity

References