Third Assignment

Christopher Gandrud, Hertie School of Governance, Spring 2016

Emilia Sicari & Rafael Lopez V.

March 25, 2016

Contents

#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, 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.

1 Processing data

1.1 Data sources and data gathering

The data that we need 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.
- Knoema. 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, meassured in singaporean dollars. 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. Since it was not possible to directly download the data base to R we requested the data, which have been sent via e-mail in csv format. This data is part of the repository.
- Singapore's open data portal. This portal offered two data bases:
  - Annual Motor Vehicle Population. The data provides the number of cars from 1995 to 2014, including: mortorbikes, rental cars, buses, taxis and other type of vehicles. The last two are considered public cars 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 (similar to the Berlin S-Bahn), Taxis (publicly run), Buses.

The following table summarizes the variables downloaded.

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%` singaporean's average income measured in singaporean dollars	1947-2009

Variable	Description	Time.frame
`anual motor vehicle`	number of: cars, rental cars, buses, taxis, buses, motorbikes	1960-2015
public transport utilization	average commuters using daily: MRT, LRT, Buses, Taxis	1995-2014

1.2 Cleaning, processing and merging of datasets

- After importing data we used the "date" variable (year) as a unique identifier for all four data sets, in order to merge them afterwards.
- Since time frames of the data were different, we selected a common span of time: 1995-2014. In one case we had to make a linear regression to forecast mising values to complete the time common frame, and in other we completed the data with we abstent for some years with "0", since we new that the problem was not lack of data but inexistence of that variable in those years.
- Cleaning the data was limited to changing column names, eliminating the unnecesarry ones, organizing the data in multile data frames so to merge them more easily afterwards. In one case we changed the format of the data from characters to integers, due to incorrect import.
- Data of inequality, namely bottom 90% and top 10% average income, were available only until 2009. Therefore, 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 the once complete data on inequality to create a simplified inequality indicator. This consis on divididing the top 10% avereage income of singaporeans by the bottom 90% average income, all for each year. Thus, the coefficient of the division shows how many times the top 10% average income of is higher than the bottom 90% throughout the considered period.

- POPULATION OF VEHICLES
- USAGE OF PT

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.

2 Descriptive and inferential statistics

2.1 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: Fri, Apr 15, 2016 - 16:54:08

% Table created by stargazer v.5.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: Fri, Apr 15, 2016 - 16:54:08

Table 2: 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 3: General data summary

labels1

2.2 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)

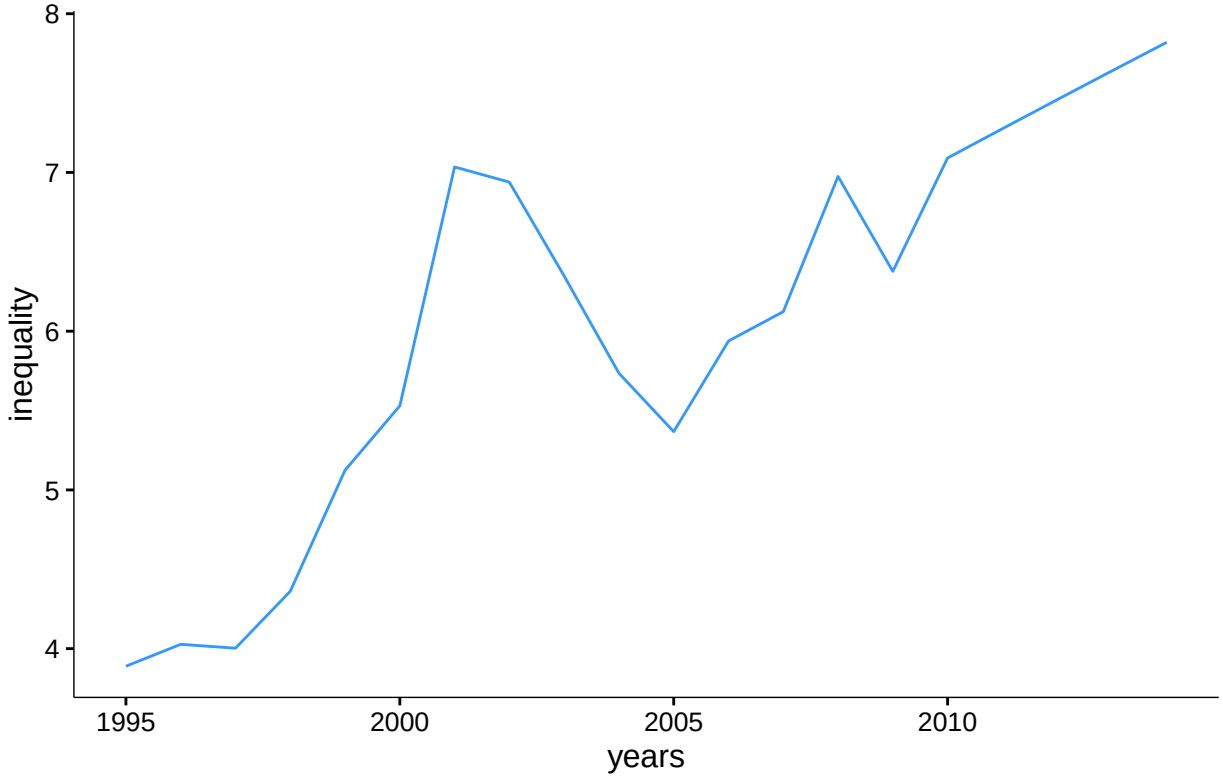


Figure 3 − Number of public and private vehicles in Singapore (1995−2014)

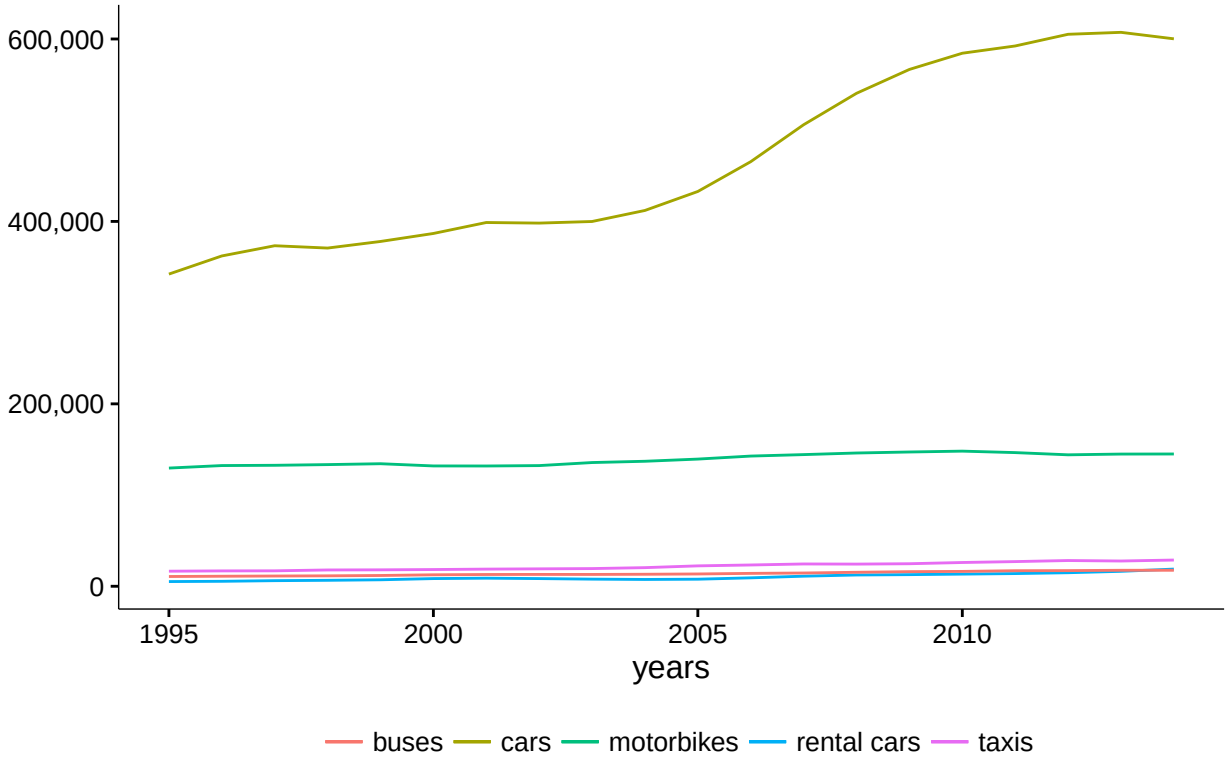
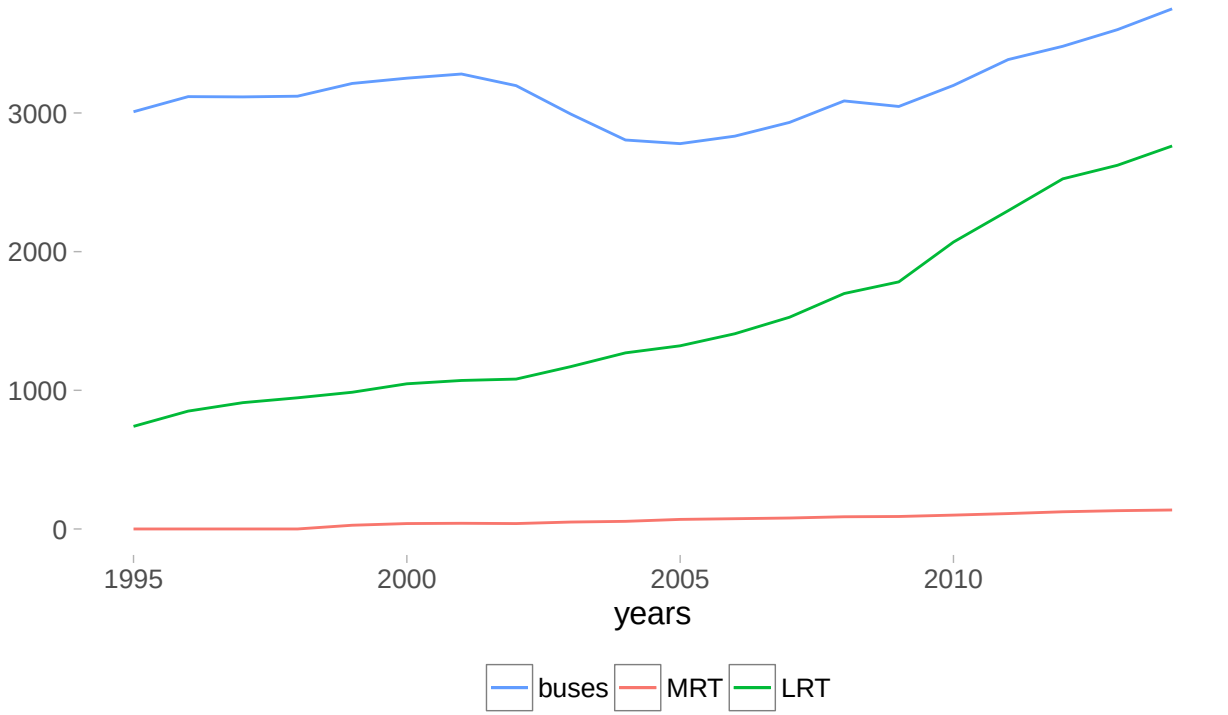


Figure 4 − Average daily passengers on public transports in Singapore in thousands (1995−2014)



2.3 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}
```

3 References

#Annex ## Annex 1: Final data set containing all the variables and observations

date	gdp.per.capita	inequality	cars	rentalcars	taxis	buses	motorbikes	other	bus.u	mrt.u	top	bottom	lrt.u
1995	35345	3.889	342245	5144	16517	10723	129587	137913	3009	740	113402	29157	0
1996	37031	4.027	362142	5451	16857	10998	132344	140512	3118	850	121752	30232	0
1997	39179	4.003	373345	6152	16933	11240	132629	142905	3116	911	126461	31592	0
1998	36525	4.362	370804	6536	17886	11429	133375	141051	3121	946	130964	30026	0
1999	36944	5.124	378024	7112	18029	11827	134346	139473	3213	986	139278	27180	27
2000	41018	5.53	386780	8438	18327	12569	131937	134756	3251	1047	156178	28243	39
2001	38660	7.035	398787	8857	18798	12902	131869	137157	3281	1071	159140	22623	41
2002	39423	6.939	398166	8423	19106	12992	132318	135951	3197	1081	156838	22602	39
2003	41070	6.348	399923	7803	19384	12951	135649	135333	2992	1171	159243	25087	50
2004	46320	5.735	412015	7455	20407	13173	137029	137316	2805	1270	161198	28106	55
2005	49715	5.367	432827	7756	22383	13494	139434	139098	2779	1321	170744	31813	69

date	gdp.per.capita	inequality	cars	rentalcars	taxis	buses	motorbikes	other	bus.u	mrt.u	top	bottom	lrt.u
2006	53355	5.938	465482	9235	23334	14120	142736	144466	2833	1408	185296	31204	74
2007	59114	6.123	505987	11054	24446	14530	144340	150979	2932	1527	208444	34043	79
2008	56201	6.975	540455	12391	24300	15327	146120	156089	3087	1698	202059	28970	88
2009	56111	6.377	566608	12763	24702	16023	147215	158207	3047	1782	192491	30185	90
2010	63498	7.091	584399	13347	26073	16309	148160	157541	3199	2069	210202	29643	100
2011	66816	7.275	592361	13919	27051	17046	146559	159768	3385	2295	216514	29760	111
2012	68205	7.458	605149	14862	28210	17162	144110	160417	3481	2525	222826	29876	124
2013	70047	7.64	607292	16396	27695	17509	144934	160344	3601	2623	229138	29993	132
2014	71318	7.82	600176	18847	28736	17554	145026	161698	3751	2762	235450	30110	137