Do Certain Unique Factors Impact a Country's Education Rank?

Alex Skipper

Econ5970

April 15,2018

Abstract

The purpose of this paper was to examine certain variables that could possibly impact a countries education rank. A simple linear regression was performed on 8 variables with 30 observations (countries). The results showed one significant dependent variable; Education Index. Reduced models were created with significant dependent variables only and performed slightly better than the original models. The original models were used to predict the models accuracy. Additional test were performed on the original models to test for multicollinearity and heteroskedasticity. The Education Index proved to show an impact but more in-depth research needs to be conducted.

1   Introduction

In many conversations today the educational system of various countries is brought up as a point of pride, strength, and success. This topic is of global interest, but in particular education ranking has been important for the United States in its effort to maintain its global education status. This project will be different from the other literature because it explores ideas that tend to be ignored. A country's education system is a useful determinant at showing the potential future progress a country will have and intellectual competitiveness. To measure a student's success or a country's education success, tests are performed to measure their abilities. Government[s], school authorities, and parents are constantly trying to improve student's performance in order to better the student's life and increase economic growth. This paper will show research of the effect that economic factors have on education as well as a few classroom climate factors. The paper will be a cross section analysis on the years 2000 and 2010. First, there needs to be way of establish ranks for the countries. The average of the reading, math and science scores on the PISA test were used form a ranking system. Eight different factors that might affect the rankings from two different years, 2000 and 2010, were included. The data will examine the top 30 countries that

perform the PISA test.

2   Literature Review

Many studies have been done on specific factors that schools can control to a degree such as, student to teacher ratio and school expenditures. Student to teacher ratio appears to be an important factor contributing to success or failure of a student's performance. A recent paper helps look into this effect, Effect of Education Policies and Institutions on Student Performance by Jieun Hong. In his paper he is looking to see whether school autonomy is being by the quality of education. Hong, suggests that "the pupil-teacher ratio is significant with negative signs, which indicate that students from a small class size can be educated effectively and efficiently" (2015). Showing that class size can have either a positive or negative impact on a child's education. Another study, The Tennessee Study of Class Size in the Early School Grades, looks at class sizes in Tennessee over a four year period, where students were comprised of "a small class of 13 and two classes of 22" Mosteller (1995). The results indicated "students who were originally in smaller classes continued to perform better than the students from regular-

sized classes with or without a teacher's aide" Mosteller (1995).

School expenditure to GDP is another studied variable because questions how been raised whether or not raising education spending increases student performance. The United States spends a smaller portion of its GDP on education but spends more per student than most countries. Many studies have looked into this controversial topic and the overall results show that increasing spending on education does not have a significant impact because it "depends on how the money is spent, not on how much money is spent"Lips, Watkins, and Fleming (2008). In the research paper, Does Spending More on Education Improve Academic Achievement?, the authors, Dan Lips, Shanea J. Watkins, Ph.D., and John Fleming, look into the per-pupil expenditures on public education by each state through grades K-12, as well as, 4th, 8th, and 12th grade NAEP math and reading scores. Their results indicated that increasing education spending does not led to significant refined student performance.

Prison population has been a recent variable to study because of the differences in criminal systems among the International community. A recent literature article, Mass incarceration and children's outcomes: Criminal justice policy is education policy by Leila Morsy and Richard Rothstein, shows

how "children with incarcerated parents are 33 percent more likely to have speech or language problems—like stuttering or stammering—than otherwise similar children whose fathers have not been incarcerated" Morsy and Rothstein (2016). Their results help shed light to a serious issue that the current criminal system can have a negative impact on a child's school performance. Many other variables have shown that there could be links between them and a student's educational performance but more research is required to provide conclusive results.

3   Data

The dependent variable will be an average of the reading, math and science scores on the PISA test. Observations and test scores for Year 2000 will be from the PISA 2000:Overview of the Study Design, Method and Results written by Stanat · Artelt · Baumert · Klieme · Neubrand · Prenzel · Schiefele · Schneider · Schumer · Tillmann · Weiß, and Organization for Economic Co-operation and Development (OECD). Observations and test scores for Year 2010 will be from the 2009 PISA test scores. This paper will include 8 independent variables and each will include observations from 2000

and 2010. Included: Enrollment Rate of both sexes as a percentage from the Data World Bank, the Education Index which is a number between 0 and 1 which accounts for the mean years of schooling in a country as percentage of the expected years of schooling for an individual from Human Development Report, School Expenditure as percentage of GDP from the Data World Bank, and the Unemployment Rate as a percentage from the Data World Bank. Also included: Health Expenditure per capita in USD from the Data World Bank, the Ratio of Pupils to Teacher from United Nations Educational, Scientific and Cultural Organization, the Prison Population Rate in per hundred thousand of the national population from World Prison Brief, and Gross Domestic Saving as a percentage of GDP from the Data World Bank.

The variable names are labeled differently for computational analysis. For Year 2000, x is the average PISA scores, x1 is Enrollment Rate, x2 is Education Index, x3 is School Expenditure as a percentage of GDP, x4 is Unemployment Rate, x5 is Health Expenditure per Capita, x6 is Pupil to Teacher Ratio, x7 Prison Population Rate, x8 is Gross Domestic Savings. For Year 2010, x0 is the average PISA scores, x11 is Enrollment Rate, x21 is Education Index, x31 is School Expenditure as a percentage of GDP, x41 is

Unemployment Rate, x51 is Health Expenditure per Capita, x61 is Pupil to Teacher Ratio, x71 Prison Population Rate, x81 is Gross Domestic Savings.

Limitations to the project include that the data was not entirely available for a select number of observations for the years 2000 and 2010. In order to have a complete data set to work with, information had to be looked at from the years immediately before or after 2000 and 2010. This project is also limited by my own personal opinions since variables that were chosen were thought to have a possible impact on the average PISA score.

The predictions were as follows leaving all else constant for all variables, it is predicted that the Enrollment Rate of both sexes will have a positive effect on the rank of education because it is believed that having a higher enrollment rate of both sexes will result in better overall test scores. The Education Index will have a positive effect on the rank of education because having high test scores is a good indicator schools are retaining students. School Expenditure will have a negative effect on the rank of education because increasing educational spending does not necessarily produce positive results. Unemployment Rate will have a negative effect on the rank of education because we believe that the more people that are unemployed would likely be tied to a less educated population. Health Expenditure per capita will

have a positive effect on the rank of education because if health expenditure is rising then the population is less likely to attend school and is as sign that people are becoming sick more often. If a student is frequently missing school that they would not perform well on standard tests. The Ratio of Pupils to Teacher will have a positive effect on the rank of education because having a low student to teacher ratio indicates there is more personal involvement between students and teachers. Prison Population Rate will have a negative effect on the rank of education because of the negative psychological effects that having a parent in prison as well as the likelihood of a child going to prison if their parent has been there. Gross Domestic Savings will have a positive effect on the rank of education because a positive gross domestic savings rate would indicate that a country has money to invest in education.

4   Methods

The method to calculate results for this interesting topic is not complex. The methods for this project consisted of using a Cross-Sectional analysis following an Ordinary Least Squares Regression on the Years 2000 and 2010.

The primary empirical model for Year 2000 can be depicted in the fol-

lowing equation:

$$y = \beta 0 + \beta 1 x 1 + \beta 2 x 2 + \beta 3 x 3 + \beta 4 x 4 + \beta 5 x 5 + \beta 6 x 6 + \beta 7 x 7 + \beta 8 x 8$$

While the primary empirical model for Year 2010 can be depicted in the following equation:

$$y = \beta 0 + \beta 11 x 11 + \beta 21 x 21 + \beta 31 x 31 + \beta 41 x 41 + \beta 51 x 51 + \beta 61 x 61 + \beta 71 x 71 + \beta 81 x 81$$

where y, labeled x for 2000 and x0 for 2010 in the data-set, is average PISA test score that dictates a country's education rank. While the x's are the variables of interest that impact y. A linear regression was used to determine the strength that the x's have on y in order to see which variables have more or less impact on PISA scores. Correlation, heteroskedasticity, and multicollinearity test were performed on both the Years 2000 and 2010 to address any possible issues with the data. Additionally, the models were tested to see how accurate they were at predicting. Predicting was done by dropping one variable from the the data-set for both years and rerunning a linear regression on their respected models. Then finding the difference between the original and prediction.

5 Findings

The following two regression models with their statistics can be seen in Tables 1 & 2. In Table 1, the Year 2000 had an adjusted R^2 of 0.5135, while in Table 2, the Year 2010 had and adjusted R^2 of 0.3252. However, there was only one significant variable in both the models less than a 5% significant level: Education Index. Additionally, the dependent variables Enrollment Rate, Unemployment Rate, and Health Expenditure had different impacts between the two years. The Enrollment Rate and Unemployment Rate had a positive impact in 2000 and a negative impact in 2010. While the Health Expenditure had a negative impact in 2000 and a positive in impact in 2010. Figures 1 & 2, show a correlation matrix graph of the individual correlation between each variables for both years. In both models, the presence of multicollinearity was checked for but none was found to exist. This was checked through the variance inflation factor (VIF) and can be seen in Tables 5 & 6.

After the original regression were performed, reduced models were made for both years with the significant variables that passed the 5% criteria level. The reduced equations for each year a depicted below:

9

Year 2000:

$$x = \beta 0 + \beta 2 x 2$$

Year 2010:

$$x0 = \beta 0 + \beta 21 x 21$$

In the reduced equations, it is noticed that there was still no presence of multicollinearity due to the fact that there is only one independent variable in each equation. In these models, there is no presence of heteroskedasticity and the adjusted R^2 value is lower than in the original model for the year 2000. There is no heteroskedasticity in 2010 for the adjusted model and has only a slightly higher adjusted R^2 value. For these reasons, the original two models are the ones that fit the data the best.

Since this is cross sectional data, testing was done on the original models for heteroskedasticity and noted that there was none as seen in Figures 3 & 4, by observing the Residuals vs Fitted graph. If there had been heteroskedasticity, it would have been concerning even though the R^2 were larger, that these models would not be efficient. Furthermore, it would imply that the best linear unbiased estimators were not correctly selected. Since the models do not have this issue, we know that we can rely on the inferential procedures. However, the limitations must still be known, due to the majority of

the variables being insignificant.

In Table 7, the prediction accuracy of the models can be seen for both years. To check the accuracy a observation was dropped in order to measure the accuracy. Australia was taken out of Year 2000 and 2010 data-sets. As a result, the number of observations decreased from 30 to 29, which impacts the models predictions from the original regression that was performed with 30 observations. A new regression was performed on the new data-set without Australia to be used to predict their average test score. This data was plugged into two new regression models, representing Year 2000 and 2010. Observing the Table 7, Year 2010 was closer at predicting Australia's average test score than Year 2000. Even though the reduced models included one dependent variable, the prediction scores were closer to Australia's original average test scores. Showing the impact the Education Index has on average PISA test scores by itself.

6   Conclusion

Though there were imperfections with the model and data, it is believed that the research that has been done is a good stepping-stone for future stud-

ies on this topic. The significant finding from this project was the significance of the Education Index. Another finding was the prediction accuracy of the models. The Year 2010 model performed better even though the adjusted R^2 was lower than Year 2000. Studying this topic more in-depth can provide clearer answers as to why the 2010 model performed better, even though it had a lower R^2. Additionally, looking further into the Education Index would bring a better understanding of its impact and its study could provide beneficial impacts.

There are other variables that could be considered and this topic lends itself to a time series studies of the data as well. Education is a subject that has many facets to consider. There are also many other factors that cannot be measured or are extremely difficult to measure. Such effects could be psychological trauma to a child during their lifetime. Such trauma can be hard to measure and taken into account for when attempting to model. However, psychological trauma does play a part of a child's education and psychological development, which can negatively impact them. Though there were not many significant variables, some of them can impact a child's learning and furthermore impact a country's education rank.

References

Education Index Human Development Reports. 2018. "Human Development Reports." URL `http://hdr.undp.org/en/content/education-index`.

Government Expenditure on Education Total Data. 2018. "Government expenditure on education, total." URL `https://data.worldbank.org/indicator/SE.XPD.TOTL.GD.ZS`.

Gross Domestic Savings. 2018. "Gross domestic savings." URL `https://data.worldbank.org/indicator/NY.GDS.TOTL.ZS`.

Gross Enrollment Ratio Primary Both Sexes Data. 2018. "Gross enrollment ratio, primary, both sexes." URL `https://data.worldbank.org/indicator/SE.PRM.ENRR?view=chart`.

Health Expenditure Per Capita. 2018. "Health expenditure per capita." URL `https://data.worldbank.org/indicator/SH.XPD.PCAP`.

Hong, Jieun. 2015. "Effects of Education Policies and Institutions on Student Performance." *Seoul Journal of Economics* .

Lips, Dan, Shanea J Watkins, and John Fleming. 2008. "Does Spending

More on Education Improve Academic Achievement? Backgrounder. No. 2179." *Heritage Foundation* .

Morsy, Leila and Richard Rothstein. 2016. "Mass incarceration and children's outcomes: Criminal justice policy is education policy." *Washington, DC: Economic Policy Institute* .

Mosteller, Frederick. 1995. "The Tennessee study of class size in the early school grades." *The future of children* :113–127.

Shepherd, Jessica. 2010. "World education rankings: which country does best at reading, maths and science?" *The Guardian* 7.

Stanat, P, C Artelt, J Baumert, E Klieme, M Neubrand, M Prenzel, U Schiefele, W Schneider, G Schümer, K Tillmann et al. 2002. "PISA 2000: Overview of the study—Design, method and results." *Berlin: Max Planck Institute for Human Development* .

UIS. 2018. URL `http://data.uis.unesco.org/?queryid=180`.

Unemployment Rate Data. 2018. "Unemployment, total (modeled ILO estimate)." URL `https://data.worldbank.org/indicator/SL.UEM.TOTL.ZS?end=2017page=5start=2000`.

World Prison Brief. 2018. "World Prison Brief data." URL http://www.prisonstudies.org/world-prison-brief-data.

7   Tables

Table 1: Year 2000 Regression

	Dependent variable:
	x
x1	0.165 (0.301)
x2	321.682*** (71.789)
x3	−0.484 (3.965)
x4	0.016 (1.370)
x5	−0.002 (0.005)
x6	1.181 (0.874)
x7	−0.043 (0.027)
x8	−0.258 (0.557)
Constant	232.685*** (50.959)
Observations	30
R^2	0.648
Adjusted R^2	0.513
Residual Std. Error	21.719 (df = 21)
F Statistic	4.826*** (df = 8; 21)

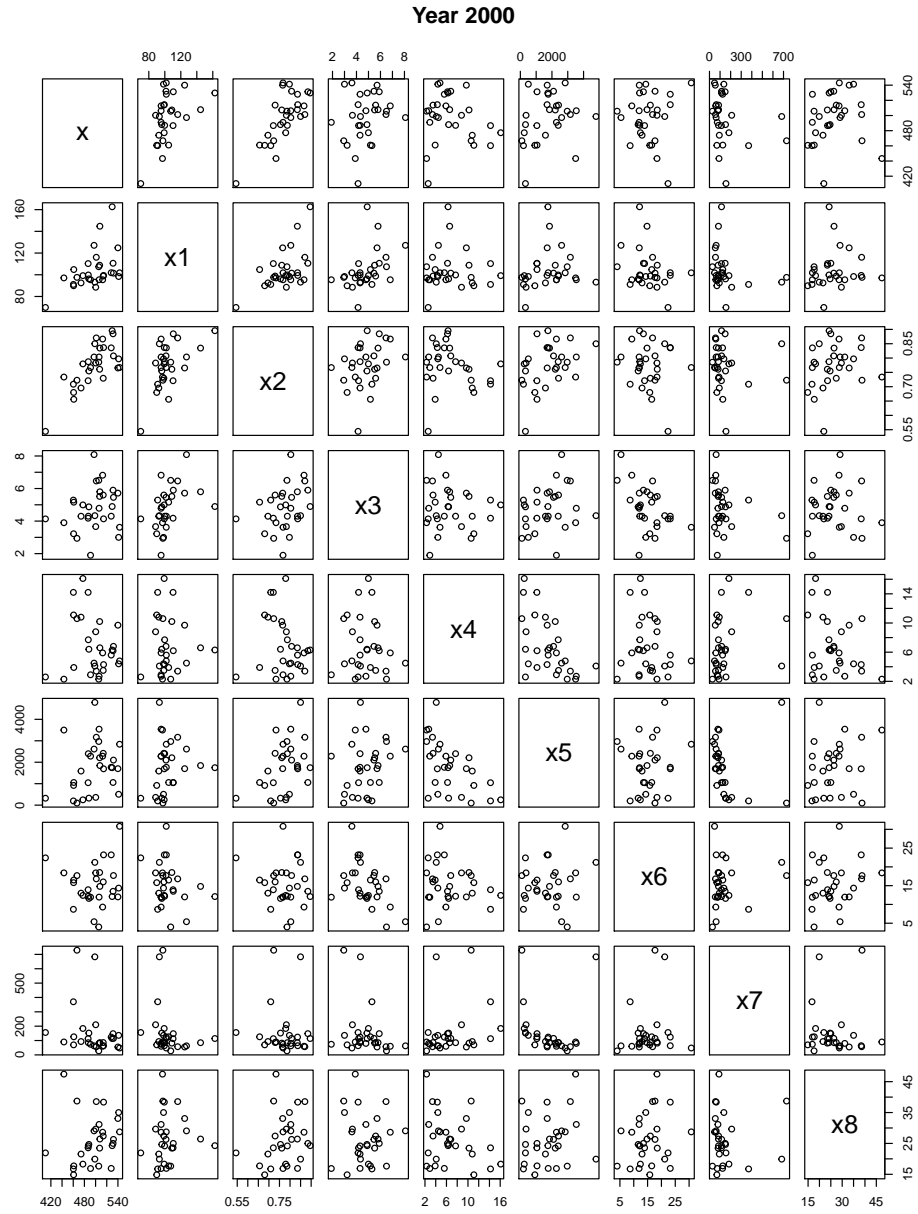*Note:* *p<0.1; **p<0.05; ***p<0.01

Figure 1: Year 2000 Correlation

Year 2000

17

Table 2: Year 2010 Regression

	Dependent variable:
	x0
x11	−0.044 (0.494)
x21	261.602*** (75.110)
x31	−1.453 (3.230)
x41	−0.656 (1.198)
x51	0.0004 (0.002)
x61	0.234 (0.778)
x71	−0.053 (0.031)
x81	−0.222 (0.500)
Constant	309.655*** (59.026)
Observations	30
R^2	0.511
Adjusted R^2	0.325
Residual Std. Error	19.888 (df = 21)
F Statistic	2.747** (df = 8; 21)

Note: *p<0.1; **p<0.05; ***p<0.01

Table 3: Reduced Year 2000 Regression

	Dependent variable:
	x
x2	312.511*** (52.279)
Constant	255.274*** (40.657)
Observations	30
R^2	0.561
Adjusted R^2	0.545
Residual Std. Error	21.004 (df = 28)
F Statistic	35.734*** (df = 1; 28)

Note:	*p<0.1; **p<0.05; ***p<0.01

Table 4: Reduced Year 2010 Regression

	Dependent variable:
	x0
x21	239.774*** (57.750)
Constant	300.760*** (47.701)
Observations	30
R^2	0.381
Adjusted R^2	0.359
Residual Std. Error	19.384 (df = 28)
F Statistic	17.239*** (df = 1; 28)

Table 5: Year 2000 VIF

x1	x2	x3	x4	x5	x6	x7	x8
1.694	1.764	1.631	1.680	1.791	1.437	1.201	1.182

Table 6: Year 2010 VIF

x11	x21	x31	x41	x51	x61	x71	x81
2.224	1.607	1.360	1.729	1.719	1.205	1.659	1.304

Table 7: Prediction Accuracy: Australia

Model	Original Observation	Prediction	Difference
Year 2000	529.7	561.1137	31.4137
Reduced 2000	529.7	535.8616	6.1616
Year 2010	518.7	529.7426	11.0427
Reduced 2010	518.7	521.3722	2.6722

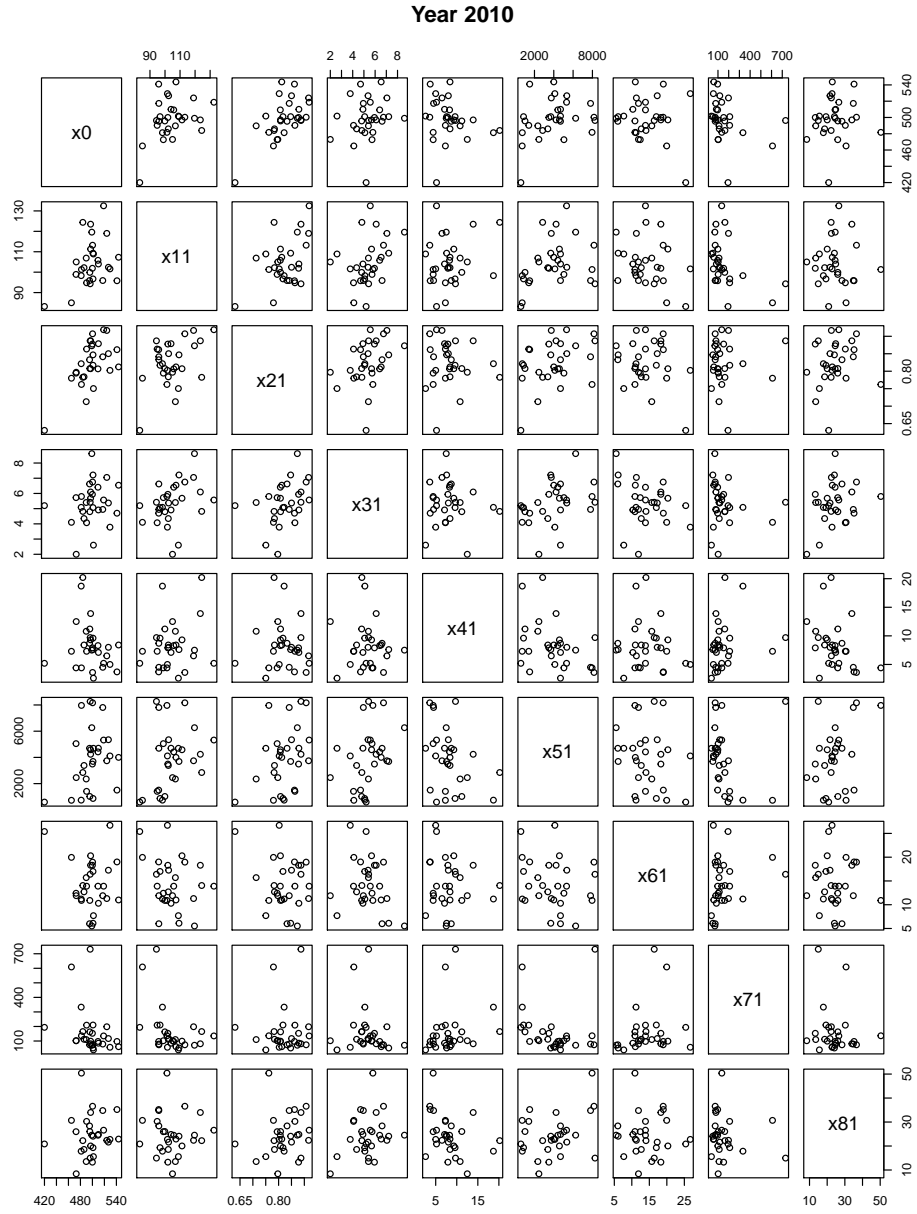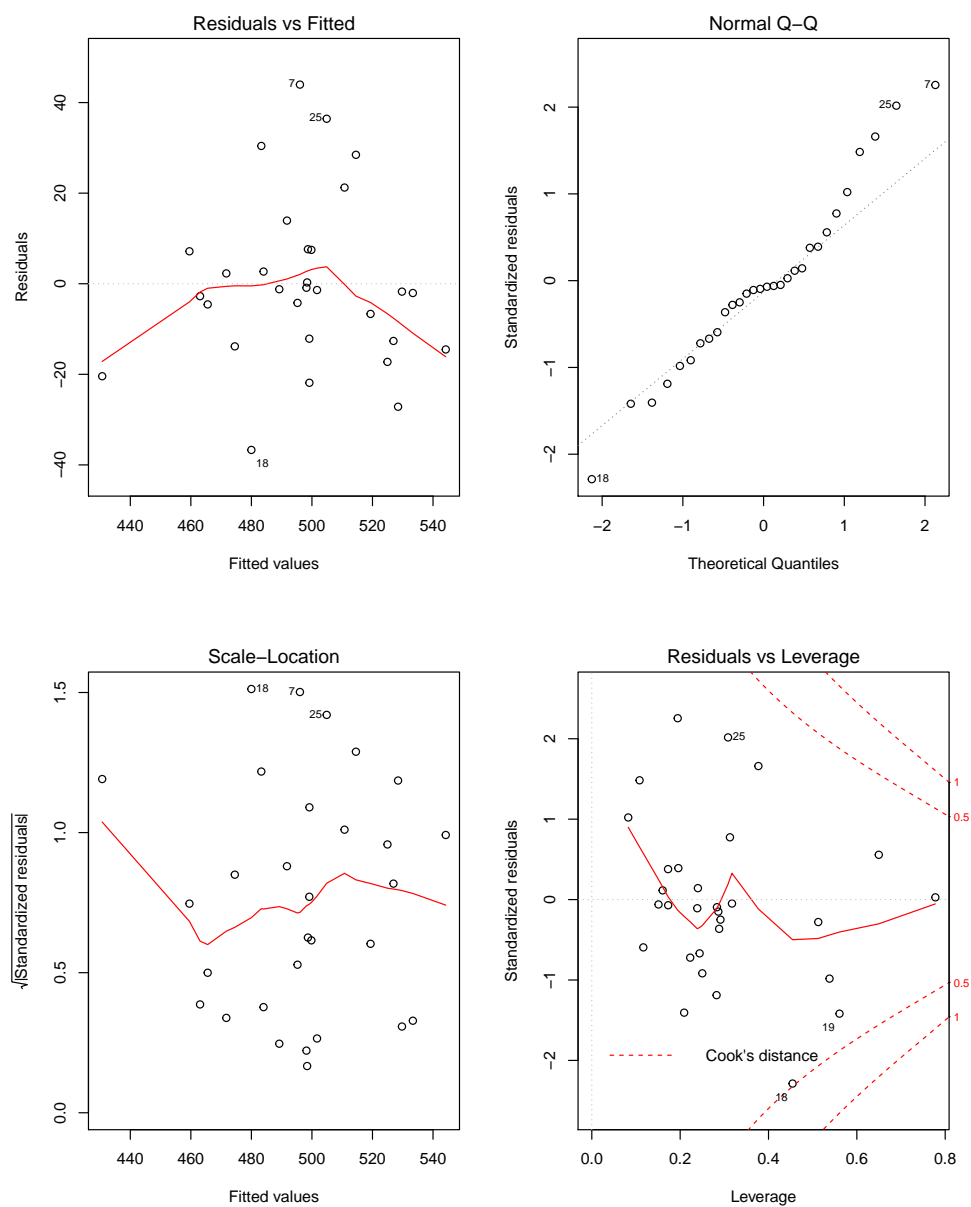Figure 2: Year 2010 Correlation

Year 2010

Figure 3: Observing Year 2000 Heteroskedasticity

Figure 4: Observing Year 2010 Heteroskedasticity