**PubPol 542 Final Report - To Paris or Not to Paris?**

Andrés Bas, Kamal Raslan, Sung ho Park, Sanjana Prabhu

## Introduction

The focus of this report is to understand how specific country level factors- namely corruption, economic freedom, government fragility and technology- influence the likelihood of a country participation in a climate agreement- focusing specifically on the Paris Agreement. Two factors that have also been considered are population and GDP per capita, as we are cognizant to the influence these two factors would have on the likelihood of participation in the agreement and wish to control for it.

We have attempted to (1) cluster countries on the basis of population and GDP per capita in order to understand if the countries are similar to one another and (2) understand the influence of the chosen variables on the likelihood to join the Paris Agreement through the use of logit regression.

The data was collected from the QOG basic database and the 2019 Index of Economic Freedom from The Heritage Foundation. Our dependent variable is if the country is part of the Paris Agreement or not, which is a binary variable. The IVs are corruption, economic freedom, government fragility and technology- all indexes and continuous variables.

In order to cluster countries on the basis of population and GDP per capita, we chose to use density based clustering, as we expected there to be outliers on the basis of population (such as China and India) and outliers on the basis of GDP per capita (such as Qatar and Luxembourg). Given this, density based clustering would be the best option as it allows us to minimize the influence of outliers on the clustering by assigning them a separate group (indicated by group 0). We ended up, in the process of clustering, dropping the following countries due to missing data- Liechtenstein, Somalia and Syrian Arab. Considering that we were using 2 variables to do the clustering, we set the minimum number of neighbours to be 3 and set the distance as 0.03.

We chose to use logistic regression specifically because we wanted to understand the likelihood of a country participating in the agreement, basis which we created a binary variable that indicates whether a country was part of the agreement or not. Given that our dependent variable is binary and given what we wanted to understand from the regression, logistic regression was our best option.

**Analyzing the results of Density-based Clustering**

Looking at the DB scan (see Exhibit 1) from the density-based clustering model, we identify 8 countries that serve as outliers. These countries were China, India, USA, Qatar, Luxembourg, Indonesia, Brazil, and Singapore. Excluding the outliers, the DB scan reveals three clusters.

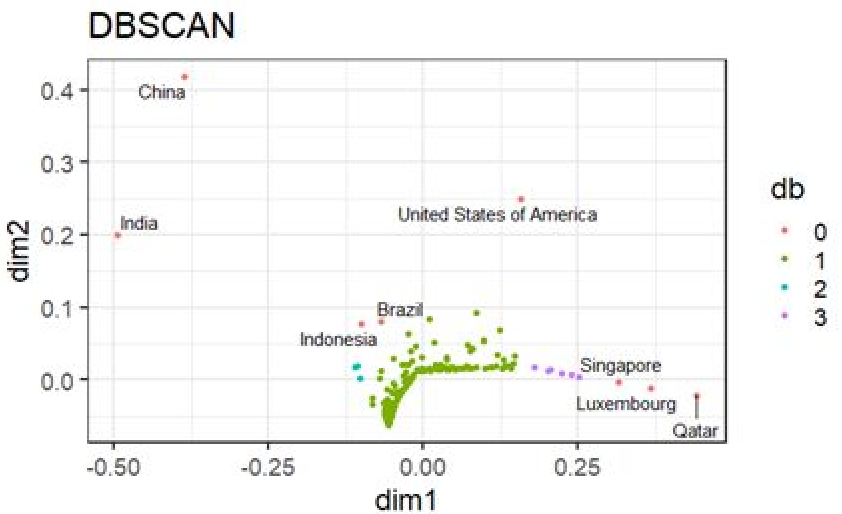



Exhibit 1. DB scan plot based on two dimensions (Population & GDP per Capita)

Looking at the results of the clustering model, the first cluster consists of 162 countries that are defined by moderate population size and GDP per capita. This cluster, on average, has a population size of 20.18 million and a GDP per capita of 16,501 (USD) (see Exhibit 2). Examples of countries included in this cluster are Cuba, Denmark, and Dominica, just to name a few.

The second cluster consists of three countries that are characterized as having a very large population size but low GDP per capita. This cluster, on average, has a population size of 183.1 million and a GDP per capita of 5,166 (USD). The countries included in this cluster are Bangladesh, Nigeria, and Pakistan.

The third cluster consists of six countries that can be defined as having a very small population size while maintaining high GDP per capita. This cluster, on average, has a population size of 5.55 million and a GDP per capita of 70,149 (USD). Examples of countries in this cluster include the United Arab Emirates and Norway.

Apart from the 171 countries that were categorized into one of the three clusters, there remain 8 noise points. These 8 countries do not fit into any of the categories because either their population size is extremely large with moderately high levels of GDP per capita (ex: China) or their population size and GDP per capita is very large (ex: United States of America).

| cluster | mean population | mean GDP per Capita |
|---|---|---|
| **0** | 438.9 | 54517 |
| **1** | 20.18 | 16501 |
| **2** | 183.1 | 5166 |
| **3** | 5.55 | 70149 |

Exhibit 2. The Mean Population & Mean GDP per Capita for each cluster category

**Hypotheses and Rationales of Chosen Regression Models**

The United Nations Framework Convention on Climate Change began in 1992 in response to the growing fear of anthropogenic climate instability. Since then, the United Nations and other global global climate conferences have created accords, protocols, and non-binding standards to combat the climate crisis. These include such conferences as The Montreal Protocol, The Kyoto Protocol, and The Paris Climate Agreement. Each of these agreements set standards for greenhouse gasses, endangered species management, and individual countries' plans to move towards a green economy.

Today, The Paris Climate Accords are the most widely adopted of all UNFCCC accords, likely due to the growing concerns regarding the impacts of climate change and reduced cost of renewable energy technology. However, it still took many years to reach full retention across the UNFCCC despite the clear directionality of the Paris Accords. Our regression is meant to test the relationship between certain characteristics of a nation and whether or not they are original adopters of the Paris Climate Accords, We determined that an original adopter would be identified as a country which ratified the 2015 agreement before the end of 2016.

We hypothesize that certain positive characteristics such as GDP per capita, score on the

Economic Freedom Index, and score on a Technological Development Index would be indicative of an original adopter. GDP per capita may correlate to more spending power and financial security leading to a country's ability to consider the global implications of climate change and to act domestically. Economic freedom is determined by whether or not the nation interferes on the free market or not and therefore may force consumers to drive the economic and political pressure to limit climate change as the costs associated are beginning to outweigh the costs of individual and state level action. Lastly, because technological solutions are necessary to achieve the goals in the Paris accords, it is reasonable to think that technologically developed nations would be more willing to act first on these accords.

We also tested the relationship between certain negative characteristics and hypothesize that countries that exhibit these characteristics will be less likely to adopt the Paris Accords in its inaugural year. These characteristics include State Fragility and Corruption. We controlled and tested for State Fragility due to the overwhelming environmental cost and financial cost of political unrest. In more fragile countries, it is assumed that there would be less pressure to adopt costly and resource intensive goals to mitigate the causes of climate change. Corruption is measured by the outwardly perceived levels of public sector corruption. We believe that nations with many corrupt actors would be less willing to adopt policy to promote environmental targets that could negatively impact corrupt actors' bottom lines.

Lastly, we believe that population will have no impact on original adoption of the Paris Climate Accords. We hypothesize this because there is very little evidence that nations with either large or small populations are more willing to act on environmental legislation or not. Non-adherence can be attributed to political dissonance in large populations or financial and economic need in smaller less populous countries. Population can also force large nations to better understand their role in the global climate framework and be inspired to action and small nations may be faster to adopt climate policy due to the existential threat of many small island nations that can potentially confound the data. It is for this reason that we believe that this factor will not impact any nation's adoption of the Paris Agreement.

**Regression Methodologies**

As mentioned earlier, we use logistic regressions to test our hypotheses, which is the appropriate model when the outcome variable is binary. We run six different regression configurations with three different sample sizes each time, totaling 18 regression runs. We start by testing the effects of economic freedom on the likelihood of signing the Paris agreement, and then add one extra variable in each subsequent regression. The six regressions that we compare read as follows:

1. Paris_OA ~ Economic_Freedom
2. Paris_OA ~ Economic_Freedom + Population
3. Paris_OA ~ Economic_Freedom + Population + GDPperCapita
4. Paris_OA ~ Economic_Freedom + Corruption + Population + GDPperCapita
5. Paris_OA ~ Economic_Freedom + Corruption + State_Fragility + Population + GDPperCapita
6. Paris_OA ~ Economic_Freedom + Corruption + State_Fragility + Gov_Technology + Population + GDPperCapita

The state fragility variable has 18 countries with missing values, and the government technology variable has 1 country with missing values. Since it would be incorrect to compare among models with different variables *and* different data sets, we need to run all our models with the three different data sets. This allows us to search for the best fitting model in a rigorous way—we cannot just assume that dropping three countries does not affect the results. In other words, we need to make sure that the regressions results in, say, model 6, are not significantly different when using 179, 178, or 161 countries. Exhibit 3 below presents a summary of the 15 regressions and its coefficients.

| Data Set | 161 Countries (Dropping 18) | | | | | | 178 Countries (Dropping Congo Only) | | | | | 179 Countries (Full Sample) | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Regression Variables | 1 | 2 | 3 | 4 | 5 | 6 | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 |
| Economic Freedom | 0.030 | 0.030 | 0.025 | (0.002) | (0.005) | (0.004) | 0.030 | 0.030 | 0.026 | (0.009) | (0.012) | 0.031 | 0.031 | 0.027 | (0.009) |
| *P-value* | *0.049** | *0.046 ** | *0.190* | *0.936* | *0.840* | *0.863* | *0.045** | *0.042** | *0.160* | *0.670* | *0.614* | *0.037** | *0.035** | *0.149* | *0.686* |
| Population | | 0.005 | 0.005 | 0.006 | 0.006 | 0.008 | | 0.003 | 0.003 | 0.004 | 0.004 | | 0.003 | 0.003 | 0.004 |
| *P-value* | | *0.192* | *0.190* | *0.130* | *0.157* | *0.104* | | *0.309* | *0.308* | *0.239* | *0.269* | | *0.312* | *0.310* | *0.265* |
| GDP per Capita | | | 4.9E-06 | -1.1E-05 | -1.3E-05 | -1.4E-05 | | | 3.8E-06 | -1.5E-05 | -1.7E-05 | | | 4.2E-06 | -1.5E-05 |
| *P-value* | | | *0.629* | *0.348* | *0.317* | *0.297* | | | *0.699* | *0.181* | *0.181* | | | *0.672* | *0.187* |
| Corruption | | | | (0.045) | (0.044) | (0.043) | | | | (0.058) | (0.057) | | | | (0.058) |
| *P-value* | | | | *0.018** | *0.021** | *0.025** | | | | *0.001*** | *0.002*** | | | | *0.001*** |
| Government Technology | | | | | 0.444 | (0.762) | | | | | 0.372 | | | | |
| *P-value* | | | | | *0.733* | *0.661* | | | | | *0.773* | | | | |
| State Fragility | | | | | | (0.057) | | | | | | | | | |
| *P-value* | | | | | | *0.287* | | | | | | | | | |

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Exhibit 3. Summary of Regression Coefficients by Model and Data Set

To compare the model fits we use the likelihood ratio test (LRT), within models that are using the same number of countries. The tests are similar to using the R-squared, but for logistic regressions. Interestingly, the best fitting model is number 4, which excludes the state fragility and government technology variables. Therefore, we can use our original, full sample size data set. Exhibit 4 presents the results from the LRTs conducted between configurations 4, 5, and 6, under the reduced data set case (for all other tests see https://github.com/shp9026/govanalytics/raw/master/Cluster%26Regression.Rmd). The test results tell us that adding the variable Corruption improves the fit of the model, but subsequently adding Government Technology and State Fragility do not. Exhibit 5 shows the detailed results of the model we choose for our analysis. See appendix for plots with a visual representation of the distribution of our selected regression.

```
Model 1: fromPy3$Paris_OA ~ +fromPy3$X2019Score + fromPy3$Population +
    fromPy3$GDPperCapita
Model 2: fromPy3$Paris_OA ~ +fromPy3$X2019Score + fromPy3$bci_bci + fromPy3$Population +
    fromPy3$GDPperCapita
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1       154     206.80
2       153     200.84  1   5.9616  0.01462 *
```

```
Model 1: fromPy3$Paris_OA ~ +fromPy3$X2019Score + fromPy3$bci_bci + fromPy3$Population +
    fromPy3$GDPperCapita
Model 2: fromPy3$Paris_OA ~ +fromPy3$X2019Score + fromPy3$bci_bci + fromPy3$egov_egov.y +
    fromPy3$Population + fromPy3$GDPperCapita
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1       153     200.84
2       152     200.72  1  0.11648   0.7329
```

```
Model 1: fromPy3$Paris_OA ~ +fromPy3$X2019Score + fromPy3$bci_bci + fromPy3$egov_egov.y +
    fromPy3$Population + fromPy3$GDPperCapita
Model 2: fromPy3$Paris_OA ~ +fromPy3$X2019Score + fromPy3$bci_bci + fromPy3$cspf_sfi +
    fromPy3$egov_egov.y + fromPy3$Population + fromPy3$GDPperCapita
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1       152     200.72
2       151     199.57  1   1.1494   0.2837
```

Exhibit 4. Selected LRTs with smaller data set

| Variables | Coefficient | Std. Error | z value | Pr(>\|z\|) |
|---|---|---|---|---|
| Intercept | 3.9440 | 1.9040 | 2.0710 | 0.03837* |
| Economic Freedom | -0.0087 | 0.0215 | -0.4050 | 0.6858 |
| **Corruption** | **-0.0582** | **0.0180** | **-3.2240** | **0.00126**** |
| Population | 0.0041 | 0.0036 | 1.1140 | 0.2655 |
| GDP per Capita | 0.0000 | 0.0000 | -1.3200 | 0.1868 |

Exhibit 5. Regression Results in Best Fitting Model (4 variables and full data set)

**Regression Results & Discussion**

The final determination of the best fitting regression that was run was that BCI was the only indicator that statistically impacted whether or not a country was an original adopter of the Paris Climate Accords. As mentioned previously other indicators such as State Fragility and Technological Development led to worse fitting models and were determined to be insignificant controls. This determination shows that there we can say with 99% certainty that a country ranked higher on the BCI Corruption Index was less likely to adopt the paris accords in the first year of enactment.

Even when controlling for more variables with slightly reduced sample size, corruption was still the only statistically significant explanatory variable for adoption of the Paris Accords. The results of these regressions with additional controls are shown in Exhibit 6. When comparing these results with our hypothesis we reject the notion that subjective positive characteristics of a country are indicative of adherence to global climate agreements, and reject the hypothesis that fragility contributes to a lower likelihood of adoption. By rejecting these hypotheses, we are able to determine that technological development, economic freedom, and GDP, which have the capacity to build a nation's ability to fight climate change, do not move the needle on the political will for collective climate action. In our results we recommend those working to create tangible changes in the face of global climate change begin by tackling corruption head-on in order to have adherence and adoption of future UNFCCC climate accords.

Exhibit 6. Unused Regression Results (Tested Hypotheses 5-10)

Hypotheses 5: Dropping no Attributes -- Technological Development Additionally Controlled

| Variables | Coefficient | Std. Error | Z Value | Pr(>\|z\|) |
|---|---|---|---|---|
| Intercept | 2.559 | 1.98 | 1.292 | 0.1963 |
| Economic Freedom | -0.004697 | 0.02323 | -0.202 | 0.8398 |
| Corruption | -0.04418 | 0.01907 | -2.317 | 0.0205* |
| Population | 5.87E-03 | 4.15E-03 | 1.415 | 0.1571 |
| GDP per Capita | -1.33E-05 | 1.33E-05 | -1.001 | 0.317 |
| Technological Development | 4.44E-01 | 1.30E+00 | 0.342 | 0.7327 |

Hypotheses 6: Dropping no Attributes -- Technological Development & Fragility Addionally Controlled

| Variables | Coefficient | Std. Error | Z Value | Pr(>\|z\|) |
|---|---|---|---|---|
| Intercept | 3.5200 | 2.1880 | 1.6090 | 0.1077 |
| Economic Freedom | -0.0040 | 0.0234 | -0.1730 | 0.8625 |
| Corruption | -0.0428 | 0.0191 | -2.2470 | 0.0247* |
| Fragility Index | -0.0566 | 0.0531 | -1.0660 | 0.2865 |
| Technological Development | -0.7620 | 1.7370 | -0.4390 | 0.6609 |
| Population | 0.0075 | 0.0046 | 1.6280 | 0.1036 |
| GDP per Capita | 0.0000 | 0.0000 | -1.0440 | 0.2965 |

Hypotheses 10: Dropping One Attribute -- Technological Development Addionally Controlled

| Variables | Coefficient | Std. Error | Z Value | Pr(>\|z\|) |
|---|---|---|---|---|
| Intercept | 3.9540 | 1.9050 | 2.0760 | 0.03793* |
| Economic Freedom | -0.0091 | 0.0215 | -0.4260 | 0.6701 |
| Corruption | -0.0577 | 0.0180 | -3.1970 | 0.00139** |
| Population | 0.0045 | 0.0038 | 1.1770 | 0.2391 |
| GDP per Capita | 0.0000 | 0.0000 | -1.3370 | 0.1811 |

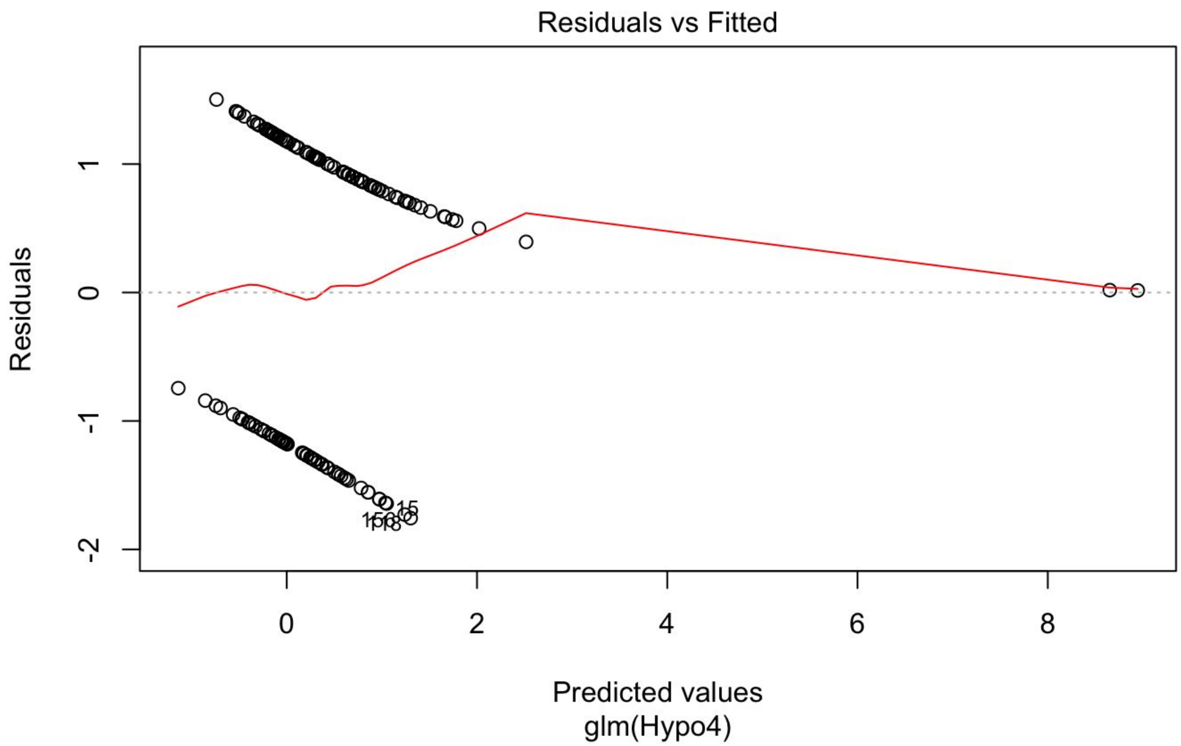
Hypotheses 11: Dropping One Attribute -- Technological Development & Fragility Addionally Controlle

| Variables | Coefficient | Std. Error | Z Value | Pr(>\|z\|) |
|---|---|---|---|---|
| Intercept | 3.9190 | 1.9080 | 2.0540 | 0.03999* |
| Economic Freedom | -0.0117 | 0.0232 | -0.5040 | 0.6140 |
| Corruption | -0.0572 | 0.0181 | -3.1580 | 0.00159** |
| Technological Development | 0.3723 | 1.2880 | 0.2890 | 0.7726 |
| Population | 0.0042 | 0.0038 | 1.1050 | 0.2690 |
| GDP per Capita | -1.70E-05 | 1.27E-05 | -1.339 | 0.18072 |

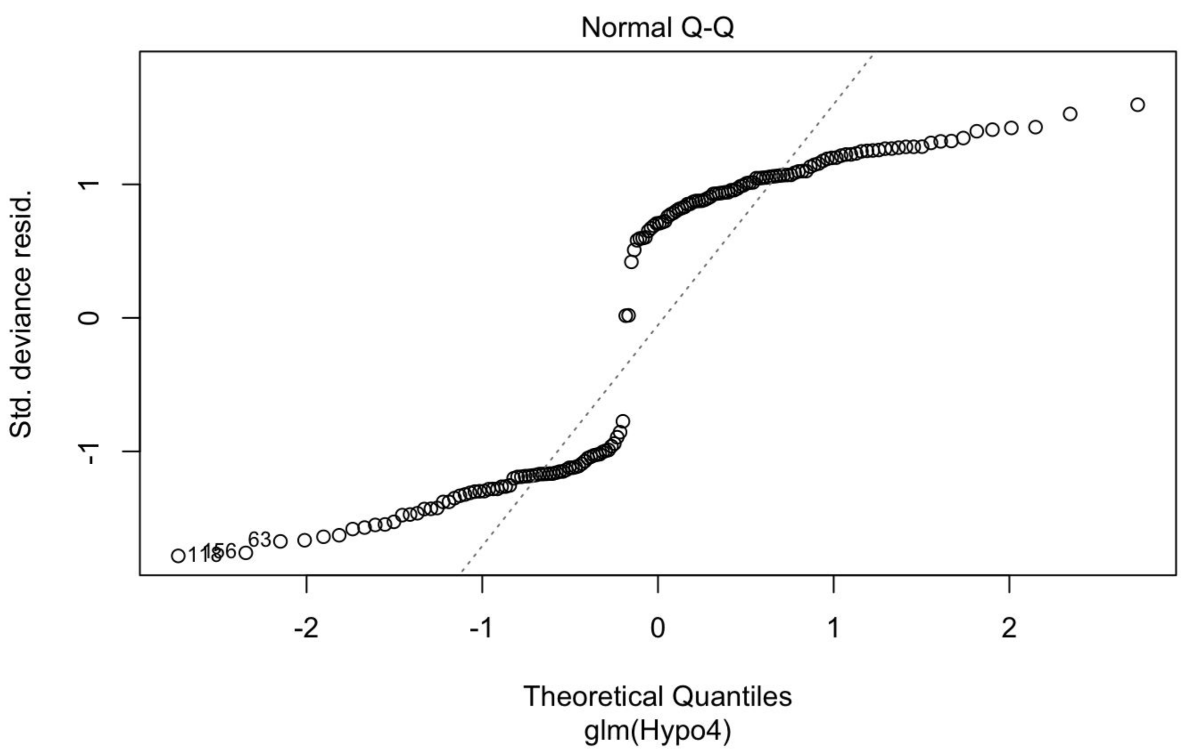
Hypotheses 15: Dropping Many Attributes -- No Addional Controls

| Variables | Coefficient | Std. Error | Z Value | Pr(>\|z\|) |
|---|---|---|---|---|
| Intercept | 3.94400 | 1.90400 | 2.07100 | 0.03837* |
| Economic Freedom | -0.00868 | 0.02147 | -0.40500 | 0.68582 |
| Corruption | -0.05815 | 0.01804 | -3.22400 | 0.00126** |
| Population | 0.00405 | 0.00364 | 1.11400 | 0.26547 |
| GDP per Capita | -0.00002 | 0.00001 | -1.32000 | 0.18680 |

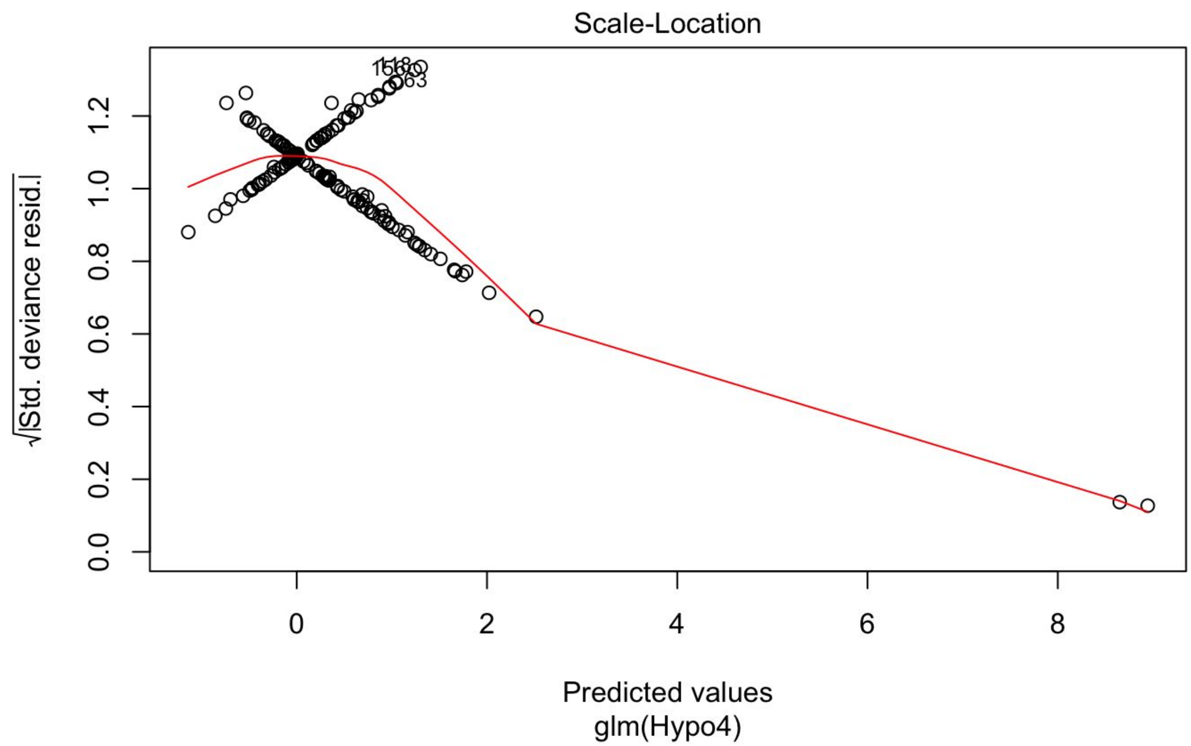
## Appendix



Residuals vs Fitted
Residuals
1
0
-1
-2
0
2
4
6
8
Predicted values
glm(Hypo4)

Normal Q-Q
Std. deviance resid.
1
0
-1
63
-2
-1
0
1
2
Theoretical Quantiles
glm(Hypo4)


Scale-Location
√|Std. deviance resid.|
Predicted values
glm(Hypo4)
118
156
163
0.0
0.2
0.4
0.6
0.8
1.0
1.2
0
2
4
6
8


Residuals vs Leverage
Std. Pearson resid.
Leverage
glm(Hypo4)
Cook's distance
88
122
133
-2
-1
0
1
0.00
0.05
0.10
0.15
0.20