What Factors Affect the Rate of Undernourishment?

An Exploratory Longitudinal Analysis

Risako Owan Il Shan Ng

Professor Laura Chihara
MATH 315: Advanced Statistical Modelling

June 6 2016

1 Introduction

Extreme hunger and undernourishment in certain parts of the world are issues that continue to top the list of most pressing global problems.[1] Despite persistent attempts to alleviate world hunger, about one in eight people on Earth are still undernourished, and one in four children experience stunted growth.[2] Unsurprisingly, undernourishment haunts the agenda of world leaders and organizations. Every U.S. President since Jimmy Carter, who signed into effect the Presidential Commission on World Hunger in 1978, has expressed a commitment to fight world hunger through a thorough review of food production and distribution policies in the poorest parts of the world.[3] Perhaps more widely known is the explicit target of the World Health Organization's Millennium Development Goals to "halve, between 1990 and 2015, the proportion of people who suffer from hunger."[4] While the rate of progress has been satisfactory in some parts of the world, other regions have fallen completely short of the pledged target.

Evidently, tackling undernourishment is not an easy task. Intuitively, world hunger could be eradicated by simply making sure that there is enough food in the world to go around. This intuition does not fare well, however, when we consider the rather puzzling fact that global per capita food supply has risen by about 27% between 1960 and 2010.[5] In other words, there must be other forces

[1]World Economic Forum, https://www.weforum.org/agenda/2016/01/what-are-the-10-biggest-global-challenges/, last accessed Jun 4 2016.

[2]World Food Program, https://www.wfp.org/hunger/stats, last accessed June 4 2016.

[3]The American Presidency Project, http://www.presidency.ucsb.edu/ws/?pid=31265, last accessed June 4 2016.

[4]World Health Organization, http://www.who.int/mediacentre/factsheets/fs290/en/, last accessed June 4 2016.

[5]Food and Agricultural Organization 2013 Yearbook, http://reliefweb.int/report/world/fao-statistical-yearbook-2013-world-food-and-agriculture, last accessed Jun 4 2016.

at work that keep this abundant food supply out of reach for those who need it the most. This paper explores these alternative factors that may influence the rate of undernourishment over four different time periods in 67 countries. We find certain indicators of food security to be important to the rate of undernourishment, after controlling for each country's level of food production.

2 Methods

The World Health Organization defines undernourishment as being "below [the] minimum level of dietary energy consumption."[6] In searching for possible factors that may affect the rate of undernourishment, we looked to the Food and Agriculture Organization of The United Nations (FAO), which classifies potential factors along the four dimensions of food security — availability, access, utilization and stability.[7] We then obtained a comprehensive data set from the FAO and chose relevant data that fell in these four categories. Figures for the variable were reported for each country in averages across overlapping three-year periods (1990-92, 1991-93, 1992-94, *etc*) which, according to the FAO, was done to reduce the impact of possible estimation errors. We are therefore working with repeated measures data, which calls for a hierarchical model with two levels: level one being the time period and level two, the country. This analysis will be restricted to only four three-year periods — 2006-08, 2007-09, 2008-10, 2009-11 — for which data was most available. The following paragraphs describe in detail our response and explanatory variables. Given in parentheses are the abbreviated names that will be used in statistical summaries and parametric representations of our final model.

Rate of undernourishment (rate). This is the response variable in our analysis. We chose to use a rate instead of simply modeling counts since we expect the number of undernourished people in a country to be directly proportional to the country's population. The data set from the FAO provided us with raw data on the number of undernourished people in each country, together with the total population of that country in different 3-year time periods. Dividing the former by the latter, we constructed the rate of undernourishment, measured in number undernourished per 10,000

[6] Millennium Development Goals Indicators, http://mdgs.un.org/unsd/mdg/Host.aspx?Content=Indicators/OfficialList.htm, last accessed June 4 2016.

[7] Food and Agriculture Organization of The United Nations, http://faostat3.fao.org/download/D/*/E, last accessed June 4 2016.

people. Since the response variable is a rate, we will use a mixed-effects Poisson regression model.

Time period (time). This captures the systematic effects of time on the rate of undernourishment in each country, and is essential to the longitudinal nature of our analysis. We code the four three-year periods 2006-08, 2007-09, 2008-10 and 2009-11 as times 1,2,3 and 4 respectively. Since the World Health Organization has reported (somewhat modest) improvements in the world hunger situation, we expect the average rate of undernourishment to fall over these four time periods.

Average value of food production (production). This measures the per capita net value of food production in constant 2004-06 international dollars, and is an indicator of food availability. Despite the puzzling observation that an ever-increasing world food supply has not managed to end world hunger, we believe that locally abundant food production tends to lower the rate of undernourishment. This variable, then, controls for the relative economic size of the food production sector in the country to expose other variables that may work against an abundant food supply. Since food production data was recorded for each three-year period, we make this a level 1 variable in our hierarchical model.

Cereal import dependency ratio (cereal). This measures the percentage of domestic food supply that has been imported by any given country, an indicator of food supply stability. A positive value indicates that the country is a net importer, while a negative value indicates that the country is a net exporter. Since a larger positive ratio represents greater dependency on external food sources, we expect countries with larger positive ratios to have higher rates of undernourishment. We make this a level 1 variable, since data is available for each three-year period.

Domestic food price volatility (volatility). This is an index that measures the amount of variation in food prices from the long run trend. Although the calculations used by the FAO to produce this index are very complicated, its interpretation is straightforward: the closer the value to 100, the more volatile food prices were in any given three-year period. Like the cereal import dependency ratio, the FAO classifies food price volatility as an indicator of food supply stability. We make this a level 1 variable.

Political stability and absence of violence and terrorism (political). This is an index, ranging from -3 to 3, that measures "the likelihood that the government will be destabilized or overthrown by unconstitutional or violent means."[8] Such calamitous events disrupt both food supply and distribution, and so we expect more politically stable countries with a more positive index to have lower rates of undernourishment. We include this variable as a measure of food supply access and stability. Because we believe political stability to be a trait of each country that should not vary too much across the years covered by this analysis, we make it a level 2 variable by aggregating the data across the four three-year time periods into a single average for each country.

Access to improved water sources (water). This is the percentage of total population with reasonable access to improved drinking sources. An "improved drinking source" includes a household connection, public standpipe, borehole, protected well or spring, and rainwater collection. The FAO considers this a measure of the degree to which an available food supply may be utilized, since food in general cannot be cooked without clean water. We make this a level 2 variable since the data suggests little variation in access to water over our chosen time periods.

Average Dietary Energy Requirement (diet). This measures the average amount of energy required to sustain a healthy lifestyle in kilo calories per person per day. The data suggests considerable variation across different countries, and having a low energy requirement may lower the incidence of undernourishment. Like access to water, this variable does not change much over the short time frame of our analysis, and so we make it a level 2 variable.

Level of development (development). This is a categorical variable with the levels Low, Medium and High. We referred to a table from the Human Development Report 2007[9] that showed the Human Development Indices (HDI) for each country in 2007. We then classified these countries into those with low development (HDI of 0—0.499), medium development (HDI of 0.5—0.799) and high development (HDI of 0.8—0.899), according to the definitions given by the Human Development Report. A country's level of development serves as a general proxy for food access, since distribution infrastructure (roads, rail lines, *etc*) would presumably be better in developed countries. Since the level of development is a trait of each country that varies little over our short 5 year time frame, we make it a level 2 variable.

---

[8] Food and Agriculture Organization of the United Nations Statistics Division, http://faostat3.fao.org/download/D/*/E, last accessed Jun 4 2016.

[9] United Nations Development Program, http://hdr.undp.org/en, last accessed Jun 5 2016.

3 Results

3.1 Exploratory Data Analysis

Although the original data set covered 227 countries, both developed and undeveloped, data for our chosen explanatory and response variables were complete for only 68 countries. Among these, one country (Saint Vincent and the Grenadines) had abbreviated figures for its number of undernourished people, which made it impossible to calculate the response variable. We therefore drop this country from the data set and work with only 67 countries. With measurements for four time periods per country, we have a total of 268 observations, a large enough data set. Table 1 below gives summary statistics for the numeric variables used in our analysis.

Table 1: Summary statistics for numeric variables

Variable	Units	Min	Max	Mean	SD
rate	number undernourished per 10,000 people	425.5	5579	2025	1139.7
production	constant 2004-06 international dollars	26	689	195.5	107.6
cereal	percent of domestic food supply	-271.2	100	35.4	44.7
volatility	index between 0 and 100	3.9	82	14.1	10.9
political	index between -3 and 3	-2.5	1.1	-0.53	0.85
water	percent of total population	44.9	99.7	77.4	15.8
diet	kilo calories per person per day	2061	2489	2238	107.1

In order to avoid issues with model-fitting, we had to center and scale the explanatory variables that had large ranges and standard deviations: *production*, *cereal*, *volatility*, *water* and *diet*. The values actually used for these variables are therefore the z-scores, or number of standard deviations from the mean, of the original values. Figure 1 shows the distribution of the response variable. The rates seem to follow a Poisson distribution reasonably well (there are no excess zeroes, for example). There are a few outlying countries with more than half their populations classified as undernourished. We identified these to be Haiti and Zambia, but include them in the analysis nonetheless until more formal diagnostic procedures are carried out.

Figure 1: Histogram of the response variable

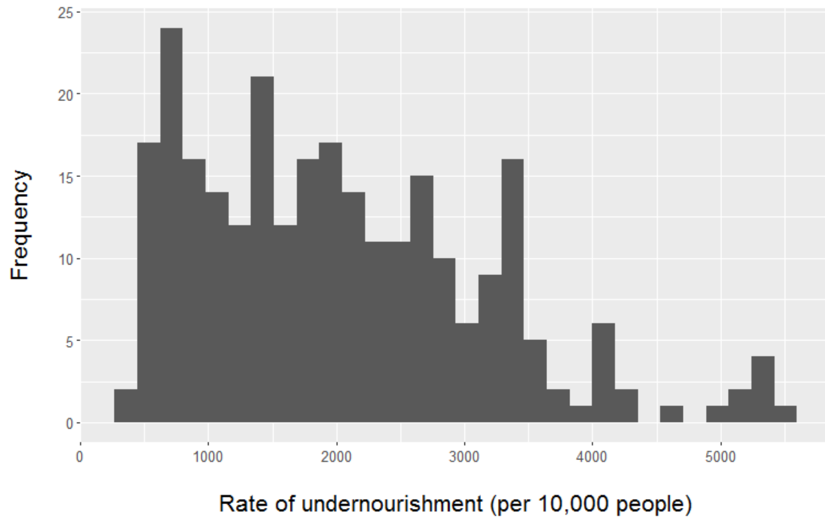
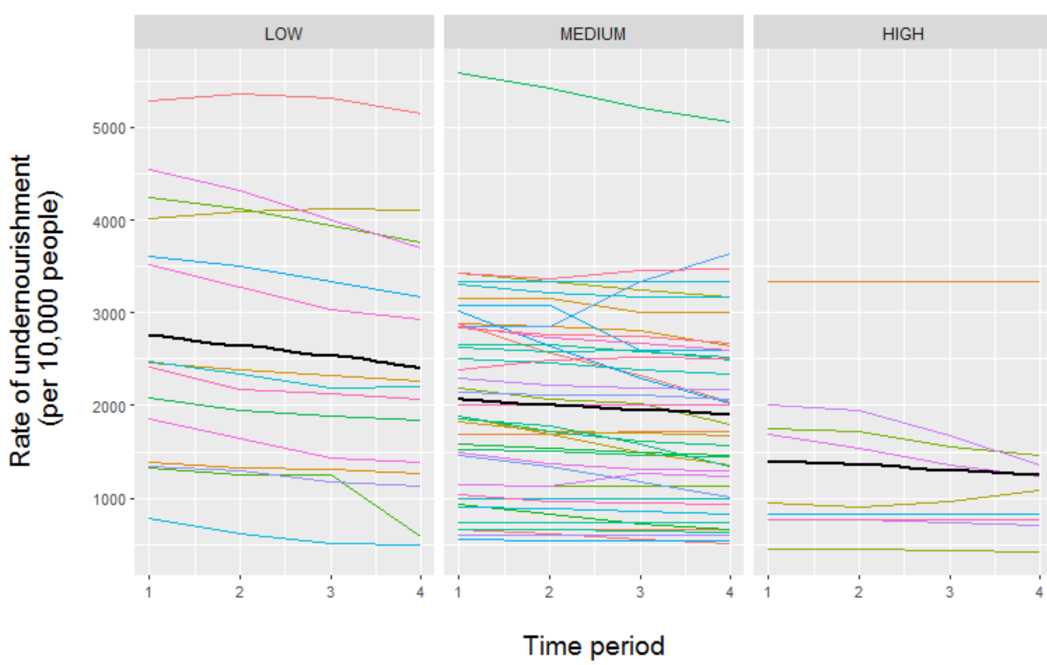


Figure 2 below shows the time trends for each country's rate of undernourishment, grouped by the country's level of development. The bold lines represent the mean trend for each group. We see a clear but modest decrease in the rate of undernourishment for all three groups — some indication of success for the World Health Organization. The rates of decrease also seem to differ among countries with different levels of development, which suggests an interaction term between *time* and *development*.

Figure 2: Time trends for the response grouped by level of development



Through more exploratory plots not shown here, we found that the data also suggests two other interaction terms: an interaction between political stability and level of development (Figure A1 in Appendix), and an interaction between food production and access to water (Figure A2 in Appendix).

3.2 Estimated Model

Our initial model included all explanatory variables and the three interaction terms identified earlier. In modeling the random effects, we could include only a random intercept and random slope for *time*, since including more random effects would cause the number of variance components to exceed the number of observations, as well as non-convergence issues in the model-fitting. We then refined our initial model by removing insignificant variables, performing likelihood ratio tests at each step to ensure that the removal of each variable did not significantly worsen the goodness of fit of the reduced model to the data. The final model is shown below in parametric form.

Level One (country i time j):

$$\log\left(\mathrm{E}[rate_{ij}]\right) = a_i + b_i time_{ij} + \gamma production_{ij}$$

Level Two (country i):

$$\begin{aligned} a_i &= \alpha_0 + \alpha_1 water_i + u_i \\ b_i &= \beta + v_i \end{aligned}$$

where u_i is the random intercept and v_i the random slope for *time*, with

$$\begin{bmatrix} u_i \\ v_i \end{bmatrix} \sim \mathrm{MVNorm}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_u^2 & \sigma_{uv} \\ \sigma_{uv} & \sigma_v^2 \end{bmatrix}\right).$$

Tables 2 and 3 give the estimates for the fixed effect parameters and variance components respectively. We see that each unit time shift to later periods (for example, from 2006-2008 to 2007-2009) is associated with an approximately 3.6% decrease in the rate of undernourishment in a country. This confirms our initial judgment from our exploratory plots that the rate of undernourishment has indeed decreased over the four time periods. Additionally, every standard deviation increase in the average value of food production is associated with a 18.8% decrease in the rate of undernourishment, which confirms our reason for including this variable. Finally, every standard deviation increase in the percentage of the population with access to improved water sources is associated with an approximately 18.6% decrease in the rate of undernourishment. A 95% confidence interval for this percentage decrease is $(6.75, 28.9)$, and so we are 95% confident that every standard deviation increase in the percentage of population with access to improved water sources is associated with a 6.75% to 28.9% decrease in the rate of undernourishment. The low p-value suggests that, after controlling for food production, access to water is strongly associated with the rate of undernourishment, and so we have identified one factor other than food production that is systematically important to the rate of undernourishment.

Table 2: Estimates for fixed effect parameters

Coefficient	Notation	Estimate	Standard error	z-value	p value
Intercept	α_0	-1.68	0.0657	-25.5	0 ***
time	β	-0.0364	0.00875	-4.16	0.0000314 ***
production	γ	-0.208	0.0649	-3.20	0.00138 **
water	α_1	-0.206	0.0682	-3.02	0.00251 **

Significance codes: 0 '***' 0.001 '**' 0.01 '*'

Table 3: Estimates for variance components

Component	Notation	Estimate
SE of random intercept	σ_u	0.476
SE of *time* random slope	σ_v	0.0154
Correlation between SE's	$\frac{\sigma_{uv}}{\sigma_u \sigma_v}$	0.01

3.3 Diagnostic Tests

We used Q-Q plots of the random effects to check for any extreme outliers that may be affecting the model estimates. Our initial Q-Q plot displayed a normal distribution for the random intercept, but showed a heavy tail for the random slope (Figure A3 in Appendix). Therefore, we removed countries with unusually small and large estimated slopes. The countries with smaller slopes included Indonesia, India, and Myanmar, while the country with the largest slope was Bangladesh. This implies that the rate of change in Bangladesh is much higher than in other countries. After removing these outliers, both our Q-Q plots displayed a nice normal distribution (Figure A4 in Appendix). A final plot of the raw residuals showed no extreme outliers in the data (Figure A5 in Appendix). Our model estimates are therefore reasonably accurate.

4 Discussion

From our results above, we see that the rate of undernourishment in a country decreases over time and as food production increases, possibly as a result of further development. Reasonable access to safe water is crucial as well, since it is determines whether the available food supply may be utilized through cooking.

Our analysis findings were consistent with others in that greater access to safe water is associated with a lower percentage of undernourishment. For example, safe water access improvement projects between 1970 and 1995 actually lead to decreased child malnutrition. [10]. However, conclusions related to average food productions were more complicated and did not always agree with our findings. For some countries, improved food distribution is more important and effective than improved production. [11] Our initial data set included information on the percentage of paved roads in a country to account for distribution but we were forced to remove this variable from our report because of the extreme sparseness in the collected data.

Our findings were also sometimes inconsistent with other similar studies in how development and political stability were not significant in our model [12]. This was very unexpected for us, and we suspect that this result is due to the fact that we did not have data on undernourishment in highly developed countries. The United Nations Development Program's Human Development Report categorizes countries into four human development index categories: low, medium, high, and very high. We will refer to these as Category 1,2,3,4 respectively in our following discussion. Our data set consisted of 60 Category 1 countries, 172 Category 2 countries, and 36 Category 3 countries, but no Category 4 countries. The fact that our analysis showed that level of development and political

[10] Explaining Child Malnutrition in Developing Countries, http://ebrary.ifpri.org/utils/getfile/collection/p15738coll2/id/48054/filename/43512.pdf, last accessed June 4, 2016

[11] Reducing Malnutrition in Developing Countries: Increasing Rice Production in South and Southeast Asia, http://trilateral.org/file/16, last accessed June 6, 2016.

[12] Food Security and Political Stability in the Asia-Pacific Region, http://apcss.org/Publications/Report_Food_Security_98.html, last accessed June 5, 2016.

stability did not affect rates of undernourishment, while other studies done with larger data sets claim otherwise suggests that there is a considerable difference in development and political stability, even between the Category 3 and Category 4 countries as categorized by the UN. This is a little surprising, considering the fact that they are still both considered "highly developed".

Since we removed a large portion of our initial data set, we recognize that we have a potential informative missingness issue. However, our original data set was sparse, and we were unable to find a similar data set containing all the information excluded in the one we acquired from FAO. This shows how difficult it is to collect a great amount of organized international data at frequent periods. The list of countries with incomplete data included countries with the highest GDPs (U.S., China, and Japan), small countries, such as Tuvalu with a population half the size of Northfield, Minnesota, and extremely poor countries, such as Burundi, which was ranked as the least happiest country in the world [13]. We suspect that some of the data was missing not because of a lack of ability to collect them, but because they were considered unnecessary due to common assumptions. We presume this is exactly why none of the most developed countries had any information related to the rate of undernourishment or percentage of paved roads. We naturally assume those numbers to be close to 0% and 100%, respectively. By excluding countries on both extremes, the severely underdeveloped and the exceptionally developed, we created a model that was different from many others, but focused more on countries in the middle range.

For future studies, we recommend that statisticians use a data set containing information on Category 4 countries as well, since we see a large difference in the effect of development and political stability between them and the rest of the world. It is safe to assert that our results would have been significantly different had we used a denser data set on all countries. We also suggest that a more detailed study be performed on food distribution, to enable a deeper comparison between how food production and food distribution affect countries of varying levels of development. Since average food production was significant in our analysis, and food development is a more crucial issue in many countries, we suspect there may be a category-specific pattern in the effect of these two variables on the rate of undernourishment.

[13] The Distribution of World Happiness, http://worldhappiness.report/wp-content/uploads/sites/2/2016/03/HR-V1Ch2_web.pdf, last accessed June 4, 2016.

5 Appendix

Figure A1: Scatter plot of response against political stability, grouped by level of development

Figure A2: Trellis plot of response against food production, grouped by quantiles of access to water

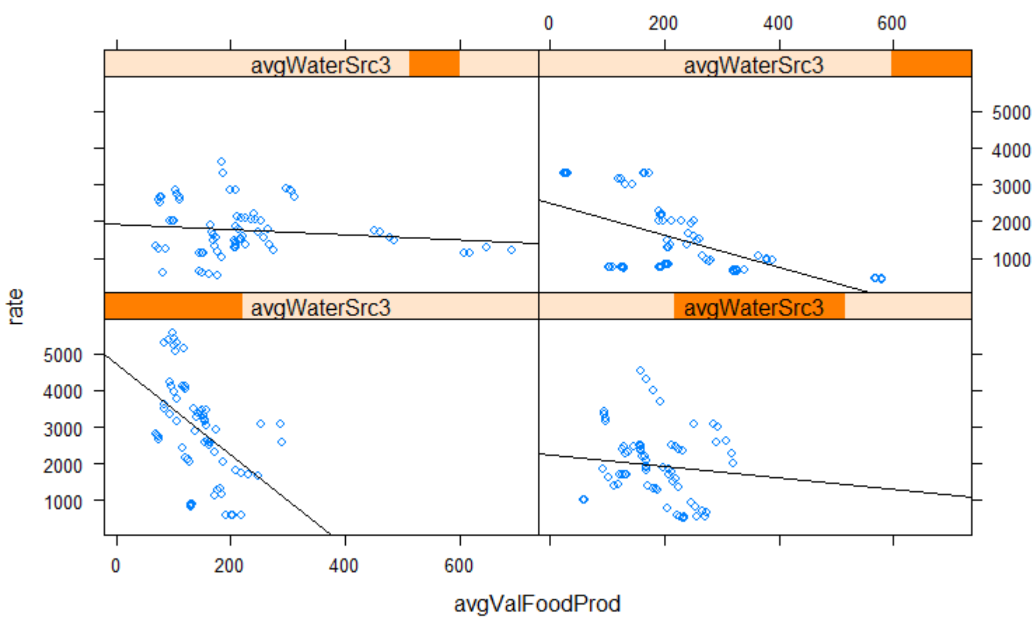


Figure A3: Initial QQ-plot of random effects for the final model

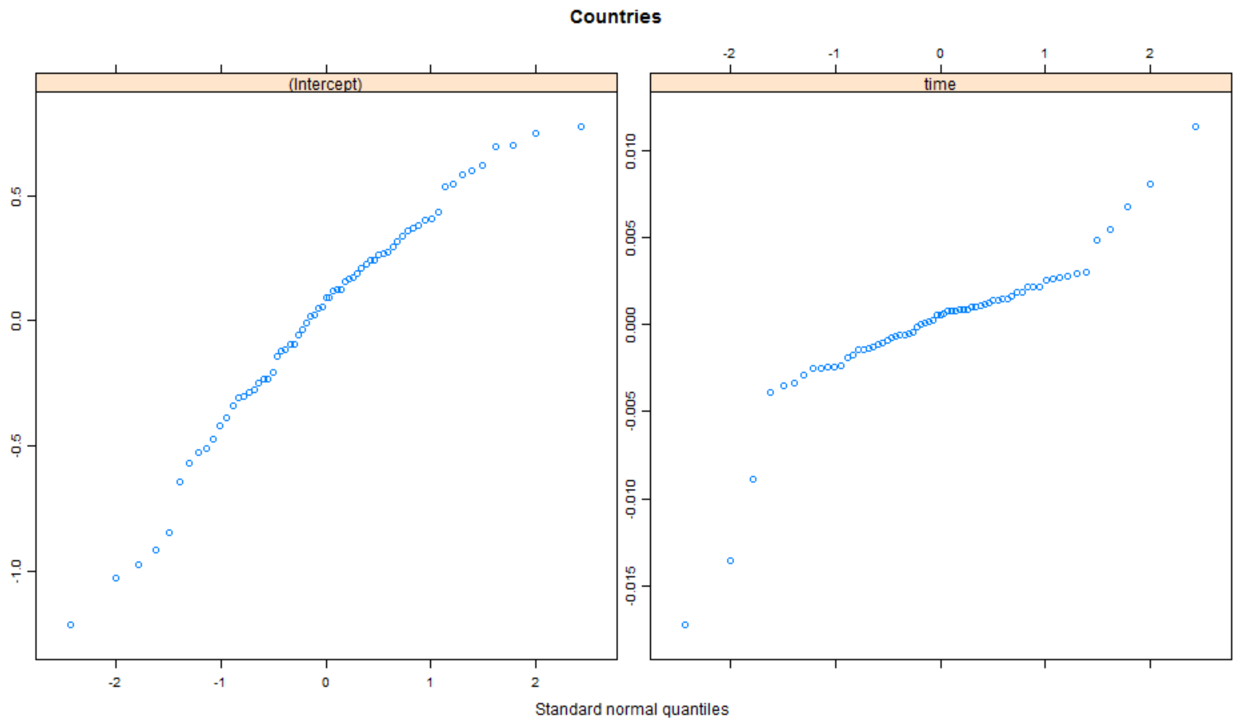


Figure A4: QQ-plot for the final model after removing outliers

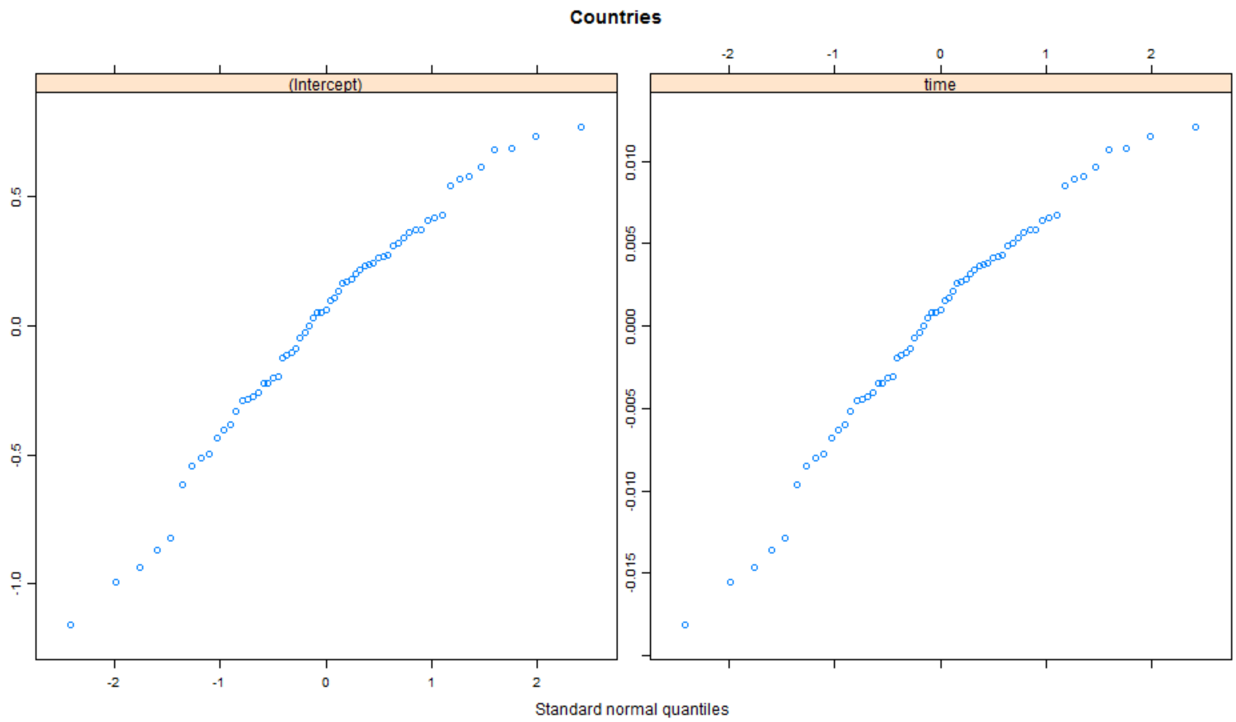


Figure A5: Plot of raw residuals for the final model