Technische Universität München

Department of Mathematics

Modelling Globalization Indices with Linear Regression and Copulas

Bachelor's Thesis

by

Alexander Kreuzer

Supervisor: Prof. Claudia Czado, Ph.D.
Advisor: Prof. Claudia Czado, Ph.D.; Daniel Kraus, M.Sc.
Submission Date: 21.08.2014

I hereby declare that this thesis is my own work and that no other sources have been used except those clearly indicated and referenced.

Fernhag, 21.08.2014

Zusammenfassung

Jedes Jahr wird für 207 Länder der KOF Index of Globalization veröffentlicht, welcher sich in den ökonomischen, den sozialen und den politischen Globalisierungsindex gliedern lässt. Diese drei Subindezes werden über die Zeit gemittelt und dienen als abhängige Variablen für drei lineare Regressionsmodelle. Zunächst werden diese Indizes und die verfügbaren unabhängigen Variablen einzeln analysiert. Als unabhängige Variablen stehen der Längen- und Breitengrad, der Kontinent, sowie die Geburtenrate eines Landes zur Verfügung. Anschließend werden Abhängigkeiten unter den unabhängigen Variablen, sowie der Einfluss einer unabhängigen Variable auf die abhängige Variable analysiert. Als letzter Schritt bevor das Modell aufgestellt wird, werden paarweise Interaktionen untersucht. Mit Hilfe der vorangegangenen Analysen werden dann die Regressionsmodelle aufgestellt, deren Koeffizienten interpretiert, sowie die Richtigkeit der bei der Regression getroffenen Annahmen überprüft. Als Grundlage für die Modellierung der Abhängigkeiten zwischen den Fehlern in den Regressionsmodellen wird anschließend auf die Grundlagen von Copulas eingegangen. Danach wird diese Theorie auf unsere Daten angewandt und die gemeinsame Vorhersage von zwei Indizes mit und ohne Copula durchgeführt.

Contents

1. Introduction

In today's world globalization as a process of international integration is ever-present. Almost everybody is contributing to a globalized world, e.g. when communicating with friends in other countries via Facebook or Skype. Big companies have many international connections and foreign branches. These are only two of many effects that drive globalization. The KOF index of globalization takes these effects into account and uses them to build an index, measuring the globalization of a country. It was introduced in 2002 and is calculated and published once a year for 207 countries. When the index is calculated, three dimensions of globalization are covered, which are economic, social and political globalization. Economic globalization is characterized as long distance flows of goods, capital and services as well as information and perceptions that accompany market exchanges. Political globalization is characterized by a diffusion of government policies and social globalization expresses the spread of ideas, information, images and people (see http://globalization.kof.ethz.ch, method of calculation). Of course the three sub indices economic, social and political globalization are not only dependent on the variables that appear in the formula for calculating them (see section 2.1). In this thesis we will have a look at such variables that do not appear in the formulas for the indices but might have an effect on the index. After having a detailed look at the available data, we use linear regression to assess if there are relationships between one of the sub indices of a country and the geographic location, the continent to which the country belongs or the birth rate of the country. E.g. one would expect that Africa has a lower economic globalization index than Europe. After fitting the linear model for each of the three sub indices, we will consider the errors made by the regression and look for dependencies among the errors of the three models. Therefore we make use of copulas. Copulas are a very valuable tool for modelling dependencies and have therefore many applications, e.g. in finance. After a brief introduction to copulas and the theoretical background, we will apply the theory to our data set.

2. Data description

2.1. Description of the response variable

As already mentioned in the introduction, the KOF index of globalization (http://globalization.kof.ethz.ch/) was introduced in 2002. The index is calculated once a year and is available from 1970 to 2011 for 207 countries. It can be separated in an economic, social and political globalization index. Due to missing data points the data set used here contains 151 countries and not every index is available from 1970 to 2011. The missing years are analyzed in more detail in section 2.2.3.
To get a first impression, the three indices of Germany are plotted in Figure 2.1.

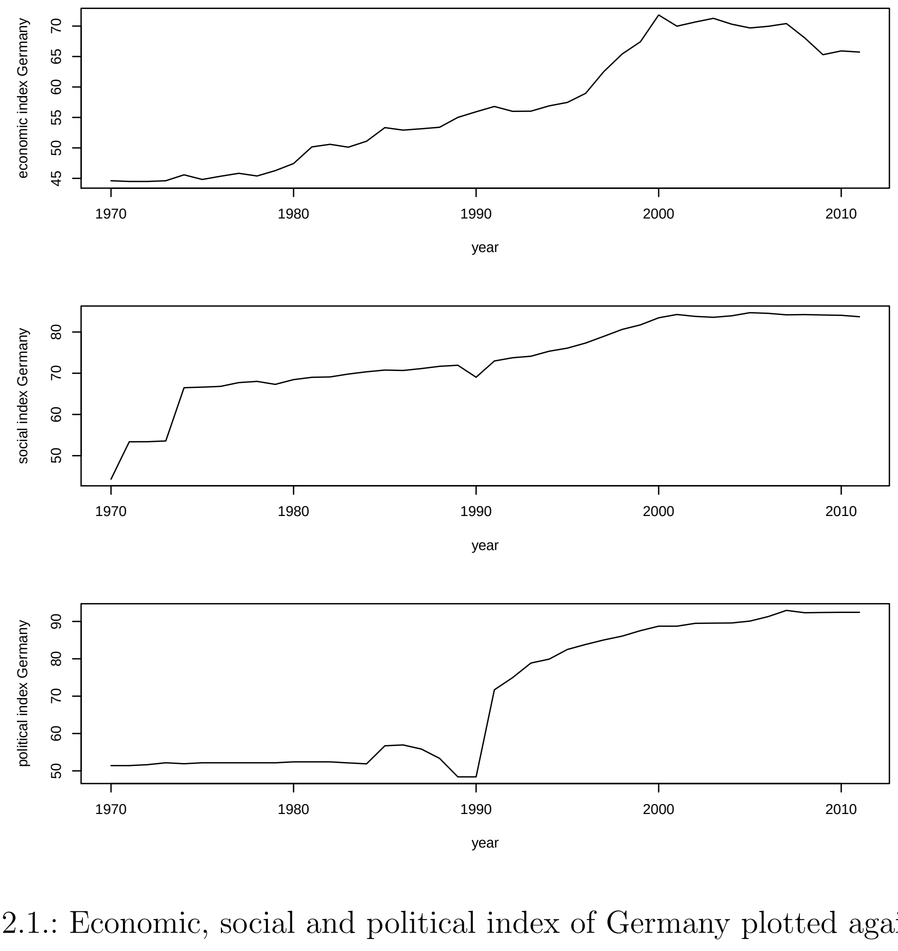

Figure 2.1.: Economic, social and political index of Germany plotted against time

The following contains detailed information about how the indices are obtained:

Globalization = 0.36·Economic Globalization + 0.38·Social Globalization + 0.26·Political Globalization

Economic Globalization = 0.5·Actual Flows + 0.5·Restrictions

Actual Flows = 0.21·Trade (percent of GDP) + 0.27·Foreign Direct Investment, stocks (percent of GDP) + 0.24·Portfolio Investment (percent of GDP)

Restrictions = 0.24·Hidden Import Barriers + 0.28·Mean Tariff Rate + 0.26·Taxes on International Trade (percent of current revenue) + 0.22·Capital Account Restrictions

Social Globalization = 0.33·Data on Personal Contact + 0.35·Data on Information Flows + 0.32·Data on Cultural Proximity

Data on Personal Contact = 0.25·Telephone Traffic + 0.04·Transfers (percent of GDP) + 0.26·International Tourism + 0.21·Foreign Population (percent of total population) + 0.24·International letters (per capita)

Data on Information Flows = 0.36·Internet Users (per 1000 people) + 0.37·Television (per 1000 people) + 0.27·Trade in Newspapers (percent of GDP)

Data on Cultural Proximity = 0.45·Number of McDonald's Restaurants (per capita) + 0.45·Number of Ikea (per capita) + 0.10·Trade in books (percent of GDP)

Political Globalization = 0.25·Embassies in Country + 0.28·Membership in International Organizations + 0.22·Participation in U.N. Security Council Missions + 0.25·International Treaties

Corresponding to the indices we introduce the quantitative variables $Econ_{jt}$, $Social_{jt}$ and $Polit_{jt}$ which describe the economic, social and political index in country j at time t. Furthermore, for all time dependent variables X_{jt} in this and the following sections we define

$$X_j = (X_{j(T_j)_1}, ..., X_{j2011}) \quad \text{and}$$
$$X = (X_{1(T_1)_1}, ..., X_{1,2011}, X_{2(T_2)_1}, ..., X_{2,2011}, ..., X_{151(T_{151})_1}, ..., X_{151,2011})$$

where the comma is just used to clarify the distinction between the values of j and t. $(T_j)_1$ is the earliest year with available data for country j. The variable T_j is defined in

section 2.2.3. For a time independent variable Y_j (j indicates the country) we define Y so that it corresponds with X, i.e. if for the i-th entry of X holds j = k, then the i-th entry of Y is Y_k.
The indices are defined so that they are between 0 and 100. Since the differences between mean and median are small, the indices seem to have a symmetric distribution.

	min	max	mean	median
Econ	9.24	99.16	50.08	49.12
Social	2.43	93.68	40.72	37.05
Polit	3.99	98.43	56.72	55.39

Table 2.1.: Summary statistics for the three response variables

2.2. Description of the available covariates

2.2.1. Longitude

For all countries in the data set the longitude at the center point is obtained from the CIA Factbook (https://www.cia.gov/library/publications/the-world-factbook/fields/2011.html). The quantitative variable Lon_j is the longitude of country j. There is no t in the index as longitude does not depend on the time.

	min	max	mean	median
Lon	-99.10	178.30	15.63	18.03

Table 2.2.: Summary statistics for *Lon*

A negative value of Lon_j indicates the amount of degrees in western direction and a positive value the amount of degrees in eastern direction. Longitude needs a special treatment since it is a circular variable (e.g. countries at 160 degrees east and 170 degrees west are near to each other). With the following function the variable Lon_j is transformed to the two-dimensional random vector $Lon_j.c = (Lon_j.c1, Lon_j.c2)$ that takes values in the unit circle, according to the notion of longitude on the globe.

$$Lon_j.c = (Lon_j.c1, Lon_j.c2) = (cos(Lon_j \cdot \frac{2\pi}{360}), sin(Lon_j \cdot \frac{2\pi}{360}))$$

When referring to the circle (Figure 2.2) 0 degrees corresponds with the x-y-coordinate (1,0) and the direction is counterclockwise. One can see that the density is very high between 0 and 20 degrees, this is where parts of Europe and Africa are located. No dots are between 180 and 270 degrees due to the pacific ocean.

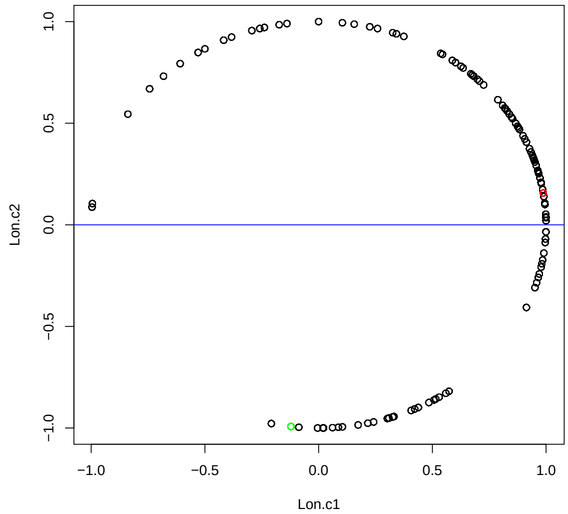

Figure 2.2.: $Lon.c1$ plotted against $Lon.c2$. The blue line is the zero meridian. Germany (9 degrees East) is marked with a red dot and USA (97 degrees West) with a green dot.

2.2.2. Latitude

Like longitude, the latitude of a country at its center point is obtained from the CIA Factbook. The quantitative variable Lat_j is the latitude of a country j. Like Lon_j, Lat_j is time independent.

	min	max	mean	median
Lat	-41.00	65.00	19.27	18.15

Table 2.3.: Summary statistics for *Lat*

A negative value of Lat_j indicates the amount of degrees in southern direction, a positive value the amount of degrees in northern direction.

2.2.3. Year

The quantitative variable year in country j at year t is denoted by T_{jt}. Of course it holds $T_{jt} = t$. The variable is defined this way to be consistent with the other variable definitions. Due to missing data points, the entries of the vector T_j do not range from 1970 to 2011 for every country, i.e. the length of the vector T_j depends on the country j. If data of a country is available at time t_1, the data is available for all years from t_1 to 2011 for this country. This is visualized in Figure 2.3, where we can see that for most countries data is available for the whole time period, some countries' earliest available data point is around 1990, and the country with the least data points available, Sri Lanka, ranges from 1998 to 2011.

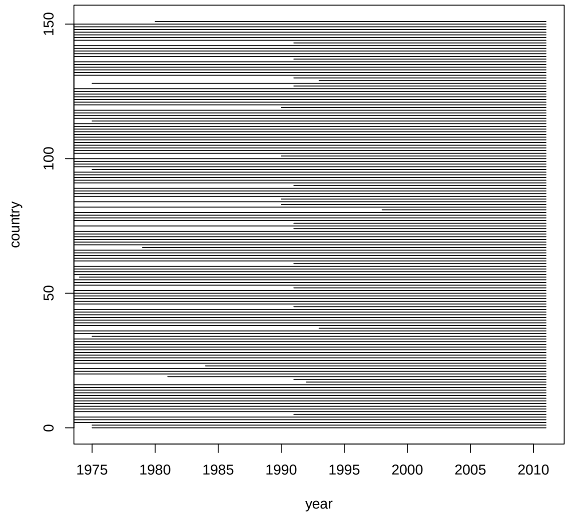

Figure 2.3.: Each integer on the y-axis corresponds with one country. For each country the line starts in the year with the earliest available data point.

2.2.4. Birth rate

The quantitative variable $Birth_{jt}$ describes the total number of births per 1000 people in country j in year t. The birth rate data is obtained from the World bank (http://data.worldbank.org/indicator/SP.DYN.CBRT.IN/countries). Three missing data points were linearly interpolated.

	min	max	mean	median
Birth	7.60	56.32	28.03	26.75

Table 2.4.: Summary statistics for *Birth*

2.2.5. Continent

In contrast to the other variables, continent is a factor. It is denoted by $Cont_j$ (j indicates the country) and is of course time independent. As there are six continents in the data set, $Cont_j$ is a factor with 6 levels. Table 2.5 shows how many data points are contained in one continent and how the 151 countries distribute over the continents.

Continent	Africa	Asia	Europe	North America	Oceania	South America
No. of data points	1757	1352	1320	658	205	541
No. of countries	43	37	37	16	5	13

Table 2.5.: Summary statistics for *Cont*

2.3. Summary of all variables

Most variables are time independent. Furthermore, we will see in the following section that the time has influence but the influence does not seem to be very big. This is why the model will later be build for the time averaged index. Therefore we define for a time dependent variable X_{jt} ($X_{jt} = Econ_{jt}, Social_{jt}, Polit_{jt}, Birth_{jt}$)

$$X_j.avg = \frac{1}{2011-(T_j)_1+1} \sum_{t=(T_j)_1}^{2011} X_{jt}$$

where $(T_j)_1$ is the first entry of the vector T_j, i.e. the earliest year of available data for country j. $X.avg$ is the vector containing $X_j.avg$ (j=1, ..., 151). For a time independent variable Y_j ($Y_j = Lon_j.c1, Lon_j.c2, Lat_j, Cont_j$) we can just define $Y.avg$ as the random vector with entries Y_{j2011} (j=1, ..., 151) and of course $Lon.c.avg = (Lon.c1.avg, Lon.c2.avg)$.

The following table gives an overview over all available variables:

	Variable	Time averaged variable	Type
Responses	$Econ_{jt}$	$Econ_j.avg$	time dependent, quantitative
	$Social_{jt}$	$Social_j.avg$	time dependent, quantitative
	$Polit_{jt}$	$Polit_j.avg$	time dependent, quantitative
Available Covariates	Lon_j	na	time independent, quantitative, circular
	Lat_j	na	time independent, quantitative
	T_{jt}	na	time dependent, quantitative
	$Birth_{jt}$	$Birth_j.avg$	time dependent, quantitative
	$Cont_j$	na	time independent, categorical

Table 2.6.: Summary of all variables

3. Exploratory data analysis

3.1. Assessing dependencies among the covariates

Because high collinearity among the covariates leads to variance inflation of the estimated regression coefficients, we take a look at the relations between the covariates.

Lon.c *Lat* *Birth*

Lon.c *Lat* *Birth*

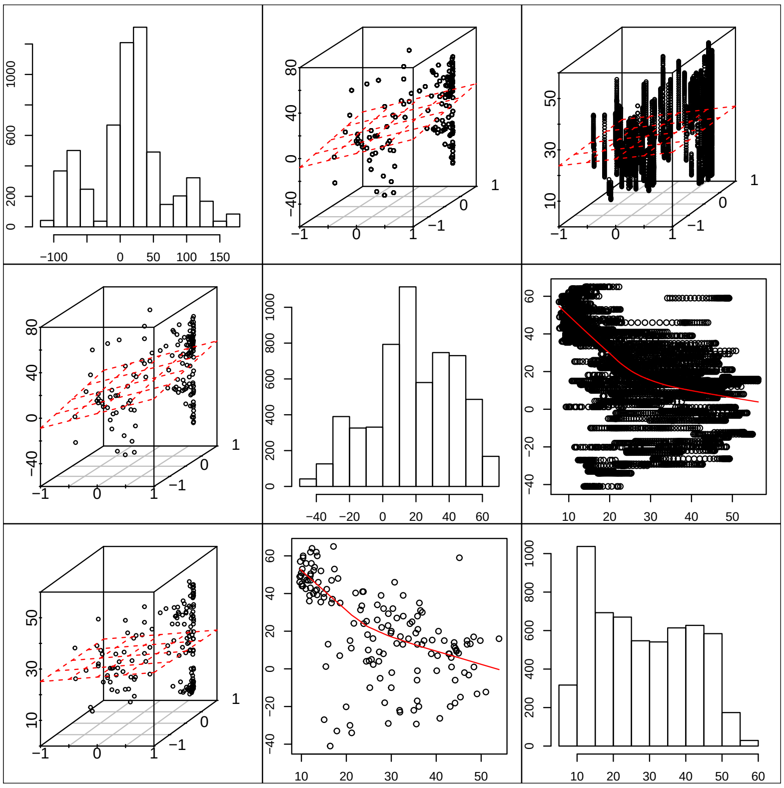

Figure 3.1.: Above the diagonal one can see scatter plots of the variable on the left (x-axis or x-y-plane) against the variable on the top (y-axis or z-axis) for the original index. On the diagonal histograms are plotted, whereas the first histogram is not the one for *Lon.c* but for *Lon*. Below the diagonal one can see scatter plots of the variable on the top (x-axis or x-y-plane) against the variable on the left (y-axis or z-axis) for the averaged data as defined in section 2.3. The corresponding regression plane is added to the three-dimensional scatter plots and the line of a locally weighted polynomial regression is added to the two-dimensional scatter plots.

First of all we can conclude from Figure 3.1 that there is not much difference observable between the original and the averaged data. Furthermore there are several relations observable among the covariates:

- There are small dependencies between $Lon_j.c$ and Lat_j. Where Europe and Africa are located (first quarter of the circle) are high latitude values, where South and North America are located (third quarter of the circle) lower values.
- There is not a significant dependency observable in the plot of $Lon.c$ against $Birth$.
- There seems to be a linear relationship between Lat_j and $Birth_{jt}$. Countries with higher latitude values have on average lower birth rates.

From Figure 3.2 we can see:

- Of course $Cont_j$ has an effect on Lat_j due to the continents' location on the globe. One can see that Europe has the highest latitude values and Oceania the lowest. This plot also corresponds with the $Lon.c$ against Lat plot in Figure 3.1 where we have seen that the average latitude of Europe and Africa is higher than the average latitude of North and South America.
- The average birth rate values per continent are about the same for $Birth_{jt}$ and $Birth_j.avg$. The ranges are smaller for $Birth_j.avg$ which is caused by taking the mean. The continent does definitely have an effect on the birth rate since Africa's birth rate is much higher than Europe's birth rate while the other continents seem to have about the same birth rate. The boxplot corresponds with the plot of latitude and birth rate in Figure 3.1 as the birth rate of Europe, North America and Asia which contain countries with higher latitude values is on average less than the average birth rate of the other continents.

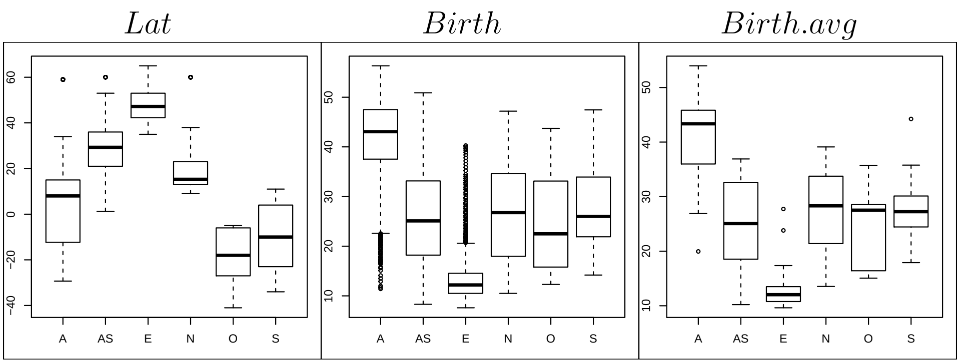

Figure 3.2.: Box plots per continent (A=Africa, AS=Asia, E=Europe, N=North America, O=Oceania, S=South America)

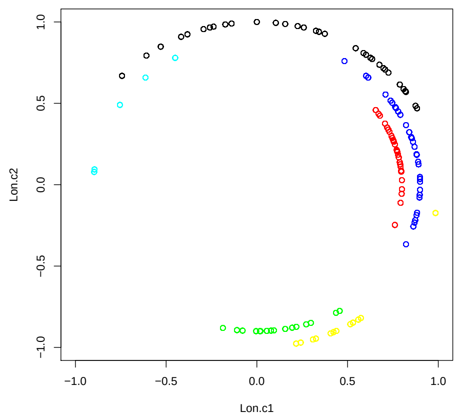

Figure 3.3.: *Lon.c* is plotted on the two-dimensional Cartesian coordinates system. Continents are marked in different colors (Africa=blue, Asia=black,Europe=red, North America=green, Oceania=light blue, South America=yellow).

Figure 3.3 visualizes which longitude values belong to which continent. This is useful for further analysis, e.g. now we have more information why longitude does not have a big effect on the birth rate what we saw in Figure 3.1. Africa and Europe have the same longitude values. From Figure 3.2 we can see that Europe has a low and Africa a high birth rate, so the average birth rate at this longitude is of medium size, comparable to the birth rates at other continents.

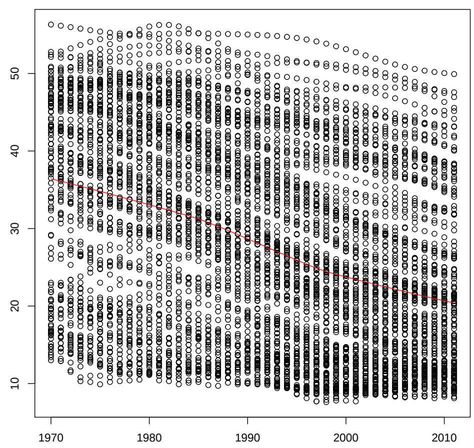

Figure 3.4.: *Birth* plotted against T

In Figure 3.4 we can see that the time does have an influence on the birth rate as there is a decreasing tendency but since the data points spread widely around the red line, the relationship is not strong.

3.2. Influence of a single covariate on the response

In the following the effects of the covariates on the response variables are analyzed to get an idea of how the covariates could appear in the linear model.

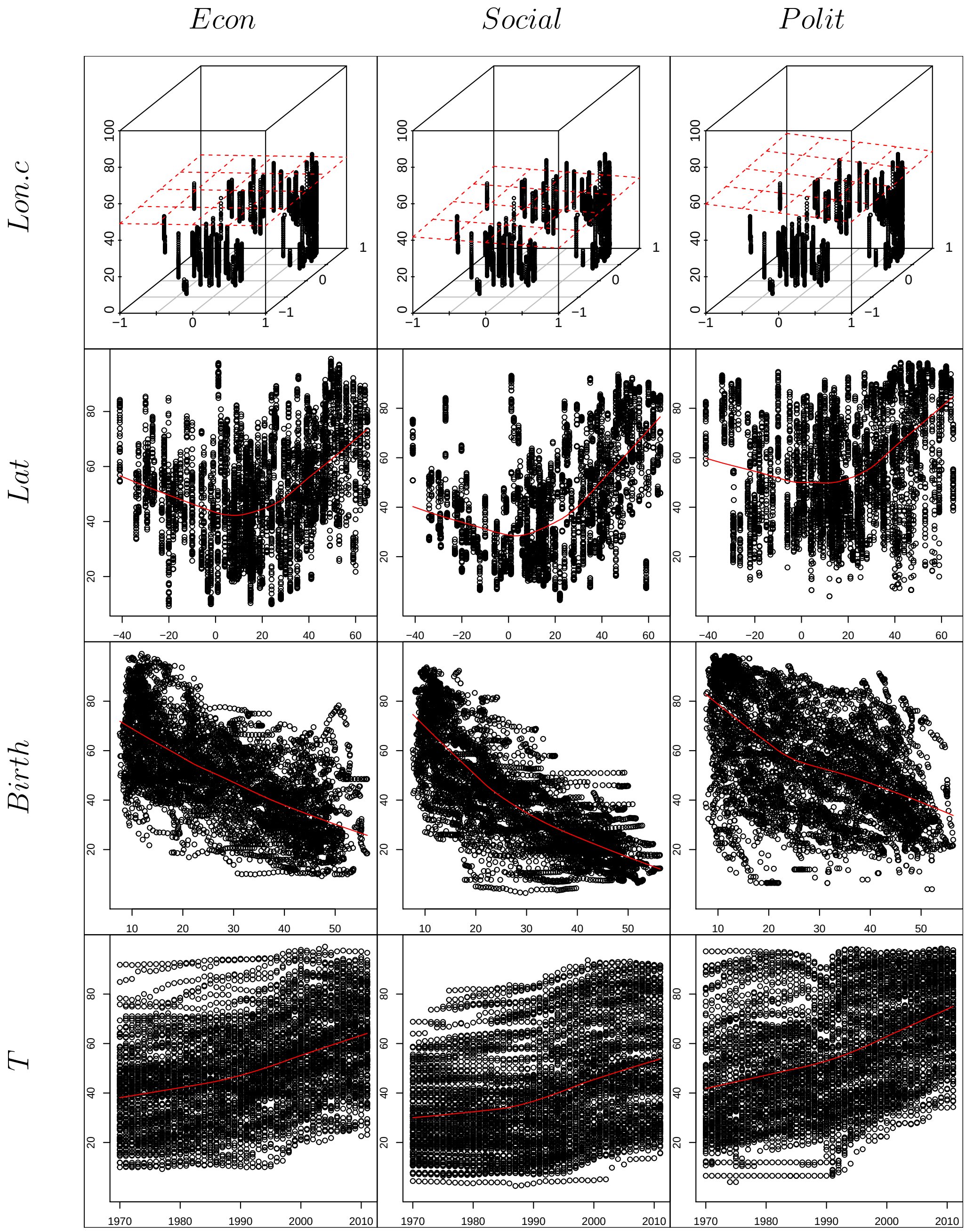

Figure 3.5.: Covariate (x-axis or x-y-plane) plotted against response (y-axis or z-axis) for the original index

Comparing Figure 3.5 with Figure 3.6 one can see that the effect of the covariate does only slightly change. We observe the following dependencies:

- $Lon_j.c$ does not have a big effect on the responses.
- All three responses could be described by Lat_j through a polynomial of degree two, while this seems to work best for the social index.
- The relation between T_{jt} and the responses seems not to be very strong, while there is an increasing tendency.
- We can see a linear relationship between $Birth_{jt}$ and all three responses. The dependency between $Birth_{jt}$ and $Polit_{jt}$ seems not to be strong as the data points spread widely.

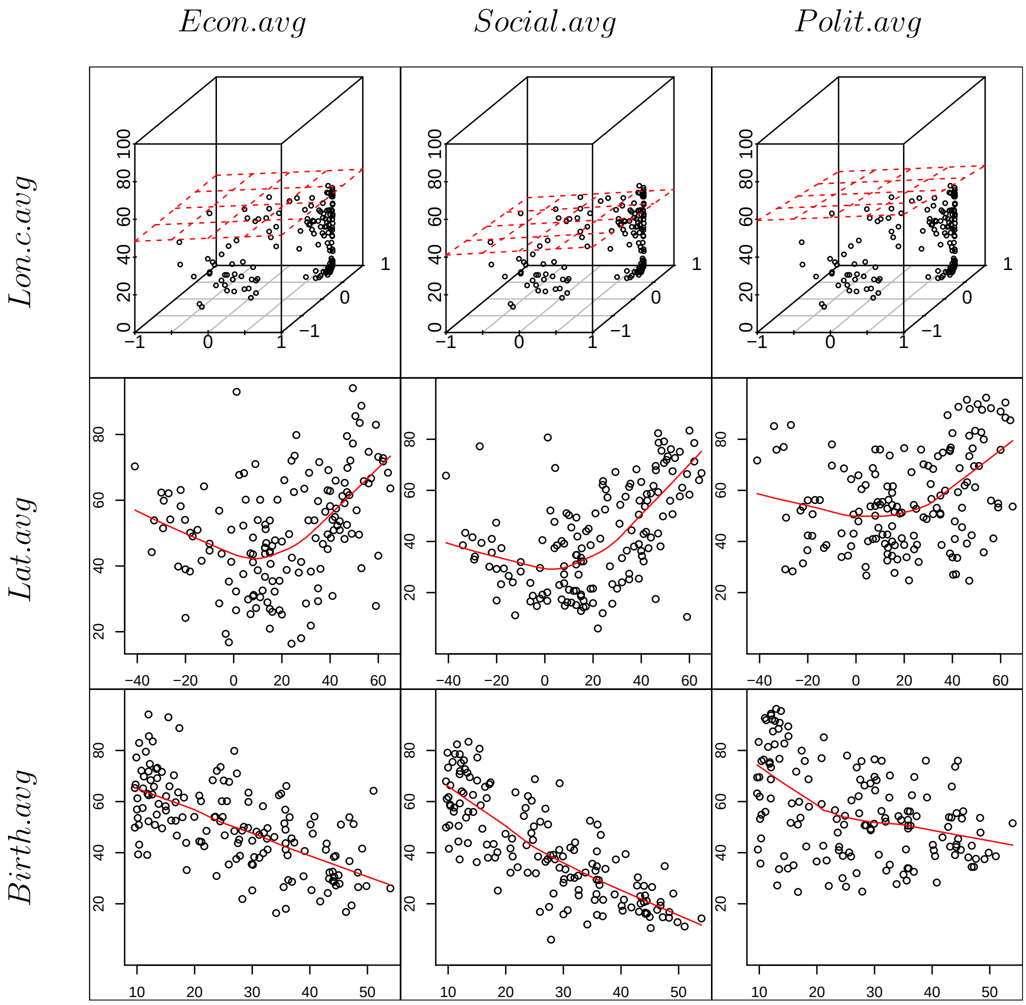

Figure 3.6.: Covariate (x-axis or x-y-plane) plotted against response (y-axis or z-axis) for the averaged data

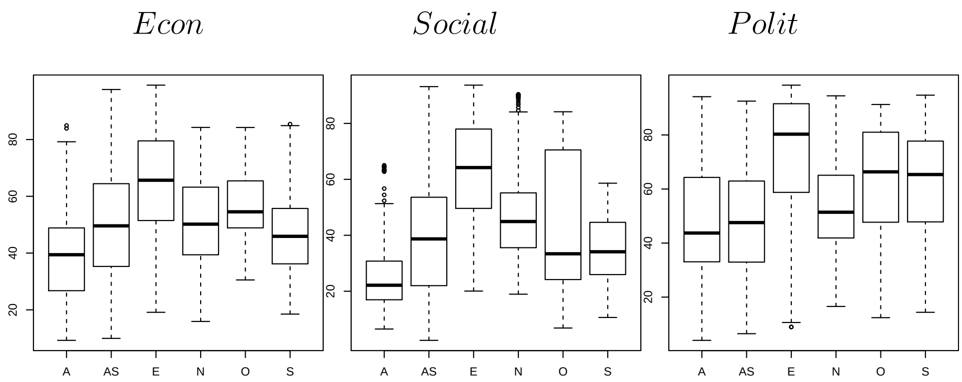

Figure 3.7.: Box plots per continent for the original index (A=Africa, AS=Asia, E=Europe, N=North America, O=Oceania, S=South America)

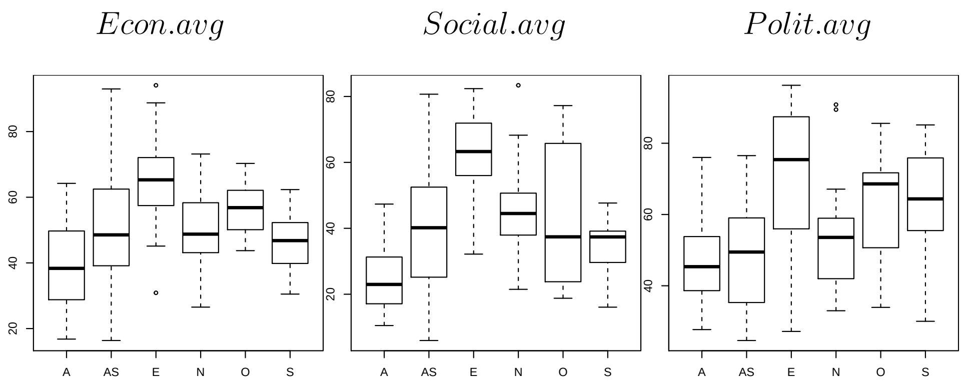

Figure 3.8.: Box plots per continent for the averaged index (A=Africa, AS=Asia, E=Europe, N=North America, O=Oceania, S=South America)

Again we observe that there are no significant differences between the plots for the original index (Figure 3.7) and the plots for the averaged index (Figure 3.8). Furthermore we can see that $Cont_j$ does definitely have an effect on all three response variables. Europe has the highest and Africa the lowest mean index (holds for all three indices). From these plots we get more information why longitude has no effect on the responses. The biggest difference is between the indices in Europe and Africa but Europe and Africa have the same longitude.

3.3. Effects of pairs of variables on the responses

As we build the regression model for the time averaged index, we only have a look at the interaction plots for the time averaged index. Interaction plots are, as well as the plots we have seen before, quite similar for the original and the time averaged data.

Remark: We will from now on often leave out the ".avg" that was part of the name of the time averaged variable to make notation easier. The original variables will no longer play a role.
When data is divided into three parts by high, medium and low values of a variable, high values mean values above the 75% quantile, low values below the 25% quantile and medium values the rest.
We can see from Figure 3.9 that there are only small interaction effects, the biggest one for $Lon.c(Lat)$ in the model for the economic index.

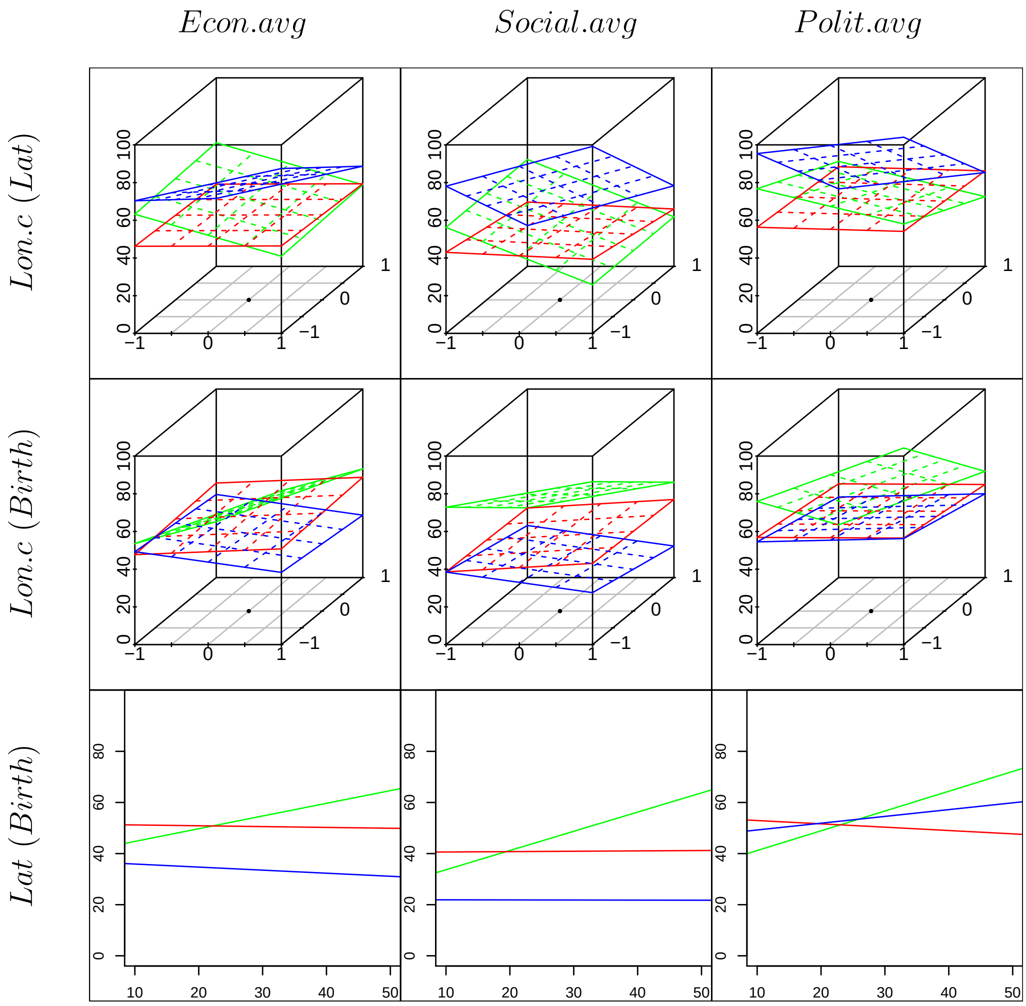

Figure 3.9.: Interaction plots. $Var1(Var2)$ means that the data set is divided into three groups: high, medium and low values of $Var2$. Then for each of the three groups the regression line/plane of the model with corresponding response variable (index on top) and covariate $Var1$ is plotted.

Having a look at Figure 3.10, there seem to be interaction effects for $Cont$ and $Lon.c$ as well as for $Cont$ and Lat. For more information we look at the mean size of a 95% prediction interval (Table 3.1). As the intervals are of appropriate size, except for the group Oceania, we will consider these interaction effects later in the linear model.

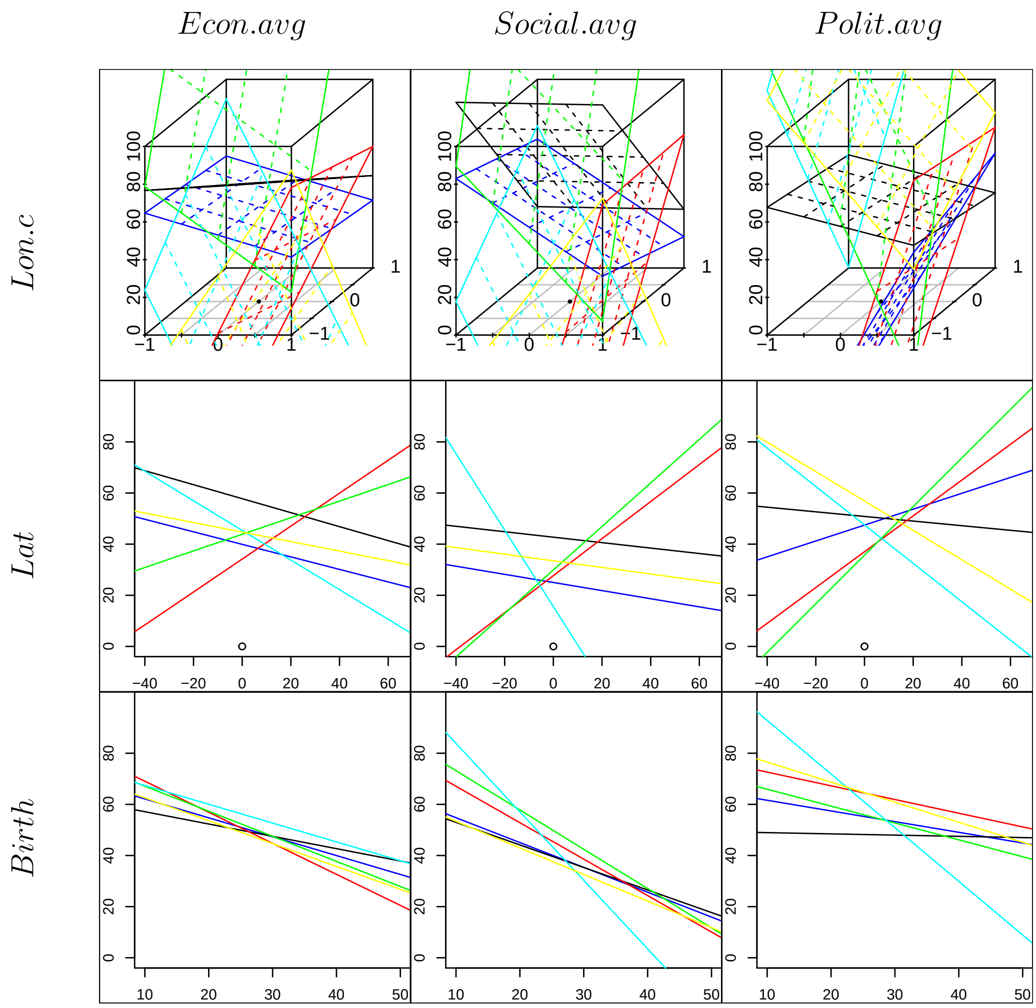

Figure 3.10.: Interaction plots. The data set is divided in six parts according to the continents: Africa=blue, Asia=black, Europe=red, North America=green, Oceania=light blue, South America=yellow. For each part the regression line/plane is drawn for the model with the corresponding response (on the top) and covariate (on the left).

Furthermore, there seems to be an interaction between *Cont* and *Birth* for the political index when we look at the plot. It seems to be important if a country is in Oceania or not. But the mean size of the 95% prediction interval of *Polit* described by *Birth* in this group is 36.94. So we consider the interaction as not significant.

	$Econ \sim Lon.c$	$Social \sim Lon.c$	$Polit \sim Lon.c$	$Econ \sim Lat$	$Social \sim Lat$	$Polit \sim Lat$
A	12.71	9.19	13.32	9.69	7.24	10.52
AS	19.69	19.36	16.28	15.76	16.78	13.68
E	13.34	11.63	19.47	11.11	10.02	16.54
N	22.43	26.63	21.69	16.51	14.50	16.69
O	76.20	230.76	164.16	24.78	56.60	75.11
S	15.90	11.90	32.59	15.68	14.12	22.35
avg	26.71	51.58	44.58	15.59	19.88	25.82

Table 3.1.: Mean size of 95% prediction interval. The first row contains the model, where $Var1 \sim Var2$ means that $Var1$ is the response and $Var2$ the covariate. The first column gives the corresponding group (A=Africa, AS=Asia, E=Europe, N=North America, O=Oceania, S=South America). The last row contains the average over all continents.

4. Regression model building

Here we will give a brief introduction to linear regression based on Guangming (2013). Two important distributions needed are the normal and Student's t distribution.

The density function of the standard normal distribution (zero mean and standard deviation 1) is given by

$$\varphi(x) = \frac{e^{-\frac{1}{2}x^2}}{\sqrt{2\pi}} \quad \forall x \in \mathbb{R}.$$

The more general normal distribution with mean $\mu \in \mathbb{R}$ and standard deviation $\sigma > 0$ has the density function

$$\varphi_{\mu,\sigma}(x) = \frac{1}{\sigma}\varphi(\frac{x-\mu}{\sigma}) \quad \forall x \in \mathbb{R}.$$

The density function of a t distribution with ν degrees of freedom is

$$f_\nu(x) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\Gamma(\frac{\nu}{2})}\left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}} \quad \forall x \in \mathbb{R}$$

where Γ is the gamma function which is for positive numbers given by $\Gamma(x) = \int_0^\infty y^{x-1}e^{-y}dy$.

In linear regression we assume a model M of the form

$$Y_i = \beta_0 + \beta_1 \cdot x_{i1} + \beta_2 \cdot x_{i2} + ... + \beta_p \cdot x_{ip} + \varepsilon_i \text{ (i = 1, ..., N)}$$

where Y_i is the response variable, $\beta_i \in \mathbb{R}$, $x_{i1}, ..., x_{ip}$ are realizations of the covariates $X_1, ..., X_p$, $\varepsilon_i \sim$ iid $N(0, \sigma^2)$ and N is the number of observations of the covariates (here N = 151). $\varepsilon_i \sim$ iid $N(0, \sigma^2)$ means that $\varepsilon_1, ..., \varepsilon_N$ are independent and identically normal distributed with zero mean and standard deviation σ.
In matrix form:

$$Y = X\beta + \epsilon$$

where $Y = (Y_1, ..., Y_N)$, $X \in \mathbb{R}^{N\times(p+1)}$ is the design matrix with i-th row $(1, x_{i1}, ..., x_{ip})$, $\varepsilon = (\varepsilon_1, ..., \varepsilon_N)$ and $\beta = (\beta_0, ..., \beta_p)$.

Furthermore we define the hat matrix

$$H = X(X^TX)^{-1}X^T \in \mathbb{R}^{N\times N}$$

β is estimated by least squares method, i.e. $\hat{\beta} = (X^T X)^{-1} X^T Y$. The estimate $\hat{Y}$ of Y is then given by $\hat{Y} = X\hat{\beta} = HY$.

The residuals are the differences between the estimated and the observed value, i.e. the residual vector is given by

$$Res = Y - \hat{Y}.$$

A very important value in regression analysis is σ^2 since it has e.g. influence on prediction intervals. σ^2 is estimated by

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{N} Res_i^2}{N-p-1}.$$

This value is also called the residual standard error of the model M.

Since

$$Var(\hat{\beta}) = (X^T X)^{-1} X^T \sigma^2 I_N X (X^T X)^{-1} = \sigma^2 (X^T X)^{-1}$$

(I_N is the N-dimensional unit matrix) the standard error of the estimate of the i-th coefficient is given by

$$s.e.(\beta_i) = \hat{\sigma}(X^T X)_{ii}^{-1}$$

where $(X^T X)_{ii}$ is the entry in the i-th row and i-th column of the matrix $X^T X$.

To check if the i-th covariate can be excluded from the model, we conduct a t-test. The null hypothesis is $\beta_i = 0$ and the alternative $\beta_i \neq 0$. The test statistic is given by

$$t^* = \frac{\hat{\beta}_i}{s.e.(\beta_i)}$$

which is t distributed with $N - p - 1$ degrees of freedom. We reject the null hypothesis at a significance level of α if $|t^*| > t_{N-p-1,1-\frac{\alpha}{2}}$ where $t_{N-p-1,1-\frac{\alpha}{2}}$ is the $1 - \frac{\alpha}{2}$ quantile of the t distribution with $N - p - 1$ degrees of freedom.

After estimating β, we will interpret the components β_i and check if the assumptions were appropriate. To check this, it is advantageous, to also consider the standardized residuals which are easier to compare since they have similar variance. To standardize the residuals we have a look at the covariance matrix of Res:

$$Cov(Res) = Cov(Y - \hat{Y}) = Cov((I_N - H)Y) = (I_N - H)Cov(Y)(I_N - H)^T = (I_N - H)I_N\sigma^2(I_N - H)^T$$

where I_N is the N-dimensional unit matrix. In the last step we used that the Y_i are independent and have the same variance σ^2 since the ε_i are independent $N(0, \sigma^2)$ distributed. We can see that the variance of the i-th residual Res_i is given by $(1 - h_{ii})\sigma^2$ where h_{ii} is the i-th diagonal element of the hat matrix. Now we can define the standardized residual as

$$std.Res_i = \frac{Res_i}{\sqrt{(1-h_{ii})\hat{\sigma}}}$$

$std.Res_i$ is approximately t distributed with $N - p - 1$ degrees of freedom (only approximately since Res_i and $\hat{\sigma}$ are not independent) and if N is big it is approximately standard normal distributed.

Further important values for a regression model are the R^2 and the adjusted R^2 of the model. The R^2 is defined as

$$R^2 = 1 - \frac{\sum_{i=1}^{N} Res_i^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2}$$

where $\bar{y} = \frac{1}{N}\sum_{i=1}^{N} y_i$. It can be shown that

$$R^2 = \frac{\sum_{i=1}^{151}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{151}(y_i - \bar{y})^2}$$

This term can easier be interpreted. In the numerator we have the variation around the mean explained by the regression equation and in the denominator we have the total variation, so the R^2 is the fraction of variation described by the regression model. By including covariates in the model the R^2 will mostly get higher (or stay the same). This is why we also have a look at the adjusted R^2 that adjusts the R^2 by penalizing higher numbers of covariates.

$$adjR^2 = 1 - \frac{\frac{\sum_{i=1}^{N} Res_i^2}{N-p-1}}{\frac{\sum_{i=1}^{N}(y_i - \bar{y})^2}{N-1}}$$

4.1. Economic index

In the following we describe how the regression model for the economic index is build and how good the fit is. Note that only time averaged variables are considered and that, like already mentioned, the ".avg" is left out.
According to section 3.2, we describe the index through latitude by a polynomial of degree two. For numerical reasons we use an orthogonal polynomial and denote the first basis polynomial evaluated at Lat_j by $Lat_j.p1$ and the second basis polynomial evaluated at Lat_j by $Lat_j.p2$. In addition we need the variable Af (respectively As, Eu, Na, Oc, Sa) whose value is 1 if the corresponding country is in Africa (respectively Asia, Europe, North America, Oceania, South America) and 0 otherwise. In a first approach we use a step wise regression algorithm. This algorithm starts with Model M0 that contains all covariates and pairwise interactions. Then it checks step by step if variables shall be included or excluded. Therefore it uses the Bayesian information criteria (BIC) which is defined as

$$BIC = N \cdot \log\left(\frac{\sum_{i=1}^{N} Res_i^2}{N}\right) + \log(N) \cdot (p+1)$$

for a model with residuals Res_i and p covariates which was fitted using N observations. A variable is excluded or included if this reduces the BIC value. This algorithm yields Model M1.econ shown in Table 4.1.

Model	Covariates	R^2	adj. R^2	Res. std. error
M0.econ	$Birth, Lon.c1, Lon.c2, Lat.p1, Lat.p2, Cont,$ $all\ pairwise\ interactions$	0.67	0.53	11.41
M1.econ	$Birth, Lon.c1, Lon.c2, Lat.p1, Lat.p2,$ $Lon.c1 \cdot Lat.p1, Lon.c1 \cdot Lat.p2,$ $Lon.c2 \cdot Lat.p1, Lon.c2 \cdot Lat.p2$	0.49	0.45	12.32

Table 4.1.: Summary of regression models for *Econ* using approach 1

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	70.46	3.69	19.08	0.00
Birth	-0.81	0.13	-6.29	0.00
Lon.c1	4.31	2.31	1.87	0.06
Lon.c2	-2.24	1.85	-1.21	0.23
Lat.p1	-37.41	24.59	-1.52	0.13
Lat.p2	7.62	18.62	0.41	0.68
Lon.c1:Lat.p1	59.94	28.31	2.12	0.04
Lon.c1:Lat.p2	29.40	22.46	1.31	0.19
Lon.c2:Lat.p1	-66.14	24.70	-2.68	0.01
Lon.c2:Lat.p2	-10.43	21.11	-0.49	0.62

Table 4.2.: Parameter estimates, standard errors, value of the t-statistics and p-value of the t-test for significance for model M1.econ

As the model contains some non significant variables, we try another approach (approach 2) with the intention of getting a model with more significant variables and higher R^2 and adj. R^2 values. This approach starts with model M2.econ that contains all covariates and pairwise interactions that seemed to be significant in section 3.3. Then step by step this model is reduced by non significant covariates which yields model M6.econ. This is shown in table 4.3. From one line to another, non significant covariates are removed from the model. Non significant means that the p-value of the t-test for significance is greater than 0.05.

Model	Covariates	R^2	adj. R^2	Res. std. error
M2.econ	$Birth, Lon.c1, Lon.c2, Lat.p1, Lat.p2, Cont,$ $Lon.c1 \cdot Cont, Lon.c2 \cdot Cont, Lat.p1 \cdot Cont,$ $Lat.p2 \cdot Cont$	0.59	0.49	11.90
M3.econ	$Birth, Lon.c1, Lon.c2, Lat.p1, Lat.p2, Cont,$ $Af, As, Eu, Na, Oc, As \cdot Lat.p1, As \cdot Lat.p2$	0.54	0.50	11.73
M4.econ	$Birth, Lon.c1, Lon.c2, Lat.p1, Lat.p2, As,$ $Eu, Na, Oc, As \cdot Lat.p1, As \cdot Lat.p2$	0.54	0.50	11.76
M5.econ	$Birth, Lon.c1, Lon.c2, Lat.p1, Lat.p2, As,$ $Na, Oc, As \cdot Lat.p1, As \cdot Lat.p2$	0.53	0.50	11.78
M6.econ	$Birth, Lon.c1, Lon.c2, Lat.p1, Lat.p2, As,$ $Oc, As \cdot Lat.p1, As \cdot Lat.p2$	0.52	0.49	11.89

Table 4.3.: Summary of regression models for *Econ* using approach 2

In table 4.4 we can see that all β_i's but two (β_4 and β_9) are significant at level 0.05. They are not excluded from the model due to interpretation reasons. β_4 is the coefficient of $Lat.p1$ and β_9 is the coefficient of $Lat.p2 \cdot As$.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	61.06	4.42	13.82	0.00
Birth	-0.70	0.12	-5.69	0.00
Lon.c1	11.27	3.41	3.30	0.00
Lon.c2	-6.35	2.51	-2.53	0.01
Lat.p1	17.42	18.52	0.94	0.35
Lat.p2	46.94	16.26	2.89	0.00
As	20.24	6.98	2.90	0.00
Oc	22.60	8.35	2.71	0.01
Lat.p1:As	-292.52	82.23	-3.56	0.00
Lat.p2:As	62.03	72.94	0.85	0.40

Table 4.4.: Parameter estimates, standard errors, value of the t-statistics and p-value of the t-test for significance for model M6.econ

Comparing M1.econ and M6.econ with respect to $R^2, adjR^2$ and residual std. error we choose M6.econ as model for $Econ_j$, i.e. we assume:

$$Econ_j = \beta_0 + \beta_1 \cdot birth_j + \beta_2 \cdot lon_j.c1 + \beta_3 \cdot lon_j.c2 + \beta_4 \cdot lat_j.p1 + \beta_5 \cdot lat_j.p2 + \beta_6 \cdot as + \beta_7 \cdot oc + \beta_8 \cdot as \cdot lat_j.p1 + \beta_9 \cdot as \cdot lat_j.p2 + \varepsilon_{econ,j}$$

where $\varepsilon_{econ,j} \sim$ iid $N(0, \sigma^2_{econ})$. (small letters: realization of the variable)

Analysis of the estimated coefficients (from Table 4.4):

- As expected (compare to Figure 3.6) there is a negative linear relation between $Econ_j$ and $Birth_j$.
- According to the fitted polynomial for Lat_j (Figure 4.1), countries with low latitude (Oceania and parts of South America) have on average a medium size index, countries with medium latitude values (Africa, parts of Asia and parts of South America) have on average a low index and countries with a high latitude (North America, Europe, parts of Asia) on average a high index. Furthermore for latitude values smaller than 10, the polynomial is decreasing. This corresponds with the interaction plot of latitude and continent (Figure 3.10) where we can see that for continents with latitude values smaller than 10 (Africa, parts of Asia, South America, Oceania) the regression line is decreasing. For latitude values bigger than 10, the polynomial is increasing. Comparing also this with the interaction plot, we can see that the regression line for Europe and North America is increasing but not the one for Asia. This is why Asia and the interaction of Asia and latitude are significant and added to the model. The corresponding coefficients are fitted so that by fixing all variables but latitude, the index is decreasing by increasing the latitude value for asian countries.

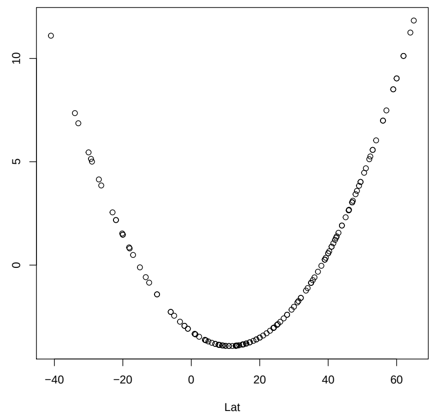

Figure 4.1.: Polynomial fitted in the regression model M6.econ for latitude

- With Figure 2.2 we can interpret β_2 and β_3. β_2 and β_3 tell us that by fixing all variables but *Lon.c*, the index is increasing by moving right or down in figure 2.2, and decreasing otherwise.
- Oceania has the same latitude as parts of South America and parts of Africa. There are only 5 countries in Oceania (see Table 2.5), so Oceania does not have much influence on estimating the coefficient. This is why the index for countries in Oceania is underestimated. This is fixed by the positive value of β_7.

- According to Figure 3.8 countries in Europe have on average the highest and countries in Africa on average the lowest index, so these continents definitely have an effect on the index. But in Figure 3.2 we can see that Europe has the lowest and Africa the highest birth rate, so the effect of these two continents is already described by the birth rate. This is the reason why As and Eu are not included in M6.econ.

The following plots help to check if the assumptions for the linear model M6.econ are appropriate.

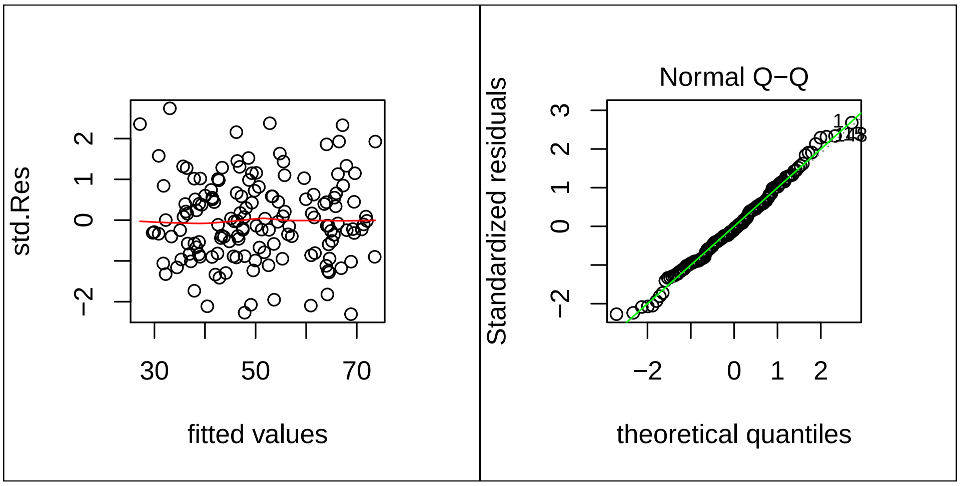

Figure 4.2.: Fitted vs. standardized residuals and QQ plot

As already mentioned, to check the standard deviation of the residuals, it is easier to have a look at the standardized residuals since they have by assumption all the same standard deviation. First of all we can see in the plot on the left of Figure 4.1 that the standardized residuals seem to have zero mean which corresponds with the assumption that the $\varepsilon_{econ,j}$'s have zero mean. Furthermore, the data points spread around the red line about the same amount for different fitted values which is evidence for the $\varepsilon_{econ,j}$'s having same variances.
The QQ plot checks if the standardized residuals are approximately $N(0,1)$ distributed by plotting the quantiles of the standard normal distribution against the quantiles of the empirical residual distribution. The data points are near to the line through origin what we expect from a good fit.

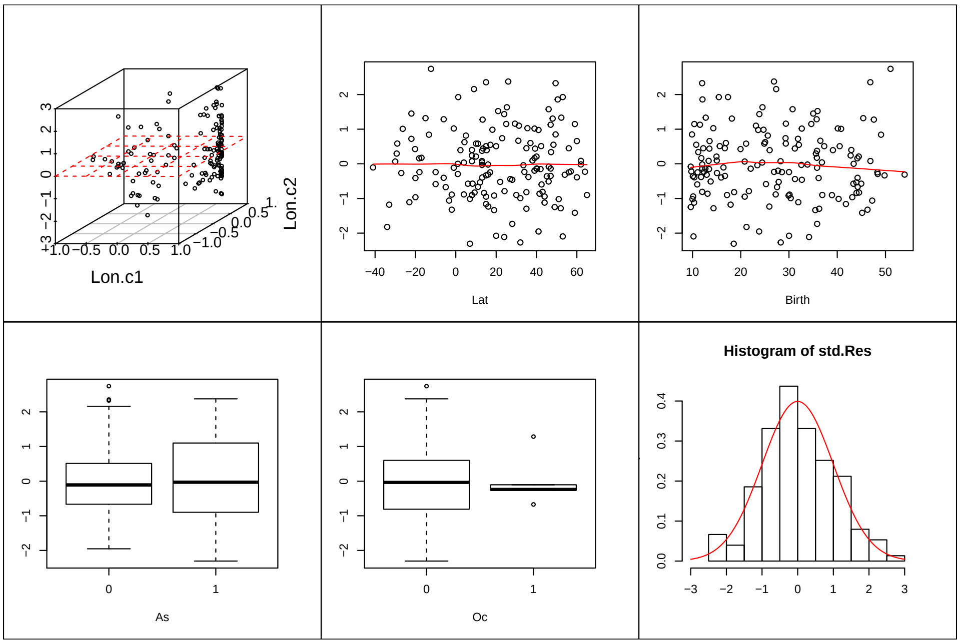

Figure 4.3.: Plots of covariate vs standardized residual and histogram (with density of the standard normal distribution added in red) of standardized residuals

As desired, we can conclude from Figure 4.3 that the variables do not have much influence on the residuals and the histogram indicates, like the QQ plot, that the standardized residuals are approximately standard normal distributed.

4.2. Social index

Here we use the same approach as in section 4.1 to build the model for the time averaged social index. So we first build regression model M1.social by applying the step wise regression algorithm that uses the BIC to decide if a variable is included or excluded. Model M1.social is shown in Table 4.5.

Model	Covariates	R^2	adj. R^2	Res. std. error
M0.social	$Birth, Lon.c1, Lon.c2, Lat.p1, Lat.p2, Cont, all\ pairwise\ interactions$	0.87	0.81	8.59
M1.social	$Birth, Lon.c1, Lon.c2, Lat.p1, Lat.p2, Cont, Birth * Lon.c1, Birth * Lon.c2, Lon.c1 \cdot Lat.p1, Lon.c1 \cdot Lat.p2, Lon.c2 \cdot Lat.p1, Lon.c2 \cdot Lat.p2, Lon.c2 \cdot Lat.p1, Lon.c2 \cdot Lat.p2, Lat.p1 \cdot Cont, Lat.p2 \cdot Cont$	0.82	0.78	9.29

Table 4.5.: Summary of regression models for *Social* using approach 1

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	57.46	7.98	7.20	0.00
Birth	-1.17	0.22	-5.29	0.00
Lon.c1	-2.57	9.02	-0.29	0.78
Lon.c2	-19.02	7.31	-2.60	0.01
Lat.p1	-41.39	79.22	-0.52	0.60
Lat.p2	126.15	68.83	1.83	0.07
Contasia	36.25	7.80	4.65	0.00
Conteurope	-12.81	33.53	-0.38	0.70
ContnorthAmerica	8.55	9.95	0.86	0.39
Contoceania	-87.77	75.00	-1.17	0.24
ContsouthAmerica	-3.57	22.84	-0.16	0.88
Birth:Lon.c1	0.41	0.30	1.39	0.17
Birth:Lon.c2	0.22	0.20	1.11	0.27
Lon.c1:Lat.p1	38.95	76.15	0.51	0.61
Lon.c1:Lat.p2	-159.98	66.63	-2.40	0.02
Lon.c2:Lat.p1	-105.62	58.42	-1.81	0.07
Lon.c2:Lat.p2	97.86	55.50	1.76	0.08
Lat.p1:Contasia	-216.29	102.92	-2.10	0.04
Lat.p1:Conteurope	483.64	482.59	1.00	0.32
Lat.p1:ContnorthAmerica	16.02	177.07	0.09	0.93
Lat.p1:Contoceania	-1009.45	778.17	-1.30	0.20
Lat.p1:ContsouthAmerica	-50.95	294.19	-0.17	0.86
Lat.p2:Contasia	20.95	80.38	0.26	0.79
Lat.p2:Conteurope	-116.76	203.96	-0.57	0.57
Lat.p2:ContnorthAmerica	15.85	153.37	0.10	0.92
Lat.p2:Contoceania	-610.58	327.08	-1.87	0.06
Lat.p2:ContsouthAmerica	26.27	159.31	0.16	0.87

Table 4.6.: Parameter estimates, standard errors, value of the t-statistics and p-value of the t-test for significance for model M1.social

Table 4.7 summarizes how the model is build using approach 2, i.e. it shows how the model containing all covariates and interactions, that seemed to be significant in section 3.3, is reduced by non significant variables.

Model	Covariates	R^2	adj. R^2	Res. std. error
M2.social	$Birth, Lon.c1, Lon.c2, Lat.p1, Lat.p2, Cont,$ $Lon.c1 \cdot Cont, Lon.c2 \cdot Cont, Lat.p1 \cdot Cont,$ $Lat.p2 \cdot Cont$	0.82	0.77	9.384
M3.social	$Birth, Lon.c1, Lon.c2, Lat.p1, Lat.p2, Af, As,$ $Eu, Na, Oc, As \cdot Lat.p1, As \cdot Lat.p2$	0.76	0.74	10.06
M4.social	$Birth, Lon.c1, Lon.c2, Lat.p1, Lat.p2, As,$ $Na, Oc, As \cdot Lat.p1, As \cdot Lat.p2$	0.75	0.73	10.16

Table 4.7.: Summary of regression models for *Social* using approach 2

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	61.36	3.93	15.63	0.00
Birth	-1.01	0.11	-9.25	0.00
Lon.c1	11.08	3.03	3.66	0.00
Lon.c2	-9.03	2.23	-4.05	0.00
Lat.p1	46.12	16.46	2.80	0.01
Lat.p2	43.48	14.45	3.01	0.00
As	21.30	6.20	3.43	0.00
Oc	23.15	7.42	3.12	0.00
Lat.p1:As	-325.36	73.07	-4.45	0.00
Lat.p2:As	104.83	64.82	1.62	0.11

Table 4.8.: Parameter estimates, standard errors, value of the t-statistics and p-value of the t-test for significance for model M4.social

Since model M1.social contains so many covariates, we decide to choose model M4.social:

$$Social_j = \beta_0 + \beta_1 \cdot birth_j + \beta_2 \cdot lon_j.c1 + \beta_3 \cdot lon_j.c2 + \beta_4 \cdot lat_j.p1 + \beta_5 \cdot lat_j.p2 + \beta_6 \cdot as + \beta_7 \cdot na + \beta_8 \cdot oc + \beta_9 \cdot as \cdot lat_j.p1 + \beta_{10} \cdot as \cdot lat_j.p2 + \varepsilon_{social,j}$$

where $\varepsilon_{social,j} \sim$ iid $N(0, \sigma^2_{social})$.

Analysis of the coefficients:

- For the covariates that also were included in model M6.econ, the analysis is similar since $Social_j$ shows a similar behavior for these variables.
- Comparing $Econ_j$ with $Social_j$, we can see that North America has a significantly higher social index than economic index, relative to the indices of the other continents. This is why, here in contrast to model M6.econ, the variable Na is significant and added to the model with a positive coefficient.

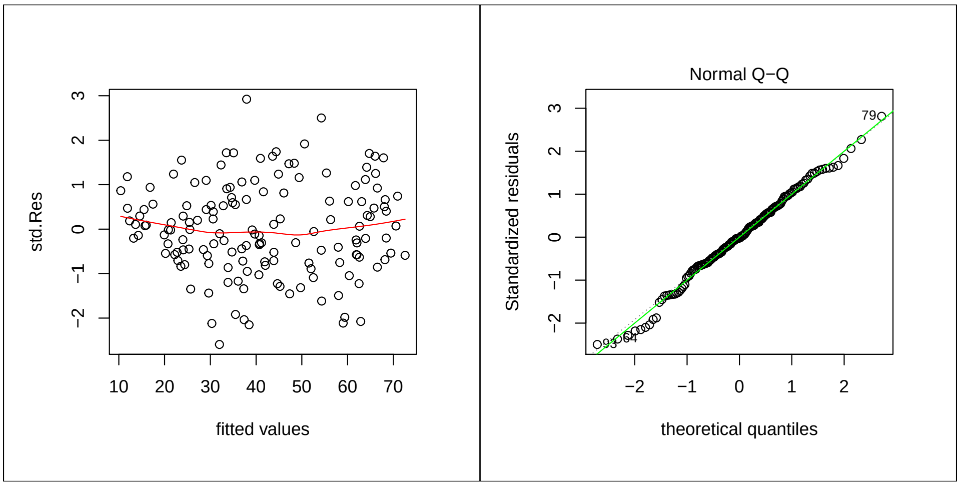

Figure 4.4.: Fitted vs standardized residuals and QQ plot

With the same reasoning as in section 4.1, we conclude that the assumptions of the errors $\varepsilon_{social,j}$ having zero mean and the same variance are appropriate. Furthermore, the QQ-plot indicates again that the standardized residuals are approximately standard normal distributed.

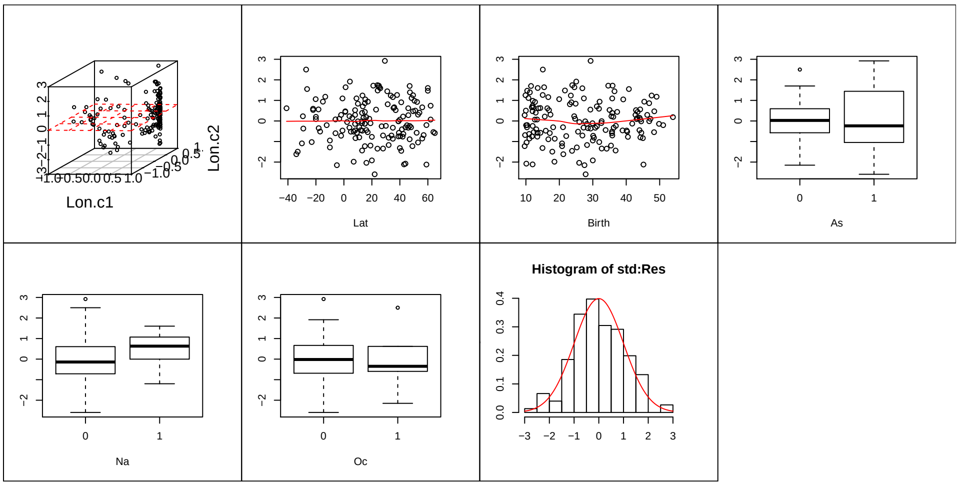

Figure 4.5.: Plots of covariate vs standardized residual and histogram of standardized residuals

Figure 4.5 is interpreted like the corresponding Figure (Figure 4.3) in section 4.1.

4.3. Political index

Here we use the same two approaches, like in section 4.1 and 4.2, to build a linear model for the time averaged political index. Approach 1 yields model M1.polit.

Model	Covariates	R^2	adj. R^2	Res. std. error
M0.polit	$Birth, Lon.c1, Lon.c2, Lat.p1, Lat.p2, Cont,$ *all pairwise interactions*	0.59	0.42	14.12
M1.polit	$Birth, Lon.c1, Lon.c2, Lat.p1, Lat.p2,$ $Birth \cdot Lat.p1, Birth \cdot Lat.p2,$	0.31	0.27	15.73

Table 4.9.: Summary of regression models for *Polit* using approach 1

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	66.68	4.15	16.08	0.00
Birth	-0.41	0.16	-2.53	0.01
Lon.c1	0.42	3.08	0.14	0.89
Lon.c2	-5.26	2.17	-2.42	0.02
Lat.p1	9.84	49.14	0.20	0.84
Lat.p2	157.72	41.88	3.77	0.00
Birth:Lat.p1	0.71	1.78	0.40	0.69
Birth:Lat.p2	-4.30	1.64	-2.62	0.01

Table 4.10.: Parameter estimates, standard errors, value of the t-statistics and p-value of the t-test for significance for model M1.polit

Model	Covariates	R^2	adj. R^2	Res. std. error
M2.polit	$Birth, Lon.c1, Lon.c2, Lat.p1, Lat.p2, Cont,$ $Lon.c1 \cdot Cont, Lon.c2 \cdot Cont, Lat.p1 \cdot Cont,$ $Lat.p2 \cdot Cont$	0.51	0.38	14.49
M3.polit	$Birth, Lon.c1, Lon.c2, Lat.p1, Lat.p2, Cont,$ $Lon.c1 \cdot Cont, Lon.c2 \cdot Cont$	0.43	0.34	14.98
M4.polit	$Birth, Lon.c1, Lon.c2, Lat.p1, Lat.p2, Af, As,$ $Eu, Na, Oc, As \cdot Lat.p1, As \cdot Lat.p2,$ $Na \cdot Lat.p1, Na \cdot Lat.p2$	0.39	0.33	15.09
M5.polit	$Birth, Lon.c1, Lon.c2, Lat.p1, Lat.p2, Af, As,$ $Eu, Na, Oc, Na \cdot Lat.p1, Na \cdot Lat.p2$	0.37	0.32	15.28
M6.polit	$Birth, Lon.c1, Lon.c2, Lat.p1, Lat.p2, Af, As,$ Eu, Na, Oc	0.35	0.30	15.42
M7.polit	$Lon.c1, Lon.c2, Lat.p1, Lat.p2, Na, Eu$	0.34	0.31	15.36

Table 4.11.: Summary of regression models for *Polit* using approach 2

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	61.31	2.84	21.56	0.00
long1_avg	-8.93	3.06	-2.92	0.00
long2_avg	-11.30	2.95	-3.83	0.00
lat1_avg	46.06	22.47	2.05	0.04
lat2_avg	42.07	17.76	2.37	0.02
Na	-14.00	6.33	-2.21	0.03
Eu	14.68	4.60	3.19	0.00

Table 4.12.: Parameter estimates, standard errors, value of the t-statistics and p-value of the t-test for significance for model M7.polit

We choose model M7.polit since it has a higher R^2 and $adj.R^2$.

$$Polit_j = \beta_0 + \beta_1 \cdot lon_j.c1 + \beta_2 \cdot lon_j.c2 + \beta_3 \cdot lat_j.p1 + \beta_4 \cdot lat_j.p2 + \beta_5 \cdot na + \varepsilon_{polit,j}$$

where $\varepsilon_{polit,j} \sim$ iid $N(0, \sigma^2_{polit})$

Analysis of the coefficients:

- The polynomial fitted for latitude is decreasing for countries with latitude smaller than 0, reaches its minimum at 0 where Africa, parts of South America and parts of Asia are located and is increasing for countries with latitude values grater than 0. The polynomial is chosen so that it fits especially Europe well, so North America and parts of Asia, which are at the same latitude, are overestimated.

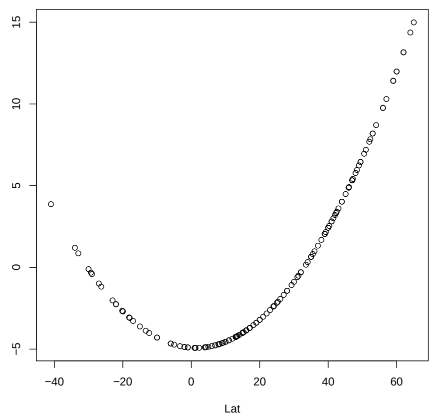

Figure 4.6.: Polynomial fitted in the regression model M7.polit for latitude

- Because of the point above na has a negative coefficient.
- When referring to the circle in Figure 2.2, the coefficients of $lon.c$ are fitted such that the plane takes the lowest values in the upper right quarter of the circle,

where most of the asian countries that have the same latitude like european countries are located. This compensates the overestimation mentioned above.

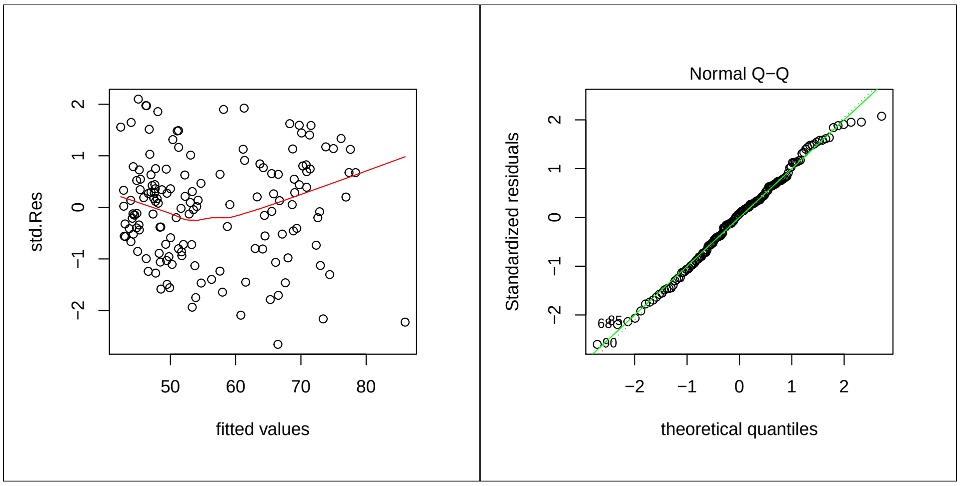

Figure 4.7.: Fitted vs standardized residuals and QQ plot

The plots in Figure 4.7 are not as satisfying as the corresponding ones in section 4.1 and 4.2. The red curve in the fitted vs residuals plot is different from 0 which indicates that the assumption of the errors having zero mean was not appropriate. In contrast to the fitted vs residuals plot, the plots of Figure 4.8 look satisfying and indicate that the errors are independent of the covariates.

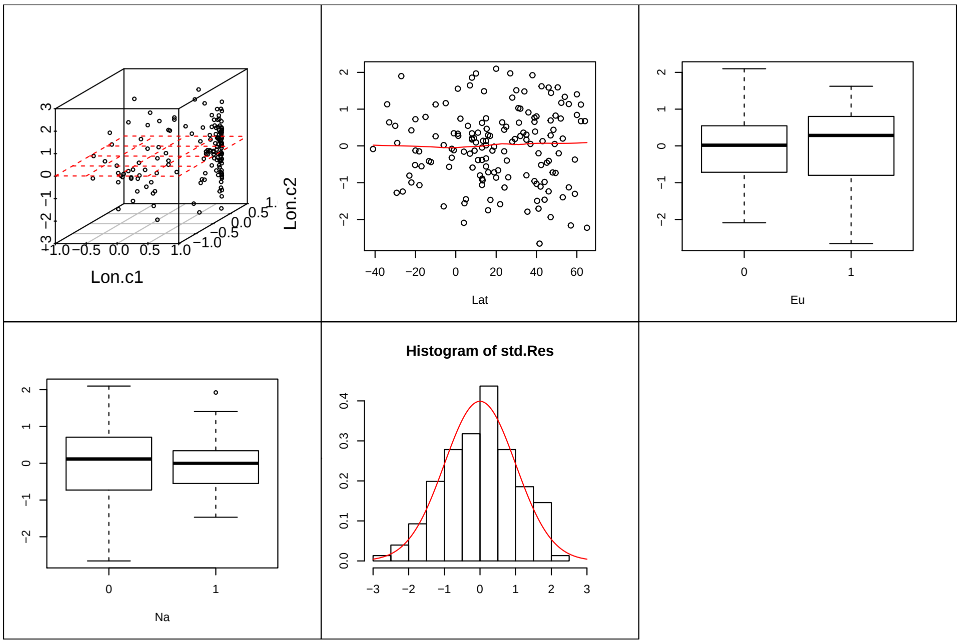

Figure 4.8.: Plots of covariate vs standardized residual and histogram of standardized residuals

5. Introduction to copulas

5.1. Theoretical background

To model dependence among the errors we use copulas. As we will see in Sklar's theorem, any bivariate joint distribution function can be expressed as a function, which is called copula, of the marginals. So the copula describes the dependence structure between the marginals. This section is based on Nelsen (2006).

Definition 5.1.1 (Copula). *$C : [0,1]^d \to [0,1]$ is a d-dimensional copula if C is a joint cumulative distribution function of a d-dimensional random vector on the unit cube $[0,1]^d$ with uniform marginals. Here bivariate copulas will be used, i.e. d=2.*

Theorem 5.1.1 (Sklar's theorem for bivariate copulas). *Let F be a bivariate joint distribution function with margins F_1 and F_2. Then there exists a copula C such that for all $x, y \in \mathbb{R}$,*

$$F(x,y) = C(F_1(x), F_2(y)).$$

If F_1 and F_2 are continuous, then C is unique. Conversely, if C is a copula and F_1 and F_2 are distribution functions, then $F(x,y) = C(F_1(x), F_2(y))$ is a joint distribution function with margins F_1 and F_2.

This theorem is proven in Nelsen (2006) (theorem 2.2.3 on page 18).

Definition 5.1.2. *Let F, F_1, F_2 and C be like in Sklar's Theorem and X and Y random variables with distribution functions F_1 and F_2 , then we call C the copula of F_1 and F_2 respectively the copula of X and Y and denote it by C_{F_1,F_2} respectively $C_{X,Y}$.*

An important property of copulas is that they are invariant with respect to strictly monotone increasing functions. This is the statement of the next theorem.

Theorem 5.1.2 (Invariance). *Let X and Y be continuous random variables, f_1 and f_2 functions that are strictly increasing on the range of X and the range of Y, respectively, then*

$$C_{f_1(X),f_2(Y)} = C_{X,Y}.$$

A proof of this theorem can be found in Nelsen (2006) (theorem 2.4.3 on page 25).

Furthermore we need an appropriate measure of dependence. E.g. Pearson's correlation coefficient ρ ($\rho = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$ for two random variables X, Y with existing positive standard deviations σ_X, σ_Y and covariance $Cov(X, Y)$) would not satisfy our needs since it can only measure linear dependencies. An appropriate measure is Kendall's tau which is introduced now. For introducing Kendall's tau we first need to define concordance and discordance.

Definition 5.1.3 (concordance, discordance). *Let (x_i, y_i) and (x_j, y_j) denote two observations from a vector (X, Y) of continuous random variables. We say that (x_i, y_i) and (x_j, y_j) are concordant if $(x_i - x_j)(y_i - y_j) > 0$ and discordant if $(x_i - x_j)(y_i - y_j) < 0$.*

Definition 5.1.4 (empirical Kendall's tau). *Let $\{(x_1, y_1), ..., (x_n, y_n)\}$ be a random sample of n observations from a vector (X,Y) of continuous random variables. Let c be the number of the concordant and d be be the number of the discordant pairs of the $\binom{n}{2}$ distinct pairs (x_i, y_i), (x_j, y_j). Then empirical Kendall's tau τ is defined as*

$$\tau = \frac{c-d}{\binom{n}{2}}.$$

Definition 5.1.5 (Kendall's tau). *Let (X, Y) be a vector of two continuous random variables with joint distribution function F and (X_1, Y_1), (X_2, Y_2) be independent and identically distributed random variables with joint distribution function F. Then Kendall's tau is given by*

$$\tau = P[(X_1 - X_2)(Y_1 - Y_2) > 0] - P[(X_1 - X_2)(Y_1 - Y_2) < 0].$$

5.2. Copula families

Here we will see some examples of copulas which will later help us to model the dependence structure. These copulas can be separated into two families, the elliptical and the Archimedean copulas. Furthermore we will give the relationship between Kendall's tau and the parameters of the copula. Concepts we will need in this section are the quasi inverse, the pseudo inverse and elliptical distributions.

Definition 5.2.1 (quasi-inverse). *Let F be a distribution function. A quasi inverse of F is any function $F^{(-1)}$ with domain $[0, 1]$ such that*

$$F^{(-1)}(x) = \begin{cases} any\ number\ x \in [-\infty, \infty]\ such\ that\ F(x) = t & if\ t \in RanF \\ inf\ \{x | F(x) \geq t\} & if\ t \notin RanF \end{cases}$$

where RanF is the range of F.

(see Nelsen(2006) page 21)

Definition 5.2.2 (pseudo-inverse). *Let $\varphi : [0, 1] \to [0, \infty]$ be a continuous, strictly decreasing function such that $\varphi(1) = 0$. The pseudo-inverse of φ is the function $\varphi^{[-1]} : [0, \infty] \to [0, 1]$ given by*

$$\varphi^{[-1]}(t) = \begin{cases} \varphi^{-1}(t) & if \quad 0 \le t \le \varphi(0) \\ 0 & if \quad \varphi(0) \le t \le \infty \end{cases} .$$

(see Nelsen(2006) page 110)

The definition of elliptical distributions is based on spherical distributions.

Definition 5.2.3 (Spherical distribution)**.** *A d-dimensional random vector X has a spherical distribution if for every orthogonal matrix $O \in \mathbb{R}^{d \times d}$ holds*

$$OX \sim O.$$

(see Mai, Scherer(2012) page 161)

Definition 5.2.4 (Elliptical distribution)**.** *The d-dimensional random vector X has an elliptical distribution if*

$$X \sim \mu + A^T Y$$

where Y has a spherical distribution on $\mathbb{R}^k$, $A \in \mathbb{R}^{k \times d}$, rank of $A^T A = k \le d$ and $\mu \in \mathbb{R}^d$.

(see Mai, Scherer(2012) page 166)

Furthermore, we need in addition to the univariate normal and t distribution, defined in chapter 4, the bivariate normal and the bivariate t distribution.
The density function of the bivariate normal distribution with mean $\mu = (\mu_1, \mu_2)$, marginal standard deviations σ_1 and σ_2 ($\sigma = (\sigma_1, \sigma_2)$) and correlation ρ is

$$\varphi_{2,\mu,\sigma,\rho}(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\left[\frac{(x_1-\mu_1)^2}{\sigma_1^2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2} - \frac{2\rho(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2}\right]\right) \quad \forall x_1, x_2 \in \mathbb{R}.$$

The bivariate t distribution with mean $\mu = (\mu_1, \mu_2)$, ν degrees of freedom and correlation ρ has the density function

$$f_{2,\mu,\nu,\rho}(x_1, x_2) = \frac{\Gamma(\frac{\nu+2}{2})}{\Gamma(\frac{\nu}{2})\pi\nu\sqrt{1-\rho^2}} \left(1 + \frac{(x_1-\mu_1)^2 + (x_2-\mu_2)^2 - 2\rho(x_1-\mu_1)(x_2-\mu_2)}{\nu(1-\rho^2)}\right)^{-\frac{\nu+2}{2}} \quad \forall x_1, x_2 \in \mathbb{R}.$$

5.2.1. Elliptical copulas

Sklar's Theorem tells us a way of constructing copulas. By setting $u_1 = F_1(x)$ and $u_2 = F_2(y)$ (so $u_1, u_2 \in [0,1]$), we can conclude that for a continuous bivariate joint distribution function F with marginals F_1 and F_2 the copula of F_1 and F_2 is given by

$$C_{F_1,F_2}(u_1, u_2) = F(F_1^{(-1)}(u_1), F_2^{(-1)}(u_2)) \text{ for all } u_1, u_2 \in [0,1].$$

This can be applied to various distributions, in particular to elliptical distributions, which gives us the family of elliptical copulas. Two well known elliptical distributions are the normal and Student's t distribution.

Theorem 5.2.1. *For a bivariate elliptical copula with correlation ρ Kendall's tau is given by*

$$\tau = \frac{2}{\pi} \cdot arcsin(\rho).$$

For the proof, see Lindskog, McNeil, and Schmock (2002) (theorem 2 on page 3).

Applying the copula construction method above to a bivariate standard normal distribution, we obtain the Gaussian copula.

Definition 5.2.5 (Gaussian copula)**.** *Let Φ be the standard normal distribution function and Φ_2 be the bivariate normal distribution function with zero mean, marginal standard deviations both 1 and correlation ρ. Then the bivariate Gaussian copula with parameter ρ is the copula C such that*

$$C(u_1, u_2) = \Phi_2(\Phi^{-1}(u_1), \Phi^{-1}(u_2)).$$

(see Mai, Scherer (2012) page 176)

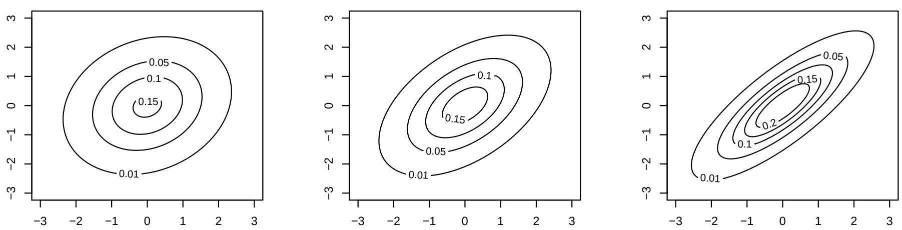

Figure 5.1.: Contour plot of the Gaussian copula with parameter $\rho = 0.3, 0.5, 0.8$ (from left to right) and corresponding values of Kendall's tau 0.13, 0.33, 0.59

Definition 5.2.6 (t copula)**.** *Let ν be the degrees of freedom and ρ the correlation of a bivariate t distribution with zero mean. T_2 denotes its distribution function. Furthermore T is the distribution function of the standard t distribution (zero mean) with ν degrees of freedom. The bivariate t-copula with parameter ν and ρ is given by*

$$C(u_1, u_2) = T_2(T^{-1}(u_1), T^{-1}(u_2)).$$

(see, Mai, Scherer (2012) page 177)

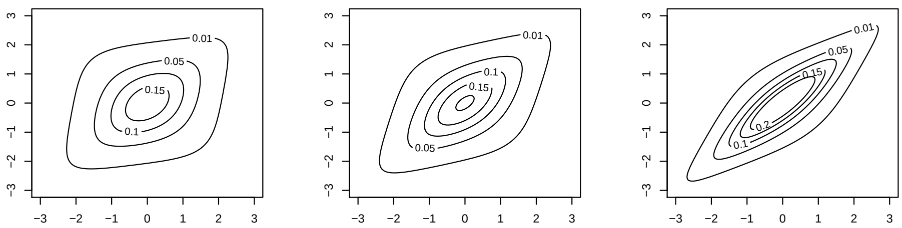

Figure 5.2.: Contour plot of the t copula with 4 degrees of freedom and $\rho = 0.3, 0.5, 0.8$ (from left to right) and corresponding values of Kendall's tau 0.19, 0.33, 0.59

Remark: The t-copula converges to the Gaussian copula as the degrees of freedom go to infinity. (see Lindner, Szimayer (2005))

5.2.2. Archimedean copulas

Definition 5.2.7 (Archimedean copula). *Let $\varphi : [0,1] \to [0,\infty]$ be a continuous strictly monotone decreasing convex function with $\varphi(1) = 0$. Then*

$$C(u_1, u_2) = \varphi^{[-1]}(\varphi(u_1) + \varphi(u_2))$$

is a copula and is called Archimedean copula with generator φ.

The proof that C is indeed a copula can be found in Nelson (2006) (theorem 4.1.4 on page 111)

Theorem 5.2.2 (Kendall's tau for Archimedean copulas). *Kendall's tau of a bivariate Archimedean copula with generator φ is given by*

$$\tau = 1 + 4 \int_0^1 \frac{\varphi(t)}{\varphi'(t)} dt.$$

The proof can be found in Nelson (2006) (corollary 5.1.4 on page 163).

For the following definitions we refer to Nelsen(2006) page 116/117.

Definition 5.2.8 (Clayton copula). *For $\delta \in [-1, \infty) \setminus \{0\}$*

$$C(u_1, u_2) = max((u_1^{-\delta} + u_2^{-\delta} - 1)^{-\frac{1}{\delta}}, 0)$$

is the Clayton copula with parameter δ. The corresponding generator is $\varphi(t) = \frac{1}{\delta}(t^{-\delta} - 1)$.

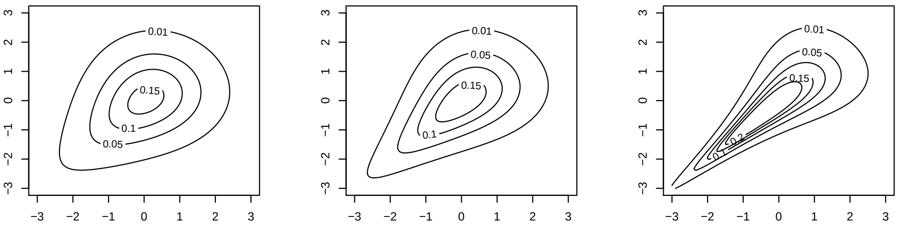

Figure 5.3.: Contour plot of the Clayton copula with $\delta = 0.5, 1, 3$ (from left to right) and corresponding values of Kendall's tau 0.20, 0.33, 0.60

Definition 5.2.9 (Gumbel copula)**.** *The Gumbel copula with parameter* $\delta \in [1, \infty)$ *is generated by* $\varphi(t) = (-\ln t)^{\delta}$*. It is given by*

$$C(u_1, u_2) = exp(-[(-ln(u_1))^{\delta} + (-ln(u_2))^{\delta}]^{\frac{1}{\delta}}).$$

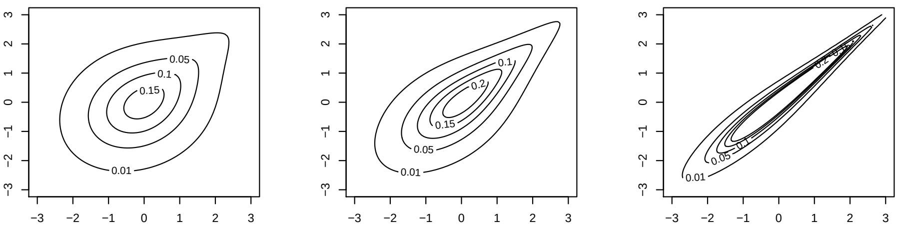

Figure 5.4.: Contour plot of the Gumbel copula with parameter $\delta = 1.3, 2, 6$ (from left to right) and corresponding values of Kendall's tau 0.23, 0.50, 0.83

Definition 5.2.10 (Frank copula)**.** *For* $\delta \in (-\infty, \infty) \setminus \{0\}$

$$C(u_1, u_2) = -\frac{1}{\delta} \ln \left(1 + \frac{(e^{-\delta u_1} - 1)(e^{-\delta u_2} - 1)}{e^{-\delta} - 1}\right)$$

is the Frank copula with parameter δ*. The generator of the Frank copula is* $\varphi(t) = -\ln(\frac{e^{-\delta t} - 1}{e^{-\delta} - 1})$.

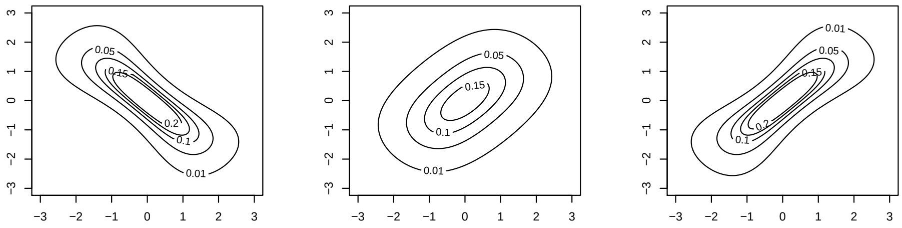

Figure 5.5.: Contour plot of the Frank copula with parameter $\delta = -10, 3, 10$ (from left to right) and corresponding values of Kendall's tau -0.67, 0.31, 0.67

6. Applying copulas to the data set

6.1. Modelling of the joint distribution function of two errors with copulas

We consider the models M6.econ, M4.social and M7.polit from chapter 4 and model the joint distribution function of two errors with a copula. Therefore, we will use maximum likelihood estimation to estimate the parameter of a copula for each family and then we will calculate the Akaike information criteria (AIC) for each of the fitted copulas. We choose the copula with the lowest AIC. To do so, we assume in this section that the values of the regression coefficients $\hat{\beta}_i$ and of the standard deviation of the error $\hat{\sigma}$ are fixed, i.e we assume the following model:

$$\begin{aligned} Econ_j = {} & \hat{\beta}_{0econ} + \hat{\beta}_{1econ} \cdot birth_j + \hat{\beta}_{2econ} \cdot lon_j.c1 + \hat{\beta}_{3econ} \cdot lon_j.c2 + \hat{\beta}_{4econ} \cdot lat_j.p1 + \\ & \hat{\beta}_{5econ} \cdot lat_j.p2 + \hat{\beta}_{6econ} \cdot as + \hat{\beta}_{7econ} \cdot oc + \hat{\beta}_{8econ} \cdot as \cdot lat_j.p1 + \hat{\beta}_{9econ} \cdot as \cdot lat_j.p2 \\ & + \hat{\varepsilon}_{econ,j} \\ = {} & econ\_mean\_pred_j + \hat{\varepsilon}_{econ,j} \end{aligned}$$

$$\begin{aligned} Social_j = {} & \hat{\beta}_{0social} + \hat{\beta}_{1social} \cdot birth_j + \hat{\beta}_{2social} \cdot lon_j.c1 + \hat{\beta}_{3social} \cdot lon_j.c2 + \hat{\beta}_{4social} \cdot lat_j.p1 \\ & + \hat{\beta}_{5social} \cdot lat_j.p2 + \hat{\beta}_{6social} \cdot as + \hat{\beta}_{7social} \cdot na + \hat{\beta}_{8social} \cdot oc + \hat{\beta}_{9social} \cdot as \cdot \\ & lat_j.p1 + \hat{\beta}_{10social} \cdot as \cdot lat_j.p2 + \hat{\varepsilon}_{social,j} \\ = {} & social\_mean\_pred_j + \hat{\varepsilon}_{social,j} \end{aligned}$$

$$\begin{aligned} Polit_j = {} & \hat{\beta}_{0polit} + \hat{\beta}_{1polit} \cdot lon_j.c1 + \hat{\beta}_{2polit} \cdot lon_j.c2 + \hat{\beta}_{3polit} \cdot lat_j.p1 + \hat{\beta}_{4polit} \cdot lat_j.p2 + \\ & \hat{\beta}_{5polit} \cdot na + \hat{\varepsilon}_{polit,j} \\ = {} & polit\_mean\_pred_j + \hat{\varepsilon}_{polit,j} \end{aligned}$$

$$\begin{aligned} F_{econ,social}(x,y) &= C_{econ,social}(\Phi_{econ}(x), \Phi_{social}(y)) \quad \forall x,y \in \mathbb{R} \\ F_{econ,polit}(x,y) &= C_{econ,polit}(\Phi_{econ}(x), \Phi_{polit}(y)) \quad \forall x,y \in \mathbb{R} \\ F_{social,polit}(x,y) &= C_{social,polit}(\Phi_{social}(x), \Phi_{polit}(y)) \quad \forall x,y \in \mathbb{R} \end{aligned}$$

where $\hat{\varepsilon}_{econ,j} \sim$ iid $N(0, \hat{\sigma}^2_{econ})$, $\hat{\varepsilon}_{social,j} \sim$ iid $N(0, \hat{\sigma}^2_{social})$, $\hat{\varepsilon}_{polit,j} \sim$ iid $N(0, \hat{\sigma}^2_{polit})$, $F_{econ,social}$, $F_{econ,polit}$ and $F_{social,polit}$ are the joint distribution functions belonging to the random vectors $(\hat{\varepsilon}_{econ,j}, \hat{\varepsilon}_{social,j}), (\hat{\varepsilon}_{econ,j}, \hat{\varepsilon}_{polit,j})$ and $(\hat{\varepsilon}_{social,j}, \hat{\varepsilon}_{polit,j})$. $C_{econ,social}$, $C_{econ,polit}$ and $C_{social,polit}$ are copulas. Sklar's theorem ensures the existence of these copulas. The $\hat{\beta}_i, \hat{\sigma}_i$ are no random variables but the estimates from chapter 4, so they are constants. $econ\_mean\_pred_j$, $social\_mean\_pred_j$ and $polit\_mean\_pred_j$ are the mean predictions. They are defined by the equation above. Since $\hat{\beta}_i$ is constant, the mean predictions are also constants. $\Phi_{econ}, \Phi_{social}, \Phi_{polit}$ are the distribution functions of $\hat{\varepsilon}_{econ,j}, \hat{\varepsilon}_{social,j}, \hat{\varepsilon}_{polit,j}$.

To estimate the copulas we need first of all appropriate data. To get this data we will use the fact that if X is a continuous random variable with distribution function F, then F(X) ~ unif[0,1] since

$$P[F(X) \leq t] = P[X \leq F^{(-1)}(t)] = F(F^{(-1)}(t)) = t$$

$\frac{\hat{\varepsilon}_{econ,j}}{\hat{\sigma}_{econ}}, \frac{\hat{\varepsilon}_{social,j}}{\hat{\sigma}_{social}}, \frac{\hat{\varepsilon}_{polit,j}}{\hat{\sigma}_{polit}}$ are by assumption standard normal distributed. So by the following transform we get unif[0,1] distributed variables that allow us to work with copulas.

$$\begin{aligned} U_{econ,j} &= \Phi(\tfrac{\hat{\varepsilon}_{econ,j}}{\hat{\sigma}_{econ}}) \\ U_{social,j} &= \Phi(\tfrac{\hat{\varepsilon}_{social,j}}{\hat{\sigma}_{social}}) \\ U_{polit,j} &= \Phi(\tfrac{\hat{\varepsilon}_{polit,j}}{\hat{\sigma}_{polit}}) \end{aligned}$$

where Φ is the standard normal distribution function. With this data we will estimate the copula $C_{U_{econ,j},U_{social,j}}$. Since Φ is strictly monotone increasing, theorem 5.1.2 tells us $C_{econ,social} = C_{U_{econ,j},U_{social,j}}$. Furthermore, we will from now on leave out the "j" in the index and only write U_{econ}, $\hat{\varepsilon}_{econ}$ and so on, if it is appropriate.
We have a look at contour plots and values of empirical Kendall's tau to decide whether there is evidence for dependence.

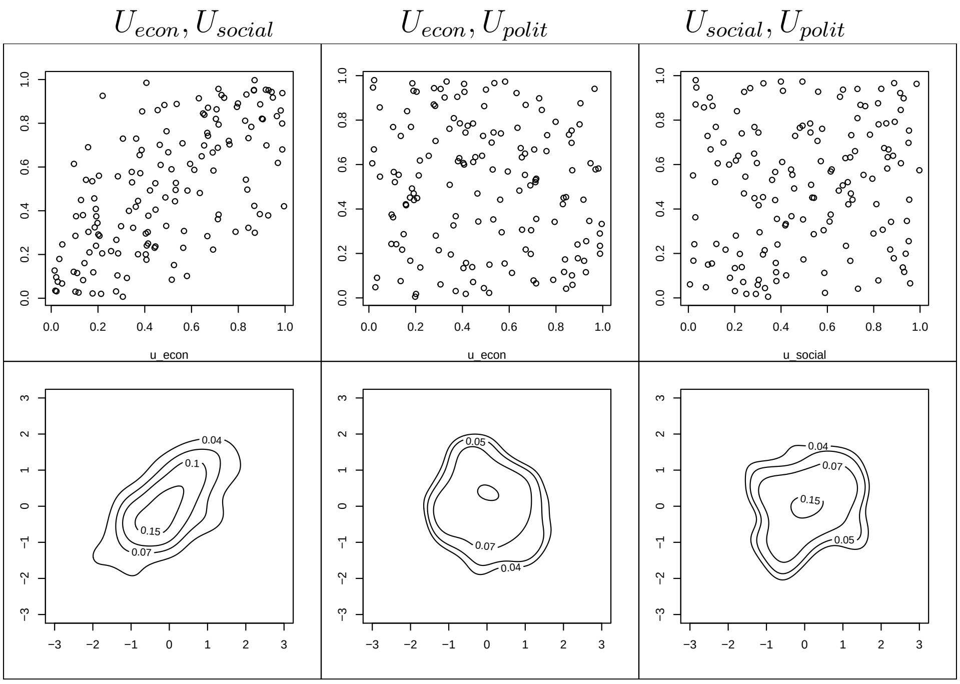

Figure 6.1.: Contour plots and scatter plots of the observations of the three bivariate distributions obtained by combining U_{econ}, U_{social} and U_{polit}.

In these plots we can see that there is not much evidence for dependencies between U_{econ} and U_{polit} as well as for U_{social} and U_{polit}. But there seems to be a dependency between U_{econ} and U_{social}. This coincides with Table 6.1 where we can see the values of Kendall's tau.

		empirical Kendall's tau
U_{econ}	U_{social}	0.45
U_{econ}	U_{polit}	-0.07
U_{social}	U_{polit}	0.09

Table 6.1.: Empirical Kendall's tau calculated for the observations of (U_{econ}, U_{social}), (U_{econ}, U_{polit}) and (U_{social}, U_{polit})

So we decide to estimate $C_{U_{econ},U_{polit}}$ and $C_{U_{social},U_{polit}}$ by the independence copula which is given by $C(u_1, u_2) = u_1 \cdot u_2$. For $C_{U_{econ},U_{social}}$ we try to find the copula that fits the dependency structure best. In a first step we estimate the parameter of all available copula families using maximum likelihood estimation, i.e. the estimate $\hat{\theta}$ of θ for a copula C is the value of θ that maximizes the log likelihood function which is given by $\sum_{i=1}^{151} ln[c((u_{i1}, u_{i2})|\theta)]$ where $c(u_{i1}, u_{i2}|\theta)$ is the density of copula C with parameter θ evaluated at observation (u_{i1}, u_{i2}).

Copula	Parameter 1	Parameter2	Kendall's tau
Gaussian	0.63		0.44
Student t	0.64	>30	0.44
Clayton	1.18		0.37
Gumbel	1.64		0.39
Frank	4.96		0.45

Table 6.2.: Maximum likelihood estimates for each copula

Since the estimated degrees of freedom for the t copula are bigger than 30, we consider the Gaussian copula instead.
Now we can calculate the Akaike information criteria (AIC) for each of these fitted copulas. We will estimate $C_{U_{econ},U_{social}}$ by the copula with the lowest AIC. The AIC of a Copula C with k-dimensional parameter θ for N observations $\{(u_{1,1}, u_{1,2}), ...(u_{N,1}, u_{N,2})\}$ is given by

$$AIC = -2 \sum_{i=1}^{N} ln[c(u_{i,1}, u_{i,2}|\theta)] + 2k.$$

Copula	AIC
Gaussian	-69.81
Clayton	-62.84
Gumbel	-55.62
Frank	-73.28

Table 6.3.: AIC calculated for each copula with copula parameter obtained from Table 6.2

According to Table 6.3, we decide to estimate $C_{U_{econ},U_{social}} = C_{econ,social}$ by a Frank copula with parameter 4.96, denoted by $\hat{C}_{econ,social}$. The corresponding Kendall's tau of the Frank copula is very close to the empirical Kendall's tau (compare Table 6.1 and Table 6.2).

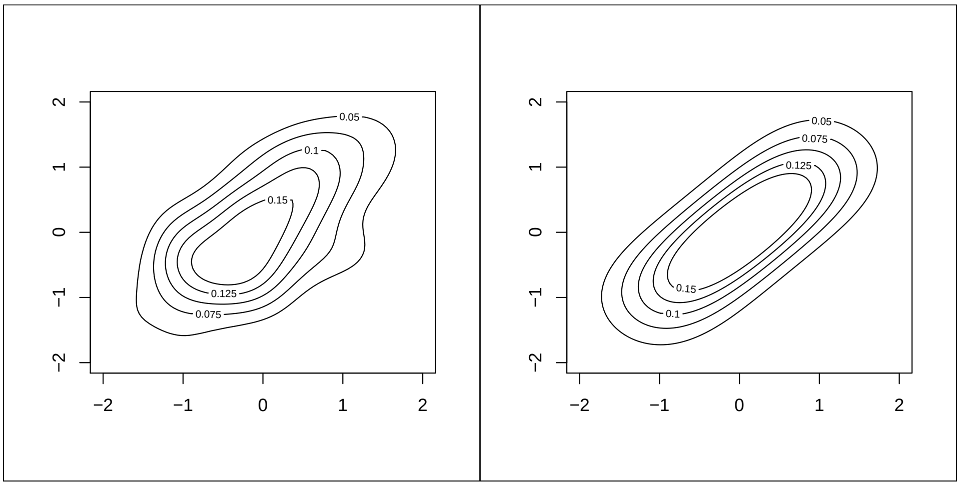

Figure 6.2.: Empirical contour plot of (U_econ, U_social) (left) and contour plot of $\hat{C}_{econ,social}$ (right)

6.2. Prediction of the economic and social index in Germany with and without using copulas

In this section $F_{X,Y}$ denotes the joint distribution function of two random variables X and Y and $f_{X,Y}$ its density function. We compare predictions for the german economic and social index using the copula with predictions without using it (independence copula).

We consider the copula model (j = 38 is Germany)

$$\begin{aligned}
&econ_ger_c = econ_mean_pred_{38} + \hat{\varepsilon}_{econ}\\
&social_ger_c = social_mean_pred_{38} + \hat{\varepsilon}_{social}\\
&F_{\hat{\varepsilon}_{econ},\hat{\varepsilon}_{social}}(x,y) = \hat{C}_{econ,social}(\Phi_{econ}(x), \Phi_{social}(y)) \quad \forall x, y \in \mathbb{R}
\end{aligned}$$

and the non copula model

$$\begin{aligned}
&econ_ger_nc = econ_mean_pred_{38} + \hat{\varepsilon}_{econ}\\
&social_ger_nc = social_mean_pred_{38} + \hat{\varepsilon}_{social}\\
&F_{\hat{\varepsilon}_{econ},\hat{\varepsilon}_{social}}(x,y) = \Phi_{econ}(x) \cdot \Phi_{social}(y) \quad \forall x, y \in \mathbb{R}
\end{aligned}$$

For the copula model we simulate 10000 times from $\hat{C}_{econ,social}$ and get simulations $(u\_c_{sim,1,1}, u\_c_{sim,2,1}), (u\_c_{sim,1,2}, u\_c_{sim,2,2}), ..., (u\_c_{sim,1,10000}, u\_c_{sim,2,10000})$. The simulated predictions for the german economic and social index are then given by

$$econ\_ger\_c_j = econ\_mean\_pred_{38} + \Phi^{-1}(u\_c_{sim,1,j}) \cdot \hat{\sigma}_{econ} = 68.9 + \Phi^{-1}(u\_c_{sim,1,j}) \cdot 11.89$$
$$social\_ger\_c_j = social\_mean\_pred_{38} + \Phi^{-1}(u\_c_{sim,2,j}) \cdot \hat{\sigma}_{social} = 68.1 + \Phi^{-1}(u\_c_{sim,2,j}) \cdot 10.57$$

for $j = 1...10000$.
When we assume independence among the errors we simulate $(u\_nc_{sim,1,1}, u\_nc_{sim,2,1}), (u\_nc_{sim,1,2}, u\_nc_{sim,2,2}), ..., (u\_nc_{sim,1,10000}, u\_nc_{sim,2,10000})$ from two independent unif[0,1] distributions and obtain the simulated predictions by

$$\begin{aligned}
econ_ger_nc_j &= econ_mean_pred_{38} + \Phi^{-1}(u_nc_{sim,1,j}) \cdot \hat{\sigma}_{econ}\\
&= 68.9 + \Phi^{-1}(u_nc_{sim,1,j}) \cdot 11.89\\
social_ger_nc_j &= social_mean_pred_{38} + \Phi^{-1}(u_nc_{sim,2,j}) \cdot \hat{\sigma}_{social}\\
&= 68.1 + \Phi^{-1}(u_nc_{sim,2,j}) \cdot 10.57
\end{aligned}$$

for $j = 1...10000$.

Furthermore we can calculate the joint density of the random vectors $(econ\_ger\_c, social\_ger\_c)$ and $(econ\_ger\_nc, social\_ger\_nc)$. We obtain it by differentiating the joint distribution function. Let φ_σ be the density function of a normal distributed random variable with zero mean and standard deviation σ and $\hat{c}_{econ,social}(x,y) = \frac{d}{dy}\frac{d}{dx}\hat{C}_{econ,social}(x,y)$ be the copula density. Then the joint density in the copula model is for all $x, y \in \mathbb{R}$ given by

$$\begin{aligned}
&f_{econ_ger_c,social_ger_c}(x,y) = \tfrac{d}{dy}\tfrac{d}{dx}F_{econ_ger_c,social_ger_c}(x,y) =\\
&\tfrac{d}{dy}\tfrac{d}{dx}F_{68.9+\hat{\varepsilon}_{econ},68.1+\hat{\varepsilon}_{social}}(x,y) = \tfrac{d}{dy}\tfrac{d}{dx}F_{\hat{\varepsilon}_{econ},\hat{\varepsilon}_{social}}(x-68.9, y-68.1) =\\
&\tfrac{d}{dy}\tfrac{d}{dx}\hat{C}_{econ,social}(\Phi_{econ}(x-68.9), \Phi_{social}(y-68.1)) =\\
&\quad \hat{c}_{econ,social}(\Phi_{econ}(x-68.9), \Phi_{social}(y-68.1)) \cdot \varphi_{\hat{\sigma}_{econ}}(x-68.9) \cdot \varphi_{\hat{\sigma}_{social}}(y-68.1).
\end{aligned}$$

If we assume independence among the errors (non copula model) the joint density is given by

$$f_{econ\_ger\_nc,social\_ger\_nc}(x, y) = \varphi_{\hat{\sigma}_{econ}}(x - 68.9) \cdot \varphi_{\hat{\sigma}_{social}}(y - 68.1) \quad \forall x, y \in \mathbb{R}.$$

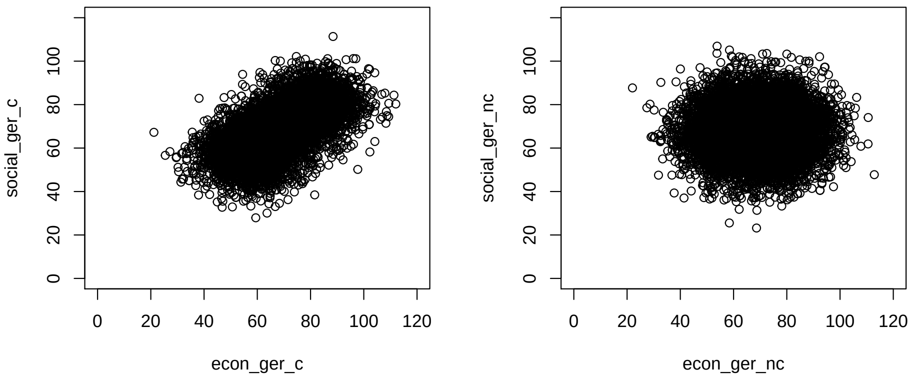

Figure 6.3.: Scatter plot of simulated predictions of the economic index against the simulated predictions of the social index, once for the copula (left) and once for the non copula model (right)

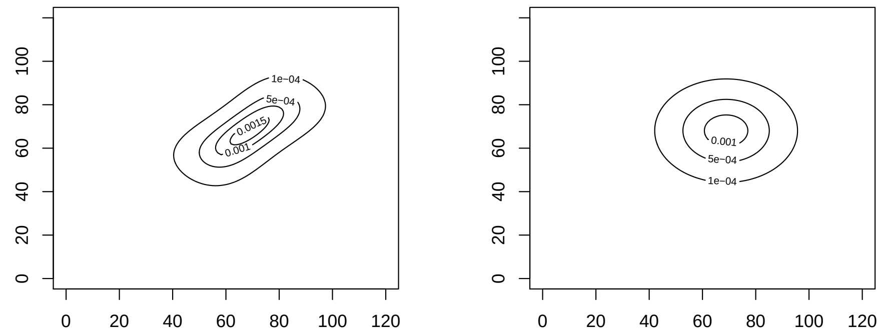

Figure 6.4.: Contour plot of $f_{econ\_ger\_c,social\_ger\_c}$ (left) and $f_{econ\_ger\_nc,social\_ger\_nc}$ (right)

In Figure 6.3 and 6.4 we observe shapes that correspond with the Frank copula and the independence copula (circles).
In addition we have a look at the 95% confidence regions of the prediction for the two models (Figure 6.5). We again observe corresponding shapes. Furthermore we can see

that the confidence region in the copula model is smaller which makes it superior to the non copula model.

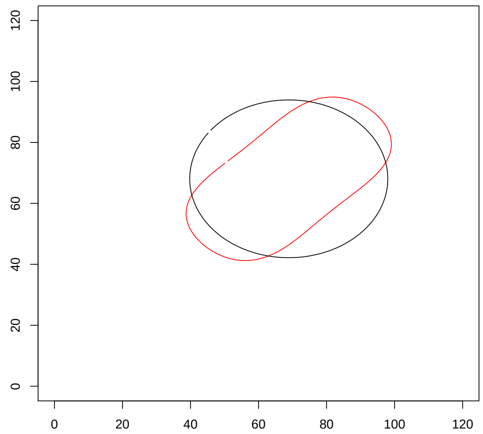

Figure 6.5.: 95% confidence regions of the prediction for the copula model (red) and the non copula model (black).

7. Outlook

As we have seen, copulas are a very useful tool when it comes to model dependency structures. We have only given a short insight in the theory of copulas and only bivariate copulas were covered. Elliptical and Archimedean copulas can easily be generalized in more dimensions. A d-dimensional elliptical copula is the copula of a d-dimensional elliptical distribution and a d-dimensional Archimedean copula is characterized by a generator φ as defined is section 5.2.2 and is given by

$$C(u_1, u_2, ..., u_d) = \varphi^{[-1]}(\varphi(u_1) + \varphi(u_2) + ... + \varphi(u_d)) \quad \forall u_1, u_2, ..., u_d \in [0, 1].$$

Another way of constructing copulas are pair copula constructions which are based on bivariate copulas (see Aast, Czado, Frigessi, Bakken (2006)). We will have a look at the three-dimensional case. Let (X_1, X_2, X_3) be a random vector with distribution function F and density f. Let furthermore f_{12} denote the joint density of X_1 and X_2 and F_{12} its distribution function, f_1 the density of X_1 and so on. $f(\ |\ )$ and $F(\ |\ )$ denote the corresponding conditional density and distribution functions. It holds

$$f(x_1, x_2, x_3) = f(x_1|x_2, x_3) \cdot f_{23}(x_2, x_3) = f(x_1|x_2, x_3) \cdot f(x_2|x_3) \cdot f_3(x_3).$$

Sklar's theorem tells us that there exists a copula density c_{23} such that

$$f(x_2, x_3) = c_{23}(F_2(x_2), F_3(x_3)) \cdot f_2(x_2) \cdot f_3(x_3).$$

So the conditional density can be expressed as follows

$$f(x_2|x_3) = c_{23}(F_2(x_2), F_3(x_3)) \cdot f_2(x_2)$$

and

$$f(x_1|x_2, x_3) = \tfrac{f(x_1,x_2|x_3)}{f(x_1|x_3)f(x_2|x_3)} \cdot f(x_1|x_3) = c_{12|3}(F(x_1|x_3), F(x_2|x_3)) \cdot f(x_1|x_3) = \\ c_{12|3}(F(x_1|x_3), F(x_2|x_3)) \cdot c_{13}(F_1(x_1), F_3(x_3)) \cdot f_1(x_1).$$

with bivariate copula densities $c_{12|3}, c_{23}$ and c_{13}. Using the results above we get for the joint density

$$f(x_1, x_2, x_3) = c_{12|3}(F(x_1|x_3), F(x_2|x_3)) \cdot c_{13}(F_1(x_1), F_3(x_3)) \cdot \\ f_1(x_1) \cdot c_{23}(F_2(x_2), F_3(x_3)) \cdot f_2(x_2) \cdot f_3(x_3).$$

We see that the joint density function can be expressed as a product of bivariate copula densities and marginal densities. This representation also holds for higher dimensions. By using different bivariate copulas this representation allows us to obtain a wide range of high-dimensional copulas. For these copulas many new exciting questions arise, e.g. which results of bivariate copulas can be generalized to the multivariate case. Here we could have used a three-dimensional copula to model the dependency structure of all three errors at once and in general high-dimensional copulas allow us to model complex dependencies of many dimensions. This is why they find application in many fields like finance, engineering or medicine and why further study in this field is desirable.

Bibliography

[1] Aast K., Czado C., Frigessi A., Bakken H. (2006) *Pair-copula constructions of multiple dependence.*

[2] Durante F., Härdle W., Jaworski P., Rychlik T. (2010). *Copula Theory and its Applications.* Springer-Verlag, Berlin Heidelberg.

[3] Embrechts P., Lindskog F., McNeil A. (2001). *Modelling Dependence with Copulas and Application to Risk Management.* ETH Zürich.

[4] Fahrmeir L., Künstler R., Pigeot I., Tutz G. (2011) *Statistik Der Weg zur Datenanalyse.* Springer-Verlag, Berlin Heidelberg.

[5] Guangming P. (2013). *Lecture notes to MH3510 Regression Analysis.* Nanyang Technological University Singapore.

[6] Lindner A., Szimayer A. (2005). *A Limit Theorem for Copulas.*

[7] Lindskog F., McNeil A., Schmock U. (2002). *Kendall's Tau for Elliptical Distributions.*

[8] Mai J., Scherer M. (2012) *Simulating copulas.* World Scientific, Singapore.

[9] Montgomery D. (2012) *Introduction to Linear Regression Analysis.* Wiley & Sons, Inc., Hoboken, New Jersey.

[10] Nelsen R. (2006). *An Introduction to Copulas.* Springer, New York.

A. R-Code

A.1. Definition of basic funtions

```
#function to eliminate leading NAs per country

NA.eliminate=function(country,x)
{
  NA_index=0
  k=1
  i=1
  while(i<=length(country))
  {
    if(is.na(x[i]))
    {
      NA_index[k]=i
      k=k+1
      while((i+1)<=length(country) && country[i+1]==country[i] && is.na(x[i+1]))
      {
        NA_index[k]=i+1
        k=k+1
        i=i+1
      }
    }
    while((i+1)<=length(country) &&
            country[i+1]==country[i])
    {
      i=i+1
    }
    i=i+1
  }
  return(NA_index)
}
#######################################
###interpolationfunction

NA.interpol=function(x){
```

```
  # Find first and last non NA entry

  if(sum(is.na(x))<length(x)){
    start=1
    while (is.na(x[start])){
      start=start+1}

    end=length(x)
    while (is.na(x[end])){
      end=end-1}

    # Interpolate linearly

    for (i in start:end){
      if (is.na(x[i])){
        if (!is.na(x[i-1])){
          tmp=1
          while (is.na(x[i+tmp])){tmp=tmp+1}
          if (tmp<4){
            interpol=seq(x[i-1],x[i+tmp],length.out=tmp+2)
            x[(i-1):(i+tmp)]=interpol
          }}}}}
  return(x)
}
####################
#mean function for vectors with NA values

mean2=function(x)
{
  return(mean(na.omit(x)))
}
##################

#function to group x by y

group=function(x,y)
{
  j=1
  z=list(NA)
  header=0
  for(i in 1:length(x))
  {
    if(!is.element(y[i],header))
    {
```

```
    index=which(y==y[i])
    z[[j]]=x[index]
    if(is.numeric(y[i]))
    {
      header[j]=y[i]
    }else{
      header[j]=as.character(y[i])
    }
    j=j+1
  }
}
#convert to matrix

f=0
length=lapply(z,length)
for(i in 1:length(z))
{
  f[i]=length[[i]]

}
nrow=max(f)
ncol=length(header)
m=matrix(NA, ncol=ncol, nrow=nrow)
colnames(m)=header
for(j in 1:ncol)
{
  for(i in 1:length[[j]])
  {
    m[i,j]=z[[j]][i]
  }
}
j=1
for(j in 1:length(m[1,])){ ###move NAs to the front
  l=length(which(is.na(m[ ,j])))
  if(l>0)
  {
    ll=l+1
    ld=length(m[ ,j])-l
    m[ll:length(m[ ,j]),j]=m[1:ld,j]
    m[1:l,j]=array(NA,l)
  }
}
result=list(m,header)
return(result)
```

```
}
```

A.2. Creating data set

```
#create dataset

dat=read.csv("C:\\Users\\Alex\\Desktop\\Bachelorarbeit\\
             globalization_2014.csv", header=TRUE, sep=";" )
##replace "." by "NA"
for(j in 1:12){
  y=which(dat[ ,j]==".")
  if(length(y)!=0){
    dat[y,j]=NA
  }
}
write.csv(dat,"C:\\Users\\Alex\\Desktop\\Bachelorarbeit\\
          globalization_2014_NA.csv")
dat=read.csv("C:\\Users\\Alex\\Desktop\\Bachelorarbeit\\
             globalization_2014_NA.csv", header=TRUE, dec="," )
dat=dat[ ,2:13]
#################
#add longitude/latitude

l=read.csv("C:\\Users\\Alex\\Desktop\\long2lat.csv",sep=";",
           header=FALSE)
#convert coordinates to numeric values (South and West correspond
#with "-", North and East correspond with "+" )
lat_dat=0
for(i in 1:length(l[ ,1])){
  a=as.character(l[i,2])
  if(nchar(a)>1){
    if(substring(a,nchar(a))=="S")
      b=-as.numeric(substring(a,1,nchar(a)-1))
    if(substring(a,nchar(a))=="N")
      b=as.numeric(substring(a,1,nchar(a)-1))
    lat_dat[i]=b
  }}
latitude_dat=data.frame(l[ ,1],lat_dat)
long_dat=0
for(i in 1:length(l[ ,1])){
  a=as.character(l[i,3])
  if(nchar(a)>1){
    if(substring(a,nchar(a))=="W")
```

```
      b=-as.numeric(substring(a,1,nchar(a)-1))
    if(substring(a,nchar(a))=="E")
      b=as.numeric(substring(a,1,nchar(a)-1))
    long_dat[i]=b
  }}
longitude_dat=data.frame(l[ ,1],long_dat)
## create longitude vector corresponding to order in dat
longitude_=rep(NA,length(dat[ ,1]))
x=as.character(longitude_dat[ ,1])
y=0
for(i in 1:length(x)){
  y[i]=x[i]
}
for( i in 1:length(dat[ ,1]))
{
  x= which(y==as.character(dat[i,1]))
  if(length(x)!=0){
    longitude_[i]= longitude_dat[x,2]
  }
}
## create latitude vector corresponding to order in dat
latitude_=rep(NA,length(dat[ ,1]))
x=as.character(latitude_dat[ ,1])
y=0
for(i in 1:length(x)){
  y[i]=x[i]
}
for( i in 1:length(dat[ ,1]))
{
  x= which(y==as.character(dat[i,1]))
  if(length(x)!=0){
    latitude_[i]= latitude_dat[x,2]
  }
}
##convert arc minutes
for( i in 1:length(longitude_))
  if(!is.na(longitude_[i])){
  if(longitude_[i]>=0)
  {
    digits=longitude_[i]-floor(longitude_[i])
    if((digits*100)%%10!=0){
      longitude_[i]=floor(longitude_[i])+(longitude_[i]-
                                      floor(longitude_[i]))*100/60}
    if((digits*100)%%10==0){
```

```
      longitude_[i]=floor(longitude_[i])+(longitude_[i]-
                                         floor(longitude_[i]))*10/60}
  }
if(longitude_[i]<0){
  longitude_[i]=-longitude_[i]
  digits=longitude_[i]-floor(longitude_[i])
  if((digits*100)%%10!=0){
    longitude_[i]=floor(longitude_[i])+(longitude_[i]-
                                       floor(longitude_[i]))*100/60}
  if((digits*100)%%10==0){
    longitude_[i]=floor(longitude_[i])+(longitude_[i]-
                                       floor(longitude_[i]))*10/60}
  longitude_[i]=-longitude_[i]}
}
if(!is.na(latitude_[i])){
if(latitude_[i]>=0)
{ digits=longitude_[i]-floor(longitude_[i])
  if((digits*100)%%10!=0){
    latitude_[i]=floor(latitude_[i])+(latitude_[i]-
                                     floor(latitude_[i]))*100/60}
  if((digits*100)%%10==0){
    latitude_[i]=floor(latitude_[i])+(latitude_[i]-
                                     floor(latitude_[i]))*10/60}
}
if(latitude_[i]<0){
  latitude_[i]=-latitude_[i]
  digits=longitude_[i]-floor(longitude_[i])
  if((digits*100)%%10!=0){
    latitude_[i]=floor(latitude_[i])+(latitude_[i]-
                                     floor(latitude_[i]))*100/60}
  if((digits*100)%%10==0){
    latitude_[i]=floor(latitude_[i])+(latitude_[i]-
                                     floor(latitude_[i]))*10/60}
  latitude_[i]=-latitude_[i]}
}
##################
#add birth_rate
birth_dat=read.csv("C:\\Users\\Alex\\Desktop\\Bachelorarbeit\\
                   birth rate\\birth rate_ges.csv", header=TRUE,
                   sep=",", dec=",")
birth=rep(NA,length(dat[ ,1]))
for( i in 1:length(dat[ ,1]))
{
  x= which(as.character(birth_dat[ ,2])==as.character(dat[i,2]))
```

```
  if(length(x)!=0){

    birth[i]= birth_dat[x,dat[i,3]-1955]

  }

}
###################
#eliminate leading NAs per country
dat_all=data.frame(dat[ ,1],dat[ ,3],dat[ ,4], dat[ ,7],
                   dat[ ,11], longitude_, latitude_, birth)
NA_index_1=NA.eliminate(dat[ ,1],dat_all[ ,3])
NA_index_2=NA.eliminate(dat[ ,1],dat_all[ ,4])
NA_index_3=NA.eliminate(dat[ ,1],dat_all[ ,5])
NA_index_4=NA.eliminate(dat[ ,1],dat_all[ ,8])
NA_index_a=union(NA_index_1,NA_index_2)
NA_index_b=union(NA_index_3,NA_index_4)
NA_index=union(NA_index_a,NA_index_b)
data=dat_all[setdiff(seq(1,length(dat_all[ ,1]),1),NA_index), ]
################################################
#3 NAs in birth_rate left -> interpolate
#NAs cannot be at beginning, so check if they are at the end:
which(is.na(data[ ,8]))
data[4906:4910, ]
###not at the end-> interpolation possible
data[4906:4910,8]=NA.interpol(data[4906:4910,8])
#get time averaged data
a=group(data[,3],data[,1])
index_econ_av=apply(a[[1]],2,mean2)
a=group(data[,4],data[,1])
index_social_av=apply(a[[1]],2,mean2)
a=group(data[,5],data[,1])
index_political_av=apply(a[[1]],2,mean2)
a=group(data[,6],data[,1])
longitude_av=apply(a[[1]],2,mean2)
a=group(data[,7],data[,1])
latitude_av=apply(a[[1]],2,mean2)
a=group(data[,8],data[,1])
birth_rate_av=apply(a[[1]],2,mean2)
data_avg=data.frame(as.factor(a[[2]]),index_econ_av,
                    index_social_av,index_political_av,
                    longitude_av,latitude_av,birth_rate_av)
head(data)
rownames(data)=seq(1,length(data[ ,1]),1)
```

```
#################
###add continent
africa_dat=read.csv("C:\\Users\\Alex\\Desktop\\Bachelorarbeit\\
                    Index_Africa.csv", header=TRUE, dec=".")
asia_dat=read.csv("C:\\Users\\Alex\\Desktop\\Bachelorarbeit\\
                  Index_Asia.csv", header=TRUE, dec=".")
europe_dat=read.csv("C:\\Users\\Alex\\Desktop\\Bachelorarbeit\\
                    Index_Europe.csv", header=TRUE, dec=".")
northAmerica_dat=read.csv("C:\\Users\\Alex\\Desktop\\
                          Bachelorarbeit\\Index_NorthAmerica.csv",
                          header=TRUE, dec=".")
southAmerica_dat=read.csv("C:\\Users\\Alex\\Desktop\\
                          Bachelorarbeit\\Index_SouthAmerica.csv",
                          header=TRUE, dec=".")
oceania_dat=read.csv("C:\\Users\\Alex\\Desktop\\Bachelorarbeit\\
                     Index_Oceania.csv", header=TRUE, dec=".")
cont=array( ,length(data[ ,2]))
for(i in 1:length(data[ ,4])){
  if(is.element(data[i,1],africa_dat[ ,2]))
  {
    cont[i]="africa"
  }

  if(is.element(data[i,1],asia_dat[ ,2]))
  {
    cont[i]="asia"
  }

  if(is.element(data[i,1],europe_dat[ ,2]))
  {
    cont[i]="europe"
  }

  if(is.element(data[i,1],northAmerica_dat[ ,2]))
  {
    cont[i]="northAmerica"
  }

  if(is.element(data[i,1],southAmerica_dat[ ,2]))
  {
    cont[i]="southAmerica"
  }

  if(is.element(data[i,1],oceania_dat[ ,2]))
```

```
  {
    cont[i]="oceania"
  }
}
cont=as.factor(cont)
data=data.frame(data,cont)
colnames(data)=c("country","year","index_econ", "index_social",
                 "index_political",  "longitude", "latitude",
                 "birth_rate","continent")
attach(data)
cont_avg=array( ,length(data_avg[ ,2]))
for(i in 1:length(data_avg[ ,4])){
  if(is.element(data_avg[i,1],africa_dat[ ,2]))
  {
    cont_avg[i]="africa"
  }

  if(is.element(data_avg[i,1],asia_dat[ ,2]))
  {
    cont_avg[i]="asia"
  }

  if(is.element(data_avg[i,1],europe_dat[ ,2]))
  {
    cont_avg[i]="europe"
  }

  if(is.element(data_avg[i,1],northAmerica_dat[ ,2]))
  {
    cont_avg[i]="northAmerica"
  }

  if(is.element(data_avg[i,1],southAmerica_dat[ ,2]))
  {
    cont_avg[i]="southAmerica"
  }

  if(is.element(data_avg[i,1],oceania_dat[ ,2]))
  {
    cont_avg[i]="oceania"
  }
}
cont_avg=as.factor(cont_avg)
data_avg=data.frame(data_avg,cont_avg)
```

```
rownames(data_avg)=seq(1,length(data_avg[ ,1]),1)
colnames(data_avg)=c("country_avg","index_econ_avg", "index_social_avg",
                     "index_political_avg",  "longitude_avg",
                     "latitude_avg", "birth_rate_avg","continent_avg")
########################################################
write.csv(data,"C:\\Users\\Alex\\Desktop\\Bachelorarbeit\\data.csv")
write.csv(data_avg,"C:\\Users\\Alex\\Desktop\\Bachelorarbeit\\
          data_avg.csv")


#read in data
data=read.csv("C:\\Users\\Alex\\Desktop\\Bachelorarbeit\\data.csv")
attach(data)
#read in time averaged data
data_avg=read.csv("C:\\Users\\Alex\\Desktop\\Bachelorarbeit\\
                  data_avg.csv")
colnames(data_avg)=c("number","country_avg","index_econ_avg",
                     "index_social_avg", "index_political_avg",
                     "longitude_avg","latitude_avg","birth_rate_avg",
                     "continent_avg")
attach(data_avg)
long1=cos(longitude/360*(2*pi))
long2=sin(longitude/360*(2*pi))
long1_avg=cos(longitude_avg/360*(2*pi))
long2_avg=sin(longitude_avg/360*(2*pi))
```

A.3. Chapter 2

```
#plot three indeces for Germany
pdf("C:/Users/Alex/Documents/pic1.pdf")
par(mfrow=c(3,1),mar=c(5,4,2,1))
index=which(country=="Germany")
x=seq(1970,2011,1)
plot(x,index_econ[index],type="l",
     ylab="economic index Germany",xlab="year")
plot(x,index_social[index],type="l",
     ylab="social index Germany",xlab="year")
plot(x,index_political[index],type="l",
     ylab="political index Germany",xlab="year")
dev.off()

#plot longitude on circle
pdf("C:/Users/Alex/Documents/pic2.pdf")
plot(cos(longitude*2*pi/360),sin(longitude*2*pi/360), xlab="Lon.c1",
```

```
     ylab="Lon.c2")
index=which(country=="Germany")
points(cos(longitude[index]*2*pi/360),sin(longitude[index]*2*pi/360),
       col="red",xlab="",ylab="")
index=which(country=="United States")
points(cos(longitude[index]*2*pi/360),sin(longitude[index]*2*pi/360),
       col="green",xlab="",ylab="")
index=which(country=="Japan")
points(cos(longitude[index]*2*pi/360),sin(longitude[index]*2*pi/360),
       col="blue",xlab="",ylab="")
abline(c(0,1),c(0,0),col="blue")
dev.off

#plot available data
pdf("C:/Users/Alex/Documents/pic3.pdf")
g=group(year,country)
m=g[[1]]
s=rep(0,42)
plot(m[ ,1],s,ylim=c(0,151),type="l",xlab="year",ylab="country")
for(i in 1:151)
{
lines(m[ ,i],rep(i,42))
}
dev.off
```

A.4. Chapter 3

```
#plot dependancies among variables
library('scatterplot3d')
pdf("C:/Users/Alex/Documents/pic4.pdf")
par(mfrow=c(3,3),mar=c(2,2,1,1))
hist(longitude, main="", ylab="",xlab="")
box("figure")
lm=lm(latitude~long1+long2)
s=scatterplot3d(long1,long2,latitude, main="",
               ylab="",xlab="", zlab="",x.ticklab=c(-1,"",0,"",1),
               y.ticklab=c(-1,"",0,"",1),
               z.ticklab=c("",-40,"",0,"",40,"",80),mar=c(2,2,1,1))
s$plane3d(lm, col="red")
box("figure")
lm=lm(birth_rate~long1+long2)
s=scatterplot3d(long1,long2,birth_rate, main="",
               ylab="",xlab="", zlab="",x.ticklab=c(-1,"",0,"",1),
```

```
                y.ticklab=c(-1,"",0,"",1),
                z.ticklab=c("",10,"",30,"",50,""),mar=c(2,2,1,1))
s$plane3d(lm, col="red")
box("figure")
lm=lm(latitude_avg~long1_avg+long2_avg)
s=scatterplot3d(long1_avg,long2_avg,latitude_avg, main="",
                ylab="",xlab="",zlab="",x.ticklab=c(-1,"",0,"",1),
                y.ticklab=c(-1,"",0,"",1),
                z.ticklab=c("",-40,"",0,"",40,"",80),mar=c(2,2,1,1))
s$plane3d(lm, col="red")
box("figure")
hist(latitude, main="", ylab="",xlab="")
box("figure")
plot(birth_rate,latitude, main="", ylab="",xlab="")
lines(lowess(birth_rate,latitude), col="red")
box("figure")
lm=lm(birth_rate_avg~long1_avg+long2_avg)
s=scatterplot3d(long1_avg,long2_avg,birth_rate_avg, main="",
                ylab="",xlab="",zlab="",x.ticklab=c(-1,"",0,"",1),
                y.ticklab=c(-1,"",0,"",1),
                z.ticklab=c("",10,"",30,"",50,""),mar=c(2,2,1,1))
s$plane3d(lm, col="red")
box("figure")
plot(birth_rate_avg,latitude_avg, main="", ylab="",xlab="")
lines(lowess(birth_rate_avg,latitude_avg), col="red")
box("figure")
hist(birth_rate, main="", ylab="",xlab="")
box("figure")
dev.off

#plot covariate-continent
pdf("C:/Users/Alex/Documents/pic5.pdf",width=9,height=3)
par(mfrow=c(1,3), mar=c(3,2,1,1))
plot(continent,latitude,xaxt="n")
axis(1,at=c(1,2,3,4,5,6),labels=c("A","AS","E","N","O","S"))
box("figure")
plot(continent,birth_rate,xaxt="n")
axis(1,at=c(1,2,3,4,5,6),labels=c("A","AS","E","N","O","S"))
box("figure")
plot(continent_avg,birth_rate_avg,xaxt="n")
axis(1,at=c(1,2,3,4,5,6),labels=c("A","AS","E","N","O","S"))
box("figure")
dev.off

#plot longitude and mark continents
```

```
pdf("C:/Users/Alex/Documents/pic6.pdf")
plot(long1[which(continent=="asia")],long2[which(continent=="asia")],
     xlim=c(-1,1),ylim=c(-1,1),xlab="Lon.c1",ylab="Lon.c2")
points(0.9*long1[which(continent=="africa")],0.9*
         long2[which(continent=="africa")],col="blue",
       xlim=c(-1,1),ylim=c(-1,1))
points(0.8*long1[which(continent=="europe")],0.8*
         long2[which(continent=="europe")],col="red",
       xlim=c(-1,1),ylim=c(-1,1))
points(0.9*long1[which(continent=="northAmerica")],0.9*
         long2[which(continent=="northAmerica")],col="green",
       xlim=c(-1,1),ylim=c(-1,1))
points(0.9*long1[which(continent=="oceania")],0.9*
         long2[which(continent=="oceania")],col="cyan",
       xlim=c(-1,1),ylim=c(-1,1))
points(long1[which(continent=="southAmerica")],
       long2[which(continent=="southAmerica")],col="yellow",
       xlim=c(-1,1),ylim=c(-1,1))
dev.off

#plot time vs year
pdf("C:/Users/Alex/Documents/pic7.pdf")
par(mfrow=c(1,1), mar=c(3,2,2,1))
plot(year, birth_rate, xlab="T", ylab="Birth")
lines(lowess(year,birth_rate), col="red")
dev.off

#plots: covariate vs rsponse

pdf("C:/Users/Alex/Documents/pic8.pdf", height=12, width=9)
mar=c(2,2,0,0)
par(mfrow=c(4,3), mar=mar)
library("scatterplot3d")
a=lm(index_econ~sin(longitude/360*2*pi)+cos(longitude/360*2*pi))
s=scatterplot3d(cos(longitude/360*2*pi), sin(longitude/360*2*pi),
                birth_rate,type="p", xlab="", ylab="", zlab="", mar=mar,
                zlim=c(0,100),x.ticklab=c(-1,"",0,"",1),
                y.ticklab=c(-1,"",0,"",1))
s$plane3d(a,col="red")
box("figure")
a=lm(index_social~sin(longitude/360*2*pi)+cos(longitude/360*2*pi))
s=scatterplot3d(cos(longitude/360*2*pi), sin(longitude/360*2*pi),
                birth_rate,type="p",xlab="", ylab="", zlab="", mar=mar,
                zlim=c(0,100),x.ticklab=c(-1,"",0,"",1),
```

```
                y.ticklab=c(-1,"",0,"",1))
s$plane3d(a,col="red")
box("figure")
a=lm(index_political~sin(longitude/360*2*pi)+cos(longitude/360*2*pi))
s=scatterplot3d(cos(longitude/360*2*pi), sin(longitude/360*2*pi),
                birth_rate,type="p",xlab="", ylab="", zlab="", mar=mar,
                zlim=c(0,100),x.ticklab=c(-1,"",0,"",1),
                y.ticklab=c(-1,"",0,"",1))
s$plane3d(a,col="red")
box("figure")
plot(latitude,index_econ, yaxt="n")
axis(2, at=c(20,40,60,80,""))
lines(lowess(latitude,index_econ), col="red")
box("figure")
plot(latitude,index_social, yaxt="n", ylim=c(0,100))
axis(2, at=c(20,40,60,80,""))
lines(lowess(latitude,index_social), col="red")
box("figure")
plot(latitude,index_political, yaxt="n", ylim=c(0,100))
axis(2, at=c(20,40,60,80,""))
lines(lowess(latitude,index_political), col="red")
box("figure")
plot(birth_rate,index_econ, yaxt="n", ylim=c(0,100))
axis(2, at=c(20,40,60,80,""))
lines(lowess(birth_rate,index_econ), col="red")
box("figure")
plot(birth_rate,index_social, yaxt="n", ylim=c(0,100))
axis(2, at=c(20,40,60,80,""))
lines(lowess(birth_rate,index_social), col="red")
box("figure")
plot(birth_rate,index_political, yaxt="n", ylim=c(0,100))
axis(2, at=c(20,40,60,80,""))
lines(lowess(birth_rate,index_political), col="red")
box("figure")
plot(year,index_econ, yaxt="n", ylim=c(0,100))
axis(2, at=c(20,40,60,80,""))
lines(lowess(year,index_econ), col="red")
box("figure")
plot(year,index_social, yaxt="n", ylim=c(0,100))
axis(2, at=c(20,40,60,80,""))
lines(lowess(year,index_social), col="red")
box("figure")
plot(year,index_political, yaxt="n", ylim=c(0,100))
axis(2, at=c(20,40,60,80,""))
```

```
lines(lowess(year,index_political), col="red")
box("figure")
dev.off()

#plots: covariate vs response, averaged

pdf("C:/Users/Alex/Documents/pic9.pdf")
mar=c(2,2,0,0)
par(mfrow=c(3,3), mar=mar)
a=lm(index_econ_avg~long1_avg+long2_avg)
s=scatterplot3d(long1_avg,long2_avg, birth_rate_avg,type="p", xlab="",
                ylab="", zlab="", mar=mar, zlim=c(0,100),
                x.ticklab=c(-1,"",0,"",1),y.ticklab=c(-1,"",0,"",1))
s$plane3d(a,col="red")
box("figure")
a=lm(index_social_avg~long1_avg+long2_avg)
s=scatterplot3d(long1_avg, long2_avg, birth_rate_avg,type="p",xlab="",
                ylab="", zlab="", mar=mar, zlim=c(0,100),
                x.ticklab=c(-1,"",0,"",1),y.ticklab=c(-1,"",0,"",1))
s$plane3d(a,col="red")
box("figure")
a=lm(index_political_avg~long1_avg+long2_avg)
s=scatterplot3d(long1_avg, long2_avg, birth_rate_avg,type="p",xlab="",
                ylab="", zlab="", mar=mar, zlim=c(0,100),
                x.ticklab=c(-1,"",0,"",1),y.ticklab=c(-1,"",0,"",1))
s$plane3d(a,col="red")
box("figure")
plot(latitude_avg,index_econ_avg, yaxt="n")
axis(2, at=c(20,40,60,80,""))
lines(lowess(latitude_avg,index_econ_avg), col="red")
box("figure")
plot(latitude_avg,index_social_avg, yaxt="n", ylim=c(0,100))
axis(2, at=c(20,40,60,80,""))
lines(lowess(latitude_avg,index_social_avg), col="red")
box("figure")
plot(latitude_avg,index_political_avg, yaxt="n", ylim=c(0,100))
axis(2, at=c(20,40,60,80,""))
lines(lowess(latitude_avg,index_political_avg), col="red")
box("figure")
plot(birth_rate_avg,index_econ_avg, yaxt="n", ylim=c(0,100))
axis(2, at=c(20,40,60,80,""))
lines(lowess(birth_rate_avg,index_econ_avg), col="red")
box("figure")
plot(birth_rate_avg,index_social_avg, yaxt="n", ylim=c(0,100))
```

```
axis(2, at=c(20,40,60,80,""))
lines(lowess(birth_rate_avg,index_social_avg), col="red")
box("figure")
plot(birth_rate_avg,index_political_avg, yaxt="n", ylim=c(0,100))
axis(2, at=c(20,40,60,80,""))
lines(lowess(birth_rate_avg,index_political_avg), col="red")
box("figure")

#plot index per continent
pdf("C:/Users/Alex/Documents/pic10.pdf",width=9,height=3)
par(mfrow=c(1,3),mar=c(3,2,0,0))
plot(continent,index_econ,xaxt="n",ylab="Y_1jt", yaxt="n")
axis(1,at=c(1,2,3,4,5,6),labels=c("A","AS","E","N","O","S"))
axis(2,at=c(20,40,60,80))
plot(continent,index_social,xaxt="n",ylab="Y_2jt", yaxt="n")
axis(1,at=c(1,2,3,4,5,6),labels=c("A","AS","E","N","O","S"))
axis(2,at=c(20,40,60,80))
plot(continent,index_political,xaxt="n",ylab="Y_3jt", yaxt="n")
axis(1,at=c(1,2,3,4,5,6),labels=c("A","AS","E","N","O","S"))
axis(2,at=c(20,40,60,80))
dev.off()

#plot index per continent, averaged
pdf("C:/Users/Alex/Documents/pic11.pdf",width=9,height=3)
par(mfrow=c(1,3),mar=c(3,2,0,0))
plot(continent_avg,index_econ_avg,xaxt="n",ylab="Y_1jt", yaxt="n")
axis(1,at=c(1,2,3,4,5,6),labels=c("A","AS","E","N","O","S"))
axis(2,at=c(20,40,60,80))
plot(continent_avg,index_social_avg,xaxt="n",ylab="Y_2jt", yaxt="n")
axis(1,at=c(1,2,3,4,5,6),labels=c("A","AS","E","N","O","S"))
axis(2,at=c(20,40,60,80))
plot(continent_avg,index_political_avg,xaxt="n",ylab="Y_3jt", yaxt="n")
axis(1,at=c(1,2,3,4,5,6),labels=c("A","AS","E","N","O","S"))
axis(2,at=c(20,40,60,80))
dev.off()

#interaction: longitude,latitude

pdf("C:/Users/Alex/Documents/pic99.pdf")
mar=c(2,2,0,0)
par(mfrow=c(3,3),mar=mar)
index_low=which(latitude_avg<quantile(latitude_avg,0.25))
index_medlat1=which(latitude_avg>=quantile(latitude_avg,0.25))
index_medlat2=which(latitude_avg<=quantile(latitude_avg,0.75))
```

```
index_med=intersect(index_medlat1,index_medlat2)
index_high=which(latitude_avg>quantile(latitude_avg,0.75))
l_econ_low=lm(index_econ_avg[index_low]~long1_avg[index_low]+
                long2_avg[index_low])
l_econ_med=lm(index_econ_avg[index_med]~long1_avg[index_med]+
                long2_avg[index_med])
l_econ_high=lm(index_econ_avg[index_high]~long1_avg[index_high]+
                 long2_avg[index_high])
#get mean prediction interval, econ index
p=predict.lm(l_econ_low,interval='confidence')
mean(p[,3]-p[,2])
p=predict.lm(l_econ_med,interval='confidence')
mean(p[,3]-p[,2])
p=predict.lm(l_econ_high,interval='confidence')
mean(p[,3]-p[,2])
pred_econ_low=predict(l_econ_low,interval='confidence' )
mean(pred_econ_low[ ,3]-pred_econ_low[ ,2])
#interactionplot, econ
s=scatterplot3d(0, 0, 0, highlight.3d = TRUE, pch = 20,xlim=c(-1,1),
                ylim=c(-1,1),zlim=c(0,100),mar=mar, xlab="", ylab="",
                zlab="",x.ticklab=c(-1,"",0,"",1),
                y.ticklab=c(-1,"",0,"",1))
s$plane3d(l_econ_low,col="green",lty="dashed",lty.box="solid")
s$plane3d(l_econ_med,col="red",lty.box="solid")
s$plane3d(l_econ_high,col="blue",lty.box="solid")
box("figure")
#interactionplot, social
l_social_low=lm(index_social_avg[index_low]~long1_avg[index_low]+
                  long2_avg[index_low])
l_social_med=lm(index_social_avg[index_med]~long1_avg[index_med]+
                  long2_avg[index_med])
l_social_high=lm(index_social_avg[index_high]~long1_avg[index_high]+
                   long2_avg[index_high])
s=scatterplot3d(0, 0, 0, highlight.3d = TRUE, pch = 20,xlim=c(-1,1),
                ylim=c(-1,1),zlim=c(0,100),mar=mar, xlab="", ylab="",
                zlab="",x.ticklab=c(-1,"",0,"",1),
                y.ticklab=c(-1,"",0,"",1))
s$plane3d(l_social_low,col="green",lty="dashed",lty.box="solid")
s$plane3d(l_social_med,col="red",lty.box="solid")
s$plane3d(l_social_high,col="blue",lty.box="solid")
box("figure")
#get mean prediction interval, social
pred_econ_low=predict(l_econ_low,interval='confidence' )
mean(pred_econ_low[ ,3]-pred_econ_low[ ,2])
```

```
#interactionplot, plotical
l_political_low=lm(index_political_avg[index_low]~long1_avg[index_low]+
                     long2_avg[index_low])
l_political_med=lm(index_political_avg[index_med]~long1_avg[index_med]+
                     long2_avg[index_med])
l_political_high=lm(index_political_avg[index_high]~long1_avg[index_high]+
                      long2_avg[index_high])
s=scatterplot3d(0, 0, 0, highlight.3d = TRUE, pch = 20,xlim=c(-1,1),
                ylim=c(-1,1),zlim=c(0,100),mar=mar, xlab="", ylab="",
                zlab="",x.ticklab=c(-1,"",0,"",1),
                y.ticklab=c(-1,"",0,"",1))
s$plane3d(l_political_low,col="green",lty="dashed",lty.box="solid")
s$plane3d(l_political_med,col="red",lty.box="solid")
s$plane3d(l_political_high,col="blue",lty.box="solid")
box("figure")
######################
#interaction: longitude,birth_rate
index_low=which(birth_rate_avg<quantile(birth_rate_avg,0.25))
index_medlat1=which(birth_rate_avg>=quantile(birth_rate_avg,0.25))
index_medlat2=which(birth_rate_avg<=quantile(birth_rate_avg,0.75))
index_med=intersect(index_medlat1,index_medlat2)
index_high=which(birth_rate_avg>quantile(birth_rate_avg,0.75))
#interactionplot, econ
l_econ_low=lm(index_econ_avg[index_low]~long1_avg[index_low]+
                long2_avg[index_low])
l_econ_med=lm(index_econ_avg[index_med]~long1_avg[index_med]+
                long2_avg[index_med])
l_econ_high=lm(index_econ_avg[index_high]~long1_avg[index_high]+
                 long2_avg[index_high])
s=scatterplot3d(0, 0, 0, highlight.3d = TRUE, pch = 20,xlim=c(-1,1),
                ylim=c(-1,1),zlim=c(0,100),mar=mar, xlab="", ylab="",
                zlab="",x.ticklab=c(-1,"",0,"",1),
                y.ticklab=c(-1,"",0,"",1))
s$plane3d(l_econ_low,col="green",lty="dashed",lty.box="solid")
s$plane3d(l_econ_med,col="red",lty.box="solid")
s$plane3d(l_econ_high,col="blue",lty.box="solid")
box("figure")
#interactionplot, social
l_social_low=lm(index_social_avg[index_low]~long1_avg[index_low]+
                  long2_avg[index_low])
l_social_med=lm(index_social_avg[index_med]~long1_avg[index_med]+
                  long2_avg[index_med])
l_social_high=lm(index_social_avg[index_high]~long1_avg[index_high]
                   +long2_avg[index_high])
```

```
s=scatterplot3d(0, 0, 0, highlight.3d = TRUE, pch = 20,xlim=c(-1,1),
                ylim=c(-1,1),zlim=c(0,100),mar=mar, xlab="", ylab="",
                zlab="",x.ticklab=c(-1,"",0,"",1),
                y.ticklab=c(-1,"",0,"",1))
s$plane3d(l_social_low,col="green",lty="dashed",lty.box="solid")
s$plane3d(l_social_med,col="red",lty.box="solid")
s$plane3d(l_social_high,col="blue",lty.box="solid")
box("figure")
#interactionplot, political
l_political_low=lm(index_political_avg[index_low]~
                     cos(longitude[index_low]/360*2*pi)+
                     long2_avg[index_low])
l_political_med=lm(index_political_avg[index_med]~
                     cos(longitude[index_med]/360*2*pi)+
                     long2_avg[index_med])
l_political_high=lm(index_political_avg[index_high]~
                      cos(longitude[index_high]/360*2*pi)+
                      long2_avg[index_high])
s=scatterplot3d(0, 0, 0, highlight.3d = TRUE, pch = 20,xlim=c(-1,1),
                ylim=c(-1,1),zlim=c(0,100),mar=mar, xlab="", ylab="",
                zlab="", x.ticklab=c(-1,"",0,"",1),
                y.ticklab=c(-1,"",0,"",1))
s$plane3d(l_political_low,col="green",lty="dashed",lty.box="solid")
s$plane3d(l_political_med,col="red",lty.box="solid")
s$plane3d(l_political_high,col="blue",lty.box="solid")
box("figure")
################
#interaction: birth_rate,latitude

index_low=which(birth_rate_avg<quantile(birth_rate_avg,0.25))
index_medlat1=which(birth_rate_avg>=quantile(birth_rate_avg,0.25))
index_medlat2=which(birth_rate_avg<=quantile(birth_rate_avg,0.75))
index_med=intersect(index_medlat1,index_medlat2)
index_high=which(birth_rate_avg>quantile(birth_rate_avg,0.75))
l_econ_low=lm(index_econ_avg[index_low]~latitude_avg[index_low])
l_econ_med=lm(index_econ_avg[index_med]~latitude_avg[index_med])
l_econ_high=lm(index_econ_avg[index_high]~latitude_avg[index_high])
#get mean predition interval, econ
pred_econ_low=predict(l_econ_low,interval='confidence' )
pred_econ_med=predict(l_econ_med,interval='confidence' )
pred_econ_high=predict(l_econ_high,interval='confidence' )
mean(pred_econ_low[ ,3]-pred_econ_low[ ,2])
mean(pred_econ_med[ ,3]-pred_econ_med[ ,2])
mean(pred_econ_high[ ,3]-pred_econ_high[ ,2])
```

```
#interactionplot, econ
plot(0,0,xlim=c(10,50),ylim=c(0,100), yaxt="n")
abline(l_econ_low,col="green")
abline(l_econ_med,col="red")
abline(l_econ_high,col="blue")
axis(2, at=c(0,20,40,60,80,""))
box("figure")
#interacionplot, social
l_social_low=lm(index_social_avg[index_low]~latitude_avg[index_low])
l_social_med=lm(index_social_avg[index_med]~latitude_avg[index_med])
l_social_high=lm(index_social_avg[index_high]~
                   latitude_avg[index_high])
plot(0,0,xlim=c(10,50),ylim=c(0,100), yaxt="n")
abline(l_social_low,col="green")
abline(l_social_med,col="red")
abline(l_social_high,col="blue")
axis(2, at=c(0,20,40,60,80,""))
box("figure")
#interactionplot, political
l_political_low=lm(index_political_avg[index_low]~
                     latitude_avg[index_low])
l_political_med=lm(index_political_avg[index_med]~
                     latitude_avg[index_med])
l_political_high=lm(index_political_avg[index_high]~
                      latitude_avg[index_high])
plot(0,0,xlim=c(10,50),ylim=c(0,100), yaxt="n")
abline(l_political_low,col="green")
abline(l_political_med,col="red")
abline(l_political_high,col="blue")
axis(2, at=c(0,20,40,60,80,""))
box("figure")
dev.off()

##################################

pdf("C:/Users/Alex/Documents/pic999.pdf")
mar=c(2,2,0,0)
par(mfrow=c(3,3),mar=mar)
af=which(continent_avg=="africa")
as=which(continent_avg=="asia")
eu=which(continent_avg=="europe")
na=which(continent_avg=="northAmerica")
oc=which(continent_avg=="oceania")
sa=which(continent_avg=="southAmerica")
```

```
#interaction: continent, longitude

l_econ_af=lm(index_econ_avg[af]~long1_avg[af]+long2_avg[af])
l_econ_as=lm(index_econ_avg[as]~long1_avg[as]+long2_avg[as])
l_econ_eu=lm(index_econ_avg[eu]~long1_avg[eu]+long2_avg[eu])
l_econ_na=lm(index_econ_avg[na]~long1_avg[na]+long2_avg[na])
l_econ_oc=lm(index_econ_avg[oc]~long1_avg[oc]+long2_avg[oc])
l_econ_sa=lm(index_econ_avg[sa]~long1_avg[sa]+long2_avg[sa])
#getting mean prediction intervals
p=predict.lm(l_econ_af, interval='confidence')
mean(p[,3]-p[,2])
p=predict.lm(l_econ_as, interval='confidence')
mean(p[,3]-p[,2])
p=predict.lm(l_econ_eu, interval='confidence')
mean(p[,3]-p[,2])
p=predict.lm(l_econ_na, interval='confidence')
mean(p[,3]-p[,2])
p=predict.lm(l_econ_oc, interval='confidence')
mean(p[,3]-p[,2])
p=predict.lm(l_econ_sa, interval='confidence')
mean(p[,3]-p[,2])
#interaction plots, econ
s=scatterplot3d(0, 0, 0, highlight.3d = TRUE, pch = 20,xlim=c(-1,1),
                ylim=c(-1,1),zlim=c(0,100),mar=mar, xlab="", ylab="",
                zlab="",x.ticklab=c(-1,"",0,"",1),
                y.ticklab=c(-1,"",0,"",1))
s$plane3d(l_econ_af,col="blue",lty="dashed",lty.box="solid")
s$plane3d(l_econ_as,lty.box="solid")
s$plane3d(l_econ_eu,col="red",lty.box="solid")
s$plane3d(l_econ_na,col="green",lty="dashed",lty.box="solid")
s$plane3d(l_econ_oc,col="cyan",lty.box="solid")
s$plane3d(l_econ_sa,col="yellow",lty.box="solid")
box("figure")
##############

l_social_af=lm(index_social_avg[af]~long1_avg[af]+long2_avg[af])
l_social_as=lm(index_social_avg[as]~long1_avg[as]+long2_avg[as])
l_social_eu=lm(index_social_avg[eu]~long1_avg[eu]+long2_avg[eu])
l_social_na=lm(index_social_avg[na]~long1_avg[na]+long2_avg[na])
l_social_oc=lm(index_social_avg[oc]~long1_avg[oc]+long2_avg[oc])
l_social_sa=lm(index_social_avg[sa]~long1_avg[sa]+long2_avg[sa])
#getting mean prediction intervals, social
p=predict.lm(l_social_af, interval='confidence')
mean(p[,3]-p[,2])
```

```
p=predict.lm(l_social_as, interval='confidence')
mean(p[,3]-p[,2])
p=predict.lm(l_social_eu, interval='confidence')
mean(p[,3]-p[,2])
p=predict.lm(l_social_na, interval='confidence')
mean(p[,3]-p[,2])
p=predict.lm(l_social_oc, interval='confidence')
mean(p[,3]-p[,2])
p=predict.lm(l_social_sa, interval='confidence')
mean(p[,3]-p[,2])
#interaction plots, social
s=scatterplot3d(0, 0, 0, highlight.3d = TRUE, pch = 20,xlim=c(-1,1),
                ylim=c(-1,1),zlim=c(0,100),mar=mar, xlab="", ylab="",
                zlab="",x.ticklab=c(-1,"",0,"",1),
                y.ticklab=c(-1,"",0,"",1))
s$plane3d(l_social_af,col="blue",lty="dashed",lty.box="solid")
s$plane3d(l_social_as,lty.box="solid")
s$plane3d(l_social_eu,col="red",lty.box="solid")
s$plane3d(l_social_na,col="green",lty="dashed",lty.box="solid")
s$plane3d(l_social_oc,col="cyan",lty.box="solid")
s$plane3d(l_social_sa,col="yellow",lty.box="solid")
box("figure")
##############

l_political_af=lm(index_political_avg[af]~long1_avg[af]+long2_avg[af])
l_political_as=lm(index_political_avg[as]~long1_avg[as]+long2_avg[as])
l_political_eu=lm(index_political_avg[eu]~long1_avg[eu]+long2_avg[eu])
l_political_na=lm(index_political_avg[na]~long1_avg[na]+long2_avg[na])
l_political_oc=lm(index_political_avg[oc]~long1_avg[oc]+long2_avg[oc])
l_political_sa=lm(index_political_avg[sa]~long1_avg[sa]+long2_avg[sa])
#getting mean prediction intervals, political
p=predict.lm(l_political_af, interval='confidence')
mean(p[,3]-p[,2])
p=predict.lm(l_political_as, interval='confidence')
mean(p[,3]-p[,2])
p=predict.lm(l_political_eu, interval='confidence')
mean(p[,3]-p[,2])
p=predict.lm(l_political_na, interval='confidence')
mean(p[,3]-p[,2])
p=predict.lm(l_political_oc, interval='confidence')
mean(p[,3]-p[,2])
p=predict.lm(l_political_sa, interval='confidence')
mean(p[,3]-p[,2])
#interaction plots, political
```

```
s=scatterplot3d(0, 0, 0, highlight.3d = TRUE, pch = 20,xlim=c(-1,1),
                ylim=c(-1,1),zlim=c(0,100),mar=mar, xlab="", ylab="",
                zlab="",x.ticklab=c(-1,"",0,"",1),
                y.ticklab=c(-1,"",0,"",1))
s$plane3d(l_political_af,col="blue",lty="dashed",lty.box="solid")
s$plane3d(l_political_as,lty.box="solid")
s$plane3d(l_political_eu,col="red",lty.box="solid")
s$plane3d(l_political_na,col="green",lty="dashed",lty.box="solid")
s$plane3d(l_political_oc,col="cyan",lty.box="solid")
s$plane3d(l_political_sa,col="yellow",lty.box="solid")
box("figure")
#################
#interaction: continent, latitude

l_econ_af=lm(index_econ_avg[af]~latitude_avg[af])
l_econ_as=lm(index_econ_avg[as]~latitude_avg[as])
l_econ_eu=lm(index_econ_avg[eu]~latitude_avg[eu])
l_econ_na=lm(index_econ_avg[na]~latitude_avg[na])
l_econ_oc=lm(index_econ_avg[oc]~latitude_avg[oc])
l_econ_sa=lm(index_econ_avg[sa]~latitude_avg[sa])
#getting mean prediction intervals, econ
p=predict.lm(l_econ_af, interval='confidence')
mean(p[,3]-p[,2])
p=predict.lm(l_econ_as, interval='confidence')
mean(p[,3]-p[,2])
p=predict.lm(l_econ_eu, interval='confidence')
mean(p[,3]-p[,2])
p=predict.lm(l_econ_na, interval='confidence')
mean(p[,3]-p[,2])
p=predict.lm(l_econ_oc, interval='confidence')
mean(p[,3]-p[,2])
p=predict.lm(l_econ_sa, interval='confidence')
mean(p[,3]-p[,2])
#interaction plots, econ
plot(0,0,xlim=c(-40,65),ylim=c(0,100), yaxt="n")
axis(2, at=c(0,20,40,60,80,""))
abline(l_econ_af,col="blue")
abline(l_econ_as)
abline(l_econ_eu,col="red")
abline(l_econ_na,col="green")
abline(l_econ_oc,col="cyan")
abline(l_econ_sa,col="yellow")
box("figure")
##############
```

```
l_social_af=lm(index_social_avg[af]~latitude_avg[af])
l_social_as=lm(index_social_avg[as]~latitude_avg[as])
l_social_eu=lm(index_social_avg[eu]~latitude_avg[eu])
l_social_na=lm(index_social_avg[na]~latitude_avg[na])
l_social_oc=lm(index_social_avg[oc]~latitude_avg[oc])
l_social_sa=lm(index_social_avg[sa]~latitude_avg[sa])
#getting mean prediction intervals, social
p=predict.lm(l_social_af, interval='confidence')
mean(p[,3]-p[,2])
p=predict.lm(l_social_as, interval='confidence')
mean(p[,3]-p[,2])
p=predict.lm(l_social_eu, interval='confidence')
mean(p[,3]-p[,2])
p=predict.lm(l_social_na, interval='confidence')
mean(p[,3]-p[,2])
p=predict.lm(l_social_oc, interval='confidence')
mean(p[,3]-p[,2])
p=predict.lm(l_social_sa, interval='confidence')
mean(p[,3]-p[,2])
#interaction plots, social
plot(0,0,xlim=c(-40,65),ylim=c(0,100), yaxt="n")
axis(2, at=c(0,20,40,60,80,""))
abline(l_social_af,col="blue")
abline(l_social_as)
abline(l_social_eu,col="red")
abline(l_social_na,col="green")
abline(l_social_oc,col="cyan")
abline(l_social_sa,col="yellow")
box("figure")
##############

l_political_af=lm(index_political_avg[af]~latitude_avg[af])
l_political_as=lm(index_political_avg[as]~latitude_avg[as])
l_political_eu=lm(index_political_avg[eu]~latitude_avg[eu])
l_political_na=lm(index_political_avg[na]~latitude_avg[na])
l_political_oc=lm(index_political_avg[oc]~latitude_avg[oc])
l_political_sa=lm(index_political_avg[sa]~latitude_avg[sa])
#getting mean prediction intervals, political
p=predict.lm(l_political_af, interval='confidence')
mean(p[,3]-p[,2])
p=predict.lm(l_political_as, interval='confidence')
mean(p[,3]-p[,2])
p=predict.lm(l_political_eu, interval='confidence')
```

```
mean(p[,3]-p[,2])
p=predict.lm(l_political_na, interval='confidence')
mean(p[,3]-p[,2])
p=predict.lm(l_political_oc, interval='confidence')
mean(p[,3]-p[,2])
p=predict.lm(l_political_sa, interval='confidence')
mean(p[,3]-p[,2])
#interaction plots, political
plot(0,0,xlim=c(-40,65),ylim=c(0,100), yaxt="n")
axis(2, at=c(0,20,40,60,80,""))
abline(l_political_af,col="blue")
abline(l_political_as)
abline(l_political_eu,col="red")
abline(l_political_na,col="green")
abline(l_political_oc,col="cyan")
abline(l_political_sa,col="yellow")
box("figure")
##################
#interaction: continent, birth
l_econ_af=lm(index_econ_avg[af]~birth_rate_avg[af])
l_econ_as=lm(index_econ_avg[as]~birth_rate_avg[as])
l_econ_eu=lm(index_econ_avg[eu]~birth_rate_avg[eu])
l_econ_na=lm(index_econ_avg[na]~birth_rate_avg[na])
l_econ_oc=lm(index_econ_avg[oc]~birth_rate_avg[oc])
l_econ_sa=lm(index_econ_avg[sa]~birth_rate_avg[sa])
#interaction plots, econ
plot(0,0,xlim=c(10,50),ylim=c(0,100), yaxt="n")
axis(2, at=c(0,20,40,60,80,""))
abline(l_econ_af,col="blue")
abline(l_econ_as)
abline(l_econ_eu,col="red")
abline(l_econ_na,col="green")
abline(l_econ_oc,col="cyan")
abline(l_econ_sa,col="yellow")
box("figure")
##############

l_social_af=lm(index_social_avg[af]~birth_rate_avg[af])
l_social_as=lm(index_social_avg[as]~birth_rate_avg[as])
l_social_eu=lm(index_social_avg[eu]~birth_rate_avg[eu])
l_social_na=lm(index_social_avg[na]~birth_rate_avg[na])
l_social_oc=lm(index_social_avg[oc]~birth_rate_avg[oc])
l_social_sa=lm(index_social_avg[sa]~birth_rate_avg[sa])
#interaction plots, social
```

```
plot(0,0,xlim=c(10,50),ylim=c(0,100), yaxt="n")
axis(2, at=c(0,20,40,60,80,""))
abline(l_social_af,col="blue")
abline(l_social_as)
abline(l_social_eu,col="red")
abline(l_social_na,col="green")
abline(l_social_oc,col="cyan")
abline(l_social_sa,col="yellow")
box("figure")
##############

l_political_af=lm(index_political_avg[af]~birth_rate_avg[af])
l_political_as=lm(index_political_avg[as]~birth_rate_avg[as])
l_political_eu=lm(index_political_avg[eu]~birth_rate_avg[eu])
l_political_na=lm(index_political_avg[na]~birth_rate_avg[na])
l_political_oc=lm(index_political_avg[oc]~birth_rate_avg[oc])
l_political_sa=lm(index_political_avg[sa]~birth_rate_avg[sa])
#mean prediction interval for oceania
p=predict.lm(l_political_oc, interval='confidence')
mean(p[ ,3]-p[,2])
#interaction plots, political
plot(0,0,xlim=c(10,50),ylim=c(0,100), yaxt="n")
axis(2, at=c(0,20,40,60,80,""))
abline(l_political_af,col="blue")
abline(l_political_as)
abline(l_political_eu,col="red")
abline(l_political_na,col="green")
abline(l_political_oc,col="cyan")
abline(l_political_sa,col="yellow")
box("figure")
dev.off()
```

A.5. Chapter 4

```
#define variables

Eu=as.numeric(continent_avg=="europe")
Af=as.numeric(continent_avg=="africa")
Oc=as.numeric(continent_avg=="oceania")
Sa=as.numeric(continent_avg=="southAmerica")
Na=as.numeric(continent_avg=="northAmerica")
As=as.numeric(continent_avg=="asia")
lat1_avg=poly(latitude_avg, 2)[,1]
```

```
lat2_avg=poly(latitude_avg, 2)[,2]
lat1=poly(latitude, 2)[,1]
lat2=poly(latitude, 2)[,2]
Birth=birth_rate_avg
Lon.c1=long1_avg
Lon.c2=long2_avg
Lat.p1=lat1_avg
Lat.p2=lat2_avg
Cont=continent_avg

# build regression model for econ index, approach 2

b=lm(index_econ_avg~birth_rate_avg+long1_avg+long2_avg+lat1_avg+lat2_avg+
       continent_avg+continent_avg*long1_avg+continent_avg*long2_avg+
       continent_avg*lat1_avg+continent_avg*lat2_avg)
b=lm(index_econ_avg~birth_rate_avg+long1_avg+long2_avg+lat1_avg+lat2_avg+
       As+Af+Eu+Na+Oc+As*lat1_avg+As*lat2_avg)
b=lm(index_econ_avg~birth_rate_avg+long1_avg+long2_avg+lat1_avg+lat2_avg+
       As+Eu+Na+Oc+As*lat1_avg+As*lat2_avg)
b=lm(index_econ_avg~birth_rate_avg+long1_avg+long2_avg+lat1_avg+lat2_avg+
       As+Na+Oc+As*lat1_avg+As*lat2_avg)
bf=lm(index_econ_avg~birth_rate_avg+long1_avg+long2_avg+lat1_avg+lat2_avg+
        As+Oc+As*lat1_avg+As*lat2_avg)
lm_econ=bf
# build regression model for econ index, approach 1

datreg=data.frame(birth_rate_avg,long1_avg,long2_avg,lat1_avg,lat2_avg,
                  continent_avg)
summary(lm(index_econ_avg~.^2, data=datreg))
lm_step_econ=lm(index_econ_avg~Birth+Lon.c1+Lon.c2+Lat.p1+Lat.p2+
                  Lon.c1*Lat.p1+Lon.c1*Lat.p2+Lon.c2*Lat.p1+Lon.c2*Lat.p2)
# final model, econ index

lm_econ=lm(index_econ_avg~Birth+Lon.c1+Lon.c2+Lat.p1+Lat.p2+As+Oc+
             As*Lat.p1+As*Lat.p2)

#summary, approach 1
library(xtable)
d=xtable(summary(lm_step_econ),align="|r|r|r|r|r|",digits=2)
print(d,floating=FALSE)

#summary, approach 2
library(xtable)
d=xtable(summary(lm_econ),align="|r|r|r|r|r|",digits=2)
print(d,floating=FALSE)
```

```
#plot polynomial fitted for latitude
pdf("C:/Users/Alex/Documents/pic100.pdf")
plot(latitude_avg,17.4058*lat1_avg+ 46.9333*lat2_avg, xlab="Lat",
     ylab="" )
dev.off

pdf("C:/Users/Alex/Documents/pic101.pdf", width=6, height=3)
library(CDVine)
par(mfrow=c(1,2), mar=c(5,4,3,3))
#plot fitted vs std res
plot(predict(lm_econ),studres(lm_econ), xlab="fitted values",
     ylab="std.Res")
lines(lowess(predict(lm_econ),studres(lm_econ)), col="red")
box("figure")
#QQ plot
plot(bf,which=2, main="", xlab="theoretical quantiles")
lines(c(-5,5),c(-5,5), col="green")
box("figure")
dev.off

#plot covariates vs std. res
pdf("C:/Users/Alex/Documents/pic102.pdf", width=9, height=6)
library('CDVine')
par(mfrow=c(2,3))
r=studres(lm_econ)
s=scatterplot3d(long1_avg,long2_avg,r, xlab="Lon.c1", ylab="Lon.c2",
                zlab="")
lm=lm(r~long1_avg+long2_avg)
s$plane3d(lm, col="red")
box("figure")
plot(latitude_avg,r, xlab="Lat")
lines(lowess(r~latitude_avg), col="red")
box("figure")
plot(birth_rate_avg,r, xlab="Birth")
lines(lowess(r~birth_rate_avg), col="red")
box("figure")
plot(as.factor(As),r, xlab="As")
box("figure")
plot(as.factor(Oc),r, xlab="Oc")
box("figure")
hist(r, xlim=c(-3,3), prob=TRUE, main="Histogram of std.Res", xlab="")
lines(seq(-3,3,0.001),dnorm(seq(-3,3,0.001)), col="red")
box("figure")
r_econ=r
```

```
#build regression model for social index, approach 2
b=lm(index_social_avg~birth_rate_avg+long1_avg+long2_avg+lat1_avg+
       lat2_avg+continent_avg+continent_avg*long1_avg+continent_avg*
       long2_avg+continent_avg*lat1_avg+continent_avg*lat2_avg)
b=lm(index_social_avg~birth_rate_avg+long1_avg+long2_avg+lat1_avg+
       lat2_avg+As+Oc+Eu+Na+Af+As*lat1_avg+As*lat2_avg)
bf=lm(index_social_avg~birth_rate_avg+long1_avg+long2_avg+lat1_avg+
        lat2_avg+As+Oc+Na+As*lat1_avg+As*lat2_avg)
#build regression model for social index, approach 1
summary(lm(index_social_avg~.^2, data=datreg))
summary(step(lm(index_social_avg~.^2, data=datreg), direction="both",
             k=log(151)))
lm_step_social=lm(index_social_avg~Birth+Lon.c1+Lon.c2+Lat.p1+Lat.p2+
                    Cont+Birth*Lon.c1+Birth*Lon.c2+Lon.c1*Lat.p1+
                    Lon.c1*Lat.p2+Lon.c2*Lat.p1+Lon.c2*Lat.p2+Lat.p1*
                    Cont+Lat.p2*Cont)
summary(lm_step_social)
#final model, social index
lm_social=lm(index_social_avg~Birth+Lon.c1+Lon.c2+Lat.p1+Lat.p2+
               As+Oc+As*Lat.p1+As*Lat.p2)

#summaty, approach 1
library(xtable)
d=xtable(summary(lm_step_social),align="|r|r|r|r|r|",digits=2)
print(d,floating=FALSE)

#summary, approach 2
library(xtable)
d=xtable(summary(lm_social),align="|r|r|r|r|r|",digits=2)
print(d,floating=FALSE)

pdf("C:/Users/Alex/Documents/pic103.pdf", width=9, height=4.5)
par(mfrow=c(1,2), mar=c(6,5,4,3))
#plot fitted vs std. res
plot(predict(lm_social), studres(lm_social), ylab="std.Res",
     xlab="fitted values")
lines(lowess(predict(lm_social), studres(lm_social)), col="red")
box("figure")
#QQ plot
plot(bf,which=2, main="", xlab="theoretical quantiles")
lines(c(-5,5),c(-5,5), col="green")
box("figure")
dev.off()

#plot covariates vs std. res
pdf("C:/Users/Alex/Documents/pic104.pdf", width=10, height=5)
```

```
par(mfrow=c(2,4))
r=studres(lm_social)
s=scatterplot3d(long1_avg,long2_avg,r, xlab="Lon.c1", ylab="Lon.c2",
                zlab="")
lm=lm(r~long1_avg+long2_avg)
s$plane3d(lm, col="red")
box("figure")
plot(latitude_avg,r, xlab="Lat")
lines(lowess(r~latitude_avg), col="red")
box("figure")
plot(birth_rate_avg,r, xlab="Birth")
lines(lowess(r~birth_rate_avg), col="red")
box("figure")
plot(as.factor(As),r, xlab="As")
box("figure")
plot(as.factor(Na),r, xlab="Na")
box("figure")
plot(as.factor(Oc),r, xlab="Oc")
box("figure")
hist(r, prob=TRUE, main="Histogram of std:Res", xlab="")
lines(seq(-3,3,0.001),dnorm(seq(-3,3,0.001)), col="red")
box("figure")
r_social=r
dev.off()

#build regression model for polit index, approach 2
b=lm(index_political_avg~birth_rate_avg+long1_avg+long2_avg+lat1_avg+
       lat2_avg+continent_avg+continent_avg*long1_avg+continent_avg*
       long2_avg+continent_avg*lat1_avg+continent_avg*lat2_avg)
b=lm(index_political_avg~birth_rate_avg+long1_avg+long2_avg+lat1_avg+
       lat2_avg+continent_avg+continent_avg*long1_avg+continent_avg*
       long2_avg)
b=lm(index_political_avg~birth_rate_avg+long1_avg+long2_avg+lat1_avg+
       lat2_avg+As+Af+Eu+Oc+Na+As*long1_avg+As*long2_avg+Na*long1_avg
     +Na*long2_avg)
b=lm(index_political_avg~birth_rate_avg+long1_avg+long2_avg+lat1_avg+
       lat2_avg+As+Af+Eu+Oc+Na+Na*long1_avg+Na*long2_avg)
b=lm(index_political_avg~birth_rate_avg+long1_avg+long2_avg+lat1_avg+
       lat2_avg+As+Af+Eu+Oc+Na)
bf=lm(index_political_avg~long1_avg+long2_avg+lat1_avg+lat2_avg+Na+Eu)
#build regression model for polit index, approach 1
step(lm(index_political_avg~.^2, data=datreg), direction="both",
     k=log(151))
lm_step_polit=lm(index_political_avg~ Birth+Lon.c1+Lon.c2+Lat.p1+Lat.p2
```

```
                 +Birth*Lat.p1+Birth*Lat.p2)
summary(lm_step_polit)
#final model, polit index
lm_polit=bf
summary(lm_polit)

#summary, approach 1
library(xtable)
d=xtable(summary(lm_step_polit),align="|r|r|r|r|r|",digits=2)
print(d,floating=FALSE)

#summary approach 2
library(xtable)
d=xtable(summary(lm_polit),align="|r|r|r|r|r|",digits=2)
print(d,floating=FALSE)

#plot polynomial fitted for latitude
pdf("C:/Users/Alex/Documents/pic1000.pdf")
plot(latitude_avg,46.043*lat1_avg+42.080*lat2_avg, xlab="Lat", ylab="" )
dev.off

pdf("C:/Users/Alex/Documents/pic105.pdf", width=9, height=4.5)
par(mfrow=c(1,2), mar=c(6,5,4,3))
#plot fitted vs std. res
plot(predict(lm_polit), studres(lm_polit), ylab="std.Res",
     xlab="fitted values")
lines(lowess(predict(lm_polit), studres(lm_polit)), col="red")
box("figure")
#QQ plot
plot(bf,which=2, main="", xlab="theoretical quantiles")
lines(c(-5,5),c(-5,5), col="green")
box("figure")
dev.off

pdf("C:/Users/Alex/Documents/pic106.pdf", width=9, height=6)
par(mfrow=c(2,3))
#plot cavarite vs std. res
r=studres(lm_polit)
s=scatterplot3d(long1_avg,long2_avg,r, xlab="Lon.c1", ylab="Lon.c2",
                zlab="")
lm=lm(r~long1_avg+long2_avg)
s$plane3d(lm, col="red")
box("figure")
plot(latitude_avg,r, xlab="Lat")
lines(lowess(r~latitude_avg), col="red")
```

```
box("figure")
plot(as.factor(Eu),r, xlab="Eu")
box("figure")
plot(as.factor(Na),r, xlab="Na")
box("figure")
hist(r, prob=TRUE, xlim=c(-3,3), main="Histogram of std.Res", xlab="")
lines(seq(-3,3,0.001),dnorm(seq(-3,3,0.001)), col="red")
box("figure")
r_political=r
dev.off()
```

A.6. Chapter 5

```
#contour plots for gaussian copula
pdf("C:/Users/Alex/Documents/pic107.pdf", width=9,height=3)
par(mfrow=c(1,3))
BiCopMetaContour(family=1, par=0.2) #gauss
BiCopMetaContour(family=1, par=0.5)
BiCopMetaContour(family=1, par=0.8)
#get Kendall's tau
BiCopPar2Tau(family=1, par=0.2)
BiCopPar2Tau(family=1, par=0.5)
BiCopPar2Tau(family=1, par=0.8)
dev.off()

#contour plots for t copula
pdf("C:/Users/Alex/Documents/pic108.pdf", width=9,height=3)
par(mfrow=c(1,3))
BiCopMetaContour(family=2, par=0.3, par2=4)
BiCopMetaContour(family=2, par=0.5, par2=4)
BiCopMetaContour(family=2, par=0.8, par2=4)
#get Kendall's tau
BiCopPar2Tau(family=2, par=0.3, par2=4)
BiCopPar2Tau(family=2, par=0.5, par2=4)
BiCopPar2Tau(family=2, par=0.8, par2=4)
dev.off()

#conotur plots for Clayton copula
pdf("C:/Users/Alex/Documents/pic109.pdf", width=9,height=3)
par(mfrow=c(1,3))
BiCopMetaContour(family=3, par=0.5)
BiCopMetaContour(family=3, par=1)
BiCopMetaContour(family=3, par=3)
#get Kendall's tau
```

```
BiCopPar2Tau(family=3, par=0.5)
BiCopPar2Tau(family=3, par=1)
BiCopPar2Tau(family=3, par=3)
dev.off()

#conotur plots for Gumbel copula
pdf("C:/Users/Alex/Documents/pic110.pdf", width=9,height=3)
par(mfrow=c(1,3))
BiCopMetaContour(family=4, par=1.3)
BiCopMetaContour(family=4, par=2)
BiCopMetaContour(family=4, par=6)
#get Kendall's tau
BiCopPar2Tau(family=4, par=1.3)
BiCopPar2Tau(family=4, par=2)
BiCopPar2Tau(family=4, par=6)
dev.off()

#conotur plots for Frank copula
pdf("C:/Users/Alex/Documents/pic111.pdf", width=9,height=3)
par(mfrow=c(1,3))
BiCopMetaContour(family=5, par=-10)
BiCopMetaContour(family=5, par=3)
BiCopMetaContour(family=5, par=10)
#get Kendall's tau
BiCopPar2Tau(family=5, par=-10)
BiCopPar2Tau(family=5, par=3)
BiCopPar2Tau(family=5, par=10)
dev.off()
```

A.7. Chapter 6

```
pdf("C:/Users/Alex/Documents/pic114.pdf", width=9, height=6)
par(mfrow=c(2,3), mar=c(4,3,2,2))
#define u-data
u_econ=pnorm(lm_econ$residual/summary(lm_econ)$sigma)
u_social=pnorm(lm_social$residual/summary(lm_social)$sigma)
u_polit=pnorm(lm_polit$residual/summary(lm_polit)$sigma)
#scatterplot, u-data
plot(u_econ,u_social)
box("figure")
plot(u_econ,u_polit)
box("figure")
plot(u_social,u_polit)
box("figure")
```

```
BiCopMetaContour(u_econ,u_social,  levels=c(0.04,0.07,0.1,0.15,0.2))
box("figure")
BiCopMetaContour(u_econ,u_polit,levels=c(0.04,0.05,0.07,0.15,0.2))
box("figure")
BiCopMetaContour(u_social,u_polit,levels=c(0.04,0.05,0.07,0.15,0.2))
box("figure")
dev.off()
#get empirical Kendall's tau, u-data
cor(u_econ, u_social, method="kendall")
cor(u_econ, u_polit, method="kendall")
cor(u_social, u_polit, method="kendall")


#get mle estimate
dd=data.frame(u_econ,u_social)
BiCopEst(u_econ,u_social, family=0)
BiCopEst(u_econ,u_social, family=1)
BiCopEst(u_econ,u_social, family=2)
BiCopEst(u_econ,u_social, family=3)
BiCopEst(u_econ,u_social, family=4)
BiCopEst(u_econ,u_social, family=5, method="mle")
BiCopEst(u_econ,u_social, family=10)
BiCopEst(u_econ,u_social, family=20)
BiCopEst(u_econ,u_social, family=30)
BiCopEst(u_econ,u_social, family=40)
BiCopPar2Tau(5,4.960853)

#get AIC
d=data.frame(u_econ,u_social)
BiCopSelect(u_econ,u_social)
CDVineAIC(d,family=1,type=1,par=0.6337182)$AIC
#CDVineAIC(d,family=2,type=1,par=0.5799164)
CDVineAIC(d,family=3,type=1,par= 1.181517)$AIC
CDVineAIC(d,family=4,type=1,par=1.642378)$AIC
CDVineAIC(d,family=5,type=1,par=4.960853)$AIC
CDVineAIC(d,family=20,type=1,par=6,par2=0.6083302)$AIC

#emptirical contourplot and contourplot of fitted Frank copula
pdf("C:/Users/Alex/Documents/pic115.pdf", width=9, height=4.5)
par(mfrow=c(1,2))
BiCopMetaContour(u_econ,u_social,
                 levels=c(0.05,0.075,0.1,0.125,0.15), xylim=c(-2,2))
box("figure")
BiCopMetaContour(family=5, par=4.960853,
```

```
                 levels=c(0.05,0.075,0.1,0.125,0.15),xylim=c(-2,2))
box("figure")
#get Kendall's tau of fitted Frank copula
BiCopPar2Tau(family=5,par=4.960853)
dev.off()

#plot simulations

pdf("C:/Users/Alex/Documents/pic116.pdf", width=9, height=4.5)
par(mfrow=c(1,2))
sim=BiCopSim(10000,family=5,par=4.960853)   # simulate from Frank copula
econsim=qnorm(sim[ ,1])*summary(lm_econ)$sigma
socialsim=qnorm(sim[ ,2])*summary(lm_social)$sigma
#simulated predictions, copula
econ_c=68.87848+econsim
social_c=68.05419+socialsim
#simulated predictions, non copula
econ_nc=68.87848+rnorm(10000)*summary(lm_econ)$sigma
social_nc=68.05419+rnorm(10000)*summary(lm_social)$sigma
#plot simulated predictions
plot(econ_c,social_c,xlim=c(0,120), ylim=c(0,120),
     xlab="econ_ger_c", ylab="social_ger_c")
plot(econ_nc,social_nc,xlim=c(0,120), ylim=c(0,120),
     xlab="econ_ger_nc", ylab="social_ger_nc")
dev.off()

#plot joint densities

pdf("C:/Users/Alex/Documents/pic117.pdf", width=9, height=4.5)
par(mfrow=c(1,2))
# joint density, copula model
fu=function(x,y)
{result=BiCopPDF(pnorm(x - 68.87848 ,0,11.89269),
                 pnorm(y - 68.05419,0,10.569), family=5,
                 par=4.960853) * dnorm(x - 68.87848,0,11.89269) *
   dnorm(y - 68.05419,0,10.569)
return(result)}
#plot density
x=seq(1,100,1)
y=x
z=outer(x,y, fu)
contour(x,y,z,xlim=c(0,120), ylim=c(0,120),
        levels=c(0.0015, 0.001, 0.0005, 0.0001))
#joint density, non copula model
```

```
f_nc=function(x,y){
  result = dnorm(x - 68.87848,0,11.89269) * dnorm(y - 68.05419,0,10.569)
  return(result)
}
#plot density
z=outer(x,y, f_nc)
contour(x,y,z,xlim=c(0,120), ylim=c(0,120),
        levels=c(0.0015, 0.001, 0.0005, 0.0001))
dev.off()

#plot confidence regions

pdf("C:/Users/Alex/Documents/pic118.pdf")
x = seq(-50,200,length.out=500)
y = seq(-50,2000,length.out=500)
#evaluate density at grid, copula model
mat=matrix(ncol=length(x), nrow=length(y))
for (i in 1:length(x)){
  for (j in 1:length(y)){
 mat[i,j]=fu(x[i], y[j])

  }
}
#evaluate density at grid, non copula model
mat_nc = matrix(ncol=length(x), nrow=length(y))
  for (i in 1:length(x)){
  for (j in 1:length(y)){
 mat_nc[i,j]=f_nc(x[i], y[j])

  }
}
# determine level of density, copula model
PDF_eval= mat
myfun = function(lev){sum(PDF_eval[PDF_eval>lev])/sum(PDF_eval)-0.95}
level_c = uniroot(myfun,interval=c(0,max(PDF_eval)),
                  tol=.Machine$double.eps)
# determine level of density, non copula model
PDF_eval= mat_nc
myfun = function(lev){sum(PDF_eval[PDF_eval>lev])/sum(PDF_eval)-0.95}
level_nc = uniroot(myfun,interval=c(0,max(PDF_eval)),
                   tol=.Machine$double.eps)
# plot densities
x_c=seq(1,100,1)
y_c=x_c
```

```
z_c=outer(x_c,y_c, fu)
x_nc=seq(1,100,1)
y_nc=x_nc
z_nc=outer(x_nc,y_nc, f_nc)
contour(x_c,y_c,z_c,xlim=c(0,120), ylim=c(0,120),
        levels=c(level_c$root), labels="", col="red")
contour(x_nc,y_nc,z_nc,xlim=c(0,120), ylim=c(0,120),
        levels=c(level_nc$root), labels="", add=TRUE)
```