Humboldt-University of Berlin
Institute of Social Sciences
SoSe 2019
Seminar: Survival Analysis
Lecturer: Alexander Schmotz

Replication analysis of “Death of international organizations”

Kseniia Teslenko
Student number: 603393
E-Mail: teslenkk@hu-berlin.de

Content

1. Introduction

Event history models have become a very popular analysis method for international policy. This method suited very well to studying political change over time. A central quantity in survival analysis is the hazard function. The most common approach to model covariate effects on survival is the Cox hazard model. This model is very popular to research the mortality or survival rate of political regimes, the length of conflicts and civil wars, the duration of alliances, the duration and effects of economic sanctions and others. At the same time the mortality of international organisations and especially the reasons of their mortality have a huge research potential because this topic less researched in comparison with all previous mentioned. Especially intergovernmental organization (IGO) have almost no publication regarding their mortality. To feel this, gap the researcher Mette Eilstrup-Sangiovanni from Cambridge University has researched "Under what conditions do international governmental organizations (IGOs) cease to exist?". To answer the question, she has used the proportional hazard method. In most papers, the authors take the proportional hazard assumption for granted and make no attempts to check that it has not been violated in their data. However, a group of researches recommending checking the data for nonproportionality. The main reason to estimate the model for nonproportionality is the avoidance of statistical models that may rise misleading conclusions.

Before I'll check the model of the research, I need to describe it. I'll start with the theoretical part of the paper and later continue with the empirical one. Next, I illustrate the description of the nonproportional hazard analysis. Then step by step I reproduce the researcher's analysis. I'll implement the nonproportional analysis to her code base and create my own analysis based on these results. At the end I draw my conclusions.

2. The organizational ecology of intergovernmental organizations 1815–2015

Applications of the statistical methods for survival data analysis are broadly used not only in biomedical field but also in social and political science. Researches of international policy have increasingly adopted a wide range of techniques for modelling duration data too. An example is the research of Mette Eilstrup-Sangiovanni "Death of international organizations. The organizational ecology of intergovernmental organizations, 1815–2015". For my term paper I have chosen this publication because the author examines the reasons why international organisations are existing in the long term. To identify this the researcher has used a combination of cross-sectional and survival analysis. By identifying the factors why internationals organisations are "dying" Ms. Eilstrup-Sangiovanni seeks to clarify why are they persisting.

The article consists of three parts. In the first part discusses the theoretical approaches. The researcher of the article focuses on the international governmental organizations (IGOs). IGOs are defined as "organizations with at least three state parties, a permanent headquarters or secretariat, as well as regular meetings and budgets" (Eilstrup-Sangiovanni 2018: 4). In a big number of scientific papers there is a statement that IGOs "rarely die" and at large they are very stable. Some theoretical approaches were examined, namely: the theory of realism, the theory of international institutions and the organizational theories based on sociology. Major theories of international relations, instead of realistic theory, explain the robustness of organisations "due to a combination of high start-up costs, increasing returns from cooperation, and general social and political intransigence" (Eilstrup-Sangiovanni 2018: 7).

The author continues with the empirical part of the analysis. The empirical analysis is based on the Dead IGO-Dataset (DIGO) that contains the information on all IGOs founded between 1815 and 2005, recording their fate as of 2015. The data confirms geographic region, issue-area, mandate, and type and size of membership of IGOs. It was used in a proportional hazard analysis to compare the relative mortality of different

IGOs. The proportional hazard analysis was used because it “measures the unique effect of each co-variate on the hazard-rate, and is thus ideally suited to isolating the independent effects of different institutional design or contextual factors” (Eilstrup-Sangiovanni 2018: 16).

The information from the Dataset was coded with five different features of IGOs. The first feature is the number of member states. The second feature is the membership which is defined by the rules to join the organisation. The next feature is its mandate. Mandate specifies the main “issue-aria” of the organisation. There are six arias used in her research: security, economic and political cooperation, social welfare cooperation, judicial functions, technical support/research, general purpose (Eilstrup-Sangiovanni 2018: 12). The feature number four is scope. The scope shows in which issues the IGO is engaged. For example, UN has economic, security and political goals at the same time. The last feature is the region. It shows from what region the IGO origins.

The last part of the article is the discussion about the findings and the description of the model. At the beginning she compares the death, birth and the population of IGOs during 1850–2005. The results are: “Of 218 IGOs that have died since 1815, fifty-eight (27%) terminated through dissolution, while forty-eight (22%) disappeared through desuetude. Forty IGOs (19%) were absorbed by or merged with another organization, while 61 (28%) were replaced by a successor” (Eilstrup-Sangiovanni 2018: 13). At the same time the population of IGOs is growing. The highest mortality rate is “during the 1910s, 1930s and 1940s—coinciding with the two World Wars and the Great Depression” (Eilstrup-Sangiovanni 2018: 15). For detailed analysis of IGOs mortality the author uses a survival analysis. A proportional hazard measures the effect of each independent variable on the hazard-rate. PH was used to measure the “variation on mortality rates across different subsets of the population” (Eilstrup-Sangiovanni 2018: 16). The researcher run the PH model twice to avoid the problem with results because of dependent variable “dead”. At first the definition for “dead” consisted of all deaths, including successions. At the second calculation “dead” were defined only as expiry, dissolution and desuetude (no succession). The results will be discussed in detailed by replaying the analysis of the author.

At the end of the research the findings were discussed within the theoretical approaches. By this way it was tried to find the explanations of stability and death of IGOs and if the theoretical approaches are still suited. At the end of the article the author gathered her conclusions and ideas for further researches.

3. Nonproportional hazard measures

As mentioned above the author has used a proportional hazard measure for her analysis. Parametric, as well as nonparametric, duration models assume proportional hazards. This means that the effect of a covariate on the hazard rate stays constant over time. By using duration models like the Cox, exponential, Weibull, and Gompertz it is possible to calculate a proportional hazards (PH). Proportional hazards models presume “that the hazard functions of all individuals differ only by a factor of proportionality. That is, if one individual’s hazard rate is 10 times higher than another’s at one point in time, it is 10 times higher at all points in time” (Box-Steffensmeier/Zorn 1998: 3). Instead of assuming that proportional hazards models for examining the duration and timing of political events stay constant, we may expect that the effect of one or more covariates on the hazard rate increases or decreases over time. Even the division of the time scale and fitting them separately with the help of the Weibull model to the observations that could fail in each period are insufficient to single out a nonproportional nature. To be sure it is recommended to test the model for non-proportionality too. The scientific literature for example, Therneau and Grambsch 2000, Box-Steffensmeier and Zorn 2001, Box-Steffensmeier, Reiter, and Zorn 2003, Box-Steffensmeier and Jones 2004, Keele 2010, Licht 2011, Park and Hendry 2015 are recommending trying to violate of the proportionality assumption.

In the international relations the robustness or relationship between states or institutions is very common. As a result, testing the non-parametric hazards is very useful.
There are often usable methods to identify the NPH: Schoenfeld residuals and Grambsch and Therneau global test.
A popular method to detect the NPH is Schoenfeld residuals "that examines the relationship between scaled Schoenfeld residuals and time" (Park/Hendry 2015: 1073). The calculated correlation between the Schoenfeld residuals for each covariate can be presented in a plot. It is expected to have the residuals which average is equal to 0. If the slope is not 0 then the proportional hazard assumption has been violated. But if the plot of Schoenfeld residuals against time shows a non-random pattern, the PH assumption has been violated.
Grambsch and Therneau have modified the Schoenfeld residuals test and after the modifying it the formula for the Therneau and Grambsch nonproportionality test is this:

$$T_k = \frac{\left\{\sum (g_k - \bar{g}) s_k^*\right\}^2}{dI_k \sum (g_k - \bar{g})^2},$$

Where s^*_k is the scaled Schoenfeld residuals, k is covariate, g_k is the time scale, and $\bar{g}$ is the average time scale, I_k is the information matrix elements for covariate k and d are the event times (Keele 2011: 4). The correlation between the covariate specific residual and event times are measured by this test. If any of the results of the test are higher than 5% a violation of the nonproportional hazards assumptions has been confirmed (Keele 2011: 4).
Examples for the nonproportional hazards could be the incorrect functional form for a covariate (Therneau/Grambsch 2000). Correcting the functional forms of covariates needs to be done and fitted before testing for nonproportional hazards (Keele 2011: 4ff).

4. Reproduction

During the reproduction of the analysis I discovered that the results were not the equal to the results of the author. I also found minor problems in the published code and data which made the reproduction process more complicated. I suppose that these singularities might have caused the differences in my results.

The dataset of the author contains information on 561 IGOs founded between 1815 and 2005, recording their fate as of 2015 (11). At the beginning the author prepares the dataset by cleaning some fields and adding new ones. The first problem is that the data set contains at least one error in dataset. The field *scopec* has one value “n” instead of “narrow”. The second problem is that in the dataset every second row is empty.

The first setback was that in the results only five continents appeared. Analysing the R code revealed that the cause for that was, that in the beginning of the code three regions were excluded by overwriting and reordering the levels of the region_c column. Only after the exclusion of Middle East and Atlantic-Oceania the LR Test value (df = 17) matched the one in the article. This doesn’t account for the other values though.

During the reproduction process I noticed that the technical organisations is not possible to estimate because during the test was got the result “NA” (not available) for this row. This means that either the dataset has exact collinearity, or the number of observations are not enough to estimate the relevant parameters. On further inspection of the dataset it appeared that it only consists of 85 observations of 34 variables. This supports the point that there is not enough information about technical organisations.

In the code other uncertainties were found as well. For example, the author creates new dataset in which she excluded the column 83, which doesn’t exists. And then never uses this new dataset. When plotting the Weibull model, I noticed a mistake in the code that delayed the reproduction process but fixing it was useful for the learning process.

My experience suggests that the published dataset is not suitable for the author’s results or the published code is not complete.
I came to the conclusion to avoid the region filter and go forward with small adjustments to the code to keep it working.
Results of the author are:

Death of international organizations. The organizational ecology of...

Table 1 Proportional hazard analysis

Independent Variables	Dependent Variables	
	All Deaths, including Successions	Death = Expiry, Dissolution, Desuetude
Number of Member States		
Number at death (or as of 2015)	−0.032*** (0.005)	−0.045*** (0.008)
Region		
Africa	1.860*** (0.599)	0.356 (0.601)
Americas	1.767*** (0.599)	0.778 (0.577)
Asia	1.372** (0.655)	0.168 (0.672)
Europe	1.975*** (0.576)	0.845 (0.536)
Middle-East	1.156 (0.721)	0.674 (0.675)
Intercontinental	1.563*** (0.552)	−0.059 (0.493)
Mandate/Function		
General Purpose	−0.823* *(0.494)*	−0.589 *(0.781)*
Judicial	0.578 (0.623)	0.916 (0.777)
Security	0.260 *(0.251)*	0.581** *(0.287)*
Social	−0.311* *(0.182)*	−0.670*** *(0.255)*
Technical/Scientific	−0.507** *(0.239)*	−0.754** *(0.331)*
Membership Form		
Geographically restricted	−2.968*** *(0.568)*	−2.287*** *(0.550)*
Purpose restricted	−2.319*** *(0.556)*	−1.152** *(0.509)*
Scope		
Narrow	−0.545 *(0.431)*	0.287 *(0.671)*
Medium	−0.818* *(0.450)*	−0.109 *(0.694)*
No. observations	553	553
R^2	0.200	0.189
Max. Possible R^2	0.989	0.935
Log Likelihood	−1187.559	−697.983
Wald Test (df = 17)	88.400***	94.480***
LR Test (df = 17)	123.353***	115.732***
Score (Logrank) Test (df = 17)	96.266***	101.876***

The overall co-efficient for regional (compared to global) IGOs is: ‘AnyRegion’ 1.672***

Statistical significance

*$p < .10$, **$p < .05$, ***$p < .01$

Figure 1

My replicated results for the first PH Model (All Death, including Successions):

Independent Variables	coef	exp.coef.	se.coef.	z	Pr(>|z|)
Number at death	-0.032	0.968	0.005	-6.305	0.000
Africa	-2.204	0.110	1.041	-2.116	0.034
Americas	-2.282	0.102	1.041	-2.191	0.028
Asia	-2.689	0.068	1.081	-2.487	0.013
Europe	-2.078	0.125	1.035	-2.008	0.045
Global	-4.047	0.017	1.175	-3.443	0.001
InterRegional	-2.456	0.086	1.040	-2.360	0.018
MiddleEast	-2.902	0.055	1.123	-2.585	0.010
OceaniaAntartica	NA	NA	0.000	NA	NA
economic	0.516	1.676	0.239	2.160	0.031
environment	-14.281	0.000	3620.116	-0.004	0.997
general	-0.287	0.751	0.544	-0.528	0.598
judicial	1.091	2.976	0.631	1.729	0.084
security	0.759	2.137	0.324	2.346	0.019
social_welf	0.191	1.211	0.251	0.761	0.447
technical	NA	NA	0.000	NA	NA
memb_formc open	2.953	19.162	0.569	5.186	0.000
memb_formc purp_resticted	0.627	1.873	0.219	2.860	0.004
broad	0.534	1.706	0.431	1.241	0.215
medium	-0.253	0.777	0.190	-1.326	0.185
n	-14.111	0.000	2091.060	-0.007	0.995
narrow	NA	NA	0.000	NA	NA

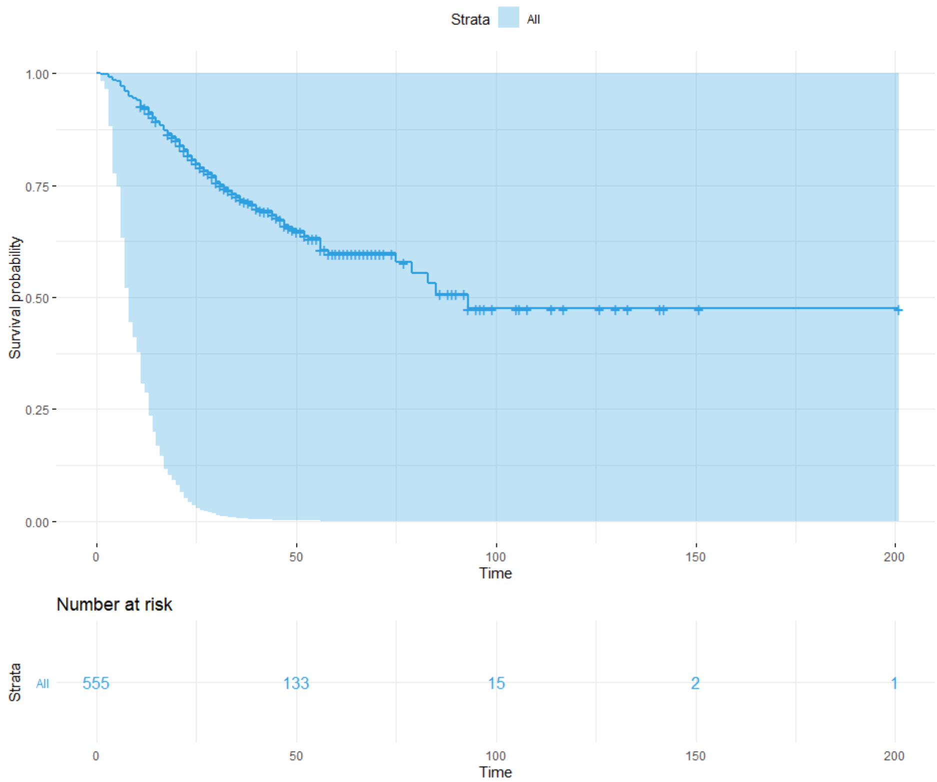

Figure 2

The plot shows that the survival probability decreases drastically during the first 90 observations and then stays slightly below 50% for the rest of the time.
The p-value for all three overall tests (likelihood, Wald, and score) are significant:

- Likelihood ratio test:
    - 125.1 on 19 df
    - p=<0.0000000000000002
- Wald test:
    - 89.24 on 19 df
    - p=0.00000000005
- Score (logrank) test:
    - 98.47 on 19 df
    - p=0.000000000001

The number of observations is 555. As we can see the variables region have less impact in the death of the organisation then their mandate. But European and African Organisations have a higher risk to die than all the other regions. We can see that the possibility to die for judicial or security focused organisations is very high in comparison with other mandates. On the other hand is in our model the judicial organisation not significant but economic organisations are significant and the possibility that they perish in comparison is also high. The membership form variable, which could only be open or restricted, are also significant. Open organisations are more tending to die than organisations with restricted membership. These trends are similar with the results of the author.
The next step was the NPH test. If the Schoenfeld Residuals plot for variable shows deviation from a straight line while it stays flat for the rest of the variables, then we have a strong indicator for an faulty model. The global test has the Null hypothesis defined as "Cox PH assumption valid". To be sure we also tested the Grambsch and Therneau test. P-values less than 0,05 indicate violation of nonproportional hazards assumption. As we can see there is no violation of nonproportionality hazards.

IV	rho	chisq	p
Membership	-0.01222	0.04328	0.835198
Africa	-0.05012	0.538042	0.463245
Americas	-0.06632	0.941604	0.331866
Asia	-0.08747	1.670314	0.196216
Europe	-0.05408	0.629203	0.427648
Global	-0.05151	0.634057	0.425871
InterRegional	-0.07007	1.063649	0.302384
MiddleEast	-0.10466	2.413354	0.120304
OceaniaAntartica	NA	NA	NA
economic	-0.00788	0.013844	0.906336
environment	0.236471	6.89E-08	0.999791
general	-0.04104	0.338543	0.560671
judicial	-0.09628	2.215636	0.136619
security	-0.0471	0.504778	0.477408
social_welf	0.056609	0.721452	0.395668
technical	NA	NA	NA
Membership open	0.065762	1.285462	0.256886
Membership resticted	0.153303	5.573094	0.018239
Broad	0.004745	0.004364	0.947329
Medium	0.008294	0.015161	0.902004
n	0.06015	8.25E-09	0.999928
Narrow	NA	NA	NA
GLOBAL	NA	26.45585	0.232625

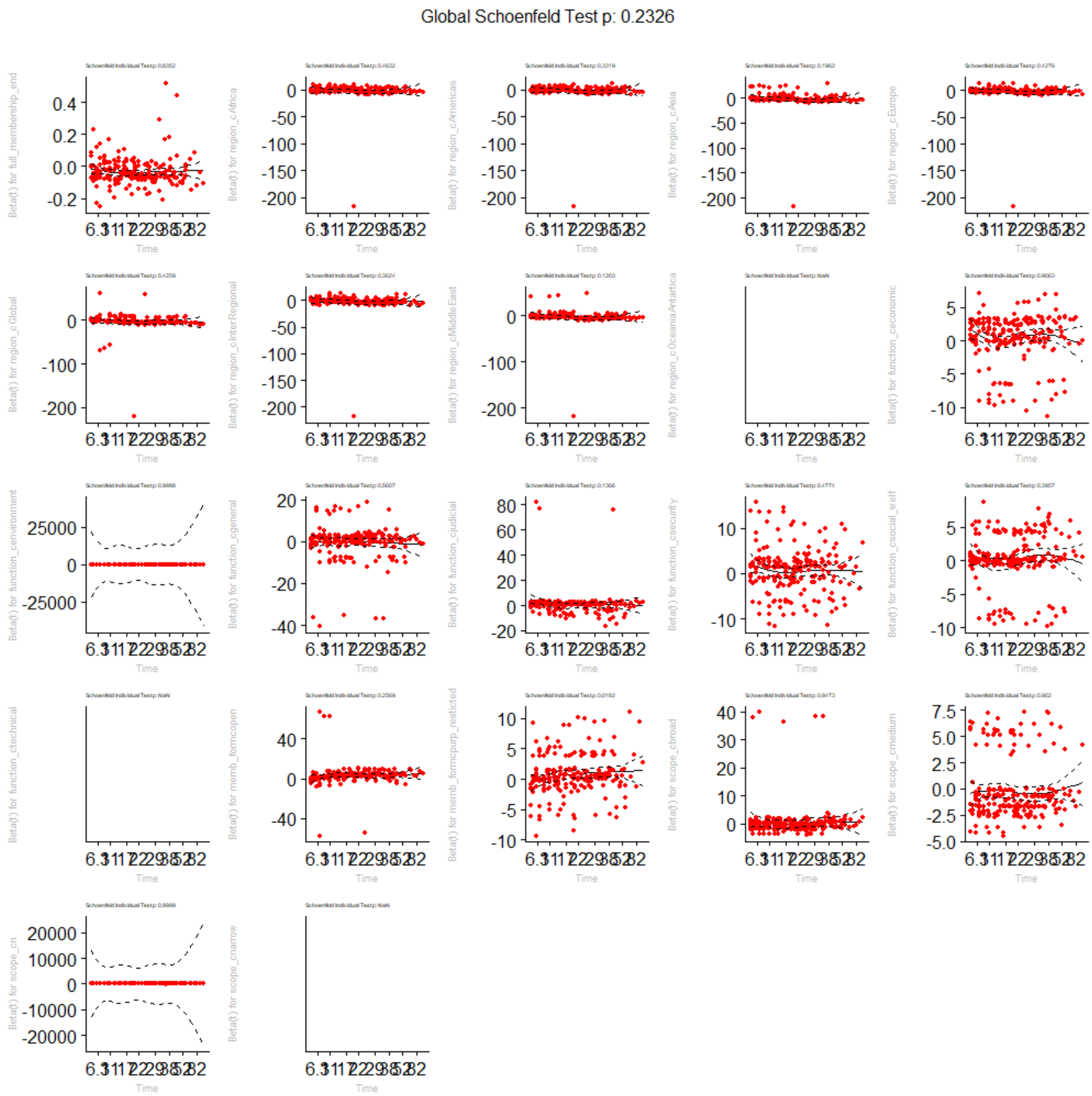

Figure 3

The second PH Model that includes expiry, dissolution, desuetude. The p-value for all three overall tests (likelihood, Wald, and score) are significant.

- Likelihood ratio test:
    - 116.4 on 19 df
    - p=5e-16
- Wald test:
    - 94.35 on 19 df

- p=6e-12
- Score (logrank) test:
    - 102.7 on 19 df
    - p=2e-13

The number of observations is 554. There is no significance with the variable region, but there is significance with the mandate/function of an organisation. It means that region does not influence the hazard or living of the organisations. At the same time security, economical and judicial organisations have more chances to die. We can see also significance in membership form. Organisations with open and restricted membership form are rending to hazard. All these results are similar with the author's results. But the numbers are different.

IV	coef	exp.coef.	se.coef.	z	Pr(>|z|)
Membership	-0.044741414	0.956244722	0.008133465	-5.50090441	3.77848E-08
Africa	13.17288809	525911.426	3511.151876	0.003751728	0.997006561
Americas	13.59435708	801593.6455	3511.151872	0.003871766	0.996910786
Asia	12.98414503	435454.2559	3511.151895	0.003697973	0.997049451
Europe	13.66198195	857676.2249	3511.151869	0.003891026	0.996895419
Global	12.81380557	367252.7474	3511.151903	0.003649459	0.997088159
InterRegional	12.76155257	348555.4385	3511.151874	0.003634577	0.997100033
MiddleEast	13.48983903	722042.3168	3511.151895	0.003841998	0.996934537
OceaniaAntartica	NA	NA	0	NA	NA
economic	0.756601516	2.131021658	0.331285676	2.283834077	0.022381284
environment	-14.44647403	5.32079E-07	5976.167795	-0.00241735	0.998071238
general	0.169155598	1.1843044	0.846705851	0.19978083	0.841651995
judicial	1.665267114	5.287085314	0.79592847	2.092232124	0.036417756
security	1.334057598	3.79641651	0.405736173	3.287992754	0.001009044
social_welf	0.084737431	1.08843124	0.350456766	0.241791396	0.808941806
technical	NA	NA	0	NA	NA
open	2.287084116	9.84618545	0.54992258	4.158920179	3.19756E-05
resticted	1.137920838	3.120274062	0.273850615	4.15526121	3.24916E-05
broad	-0.290148548	0.748152422	0.671501216	-0.43208939	0.665676448
medium	-0.39729195	0.672137766	0.247735	-1.6036973	0.108780792
n	-14.21283354	6.72117E-07	3378.948689	-0.00420629	0.996643877
narrow	NA	NA	0	NA	NA

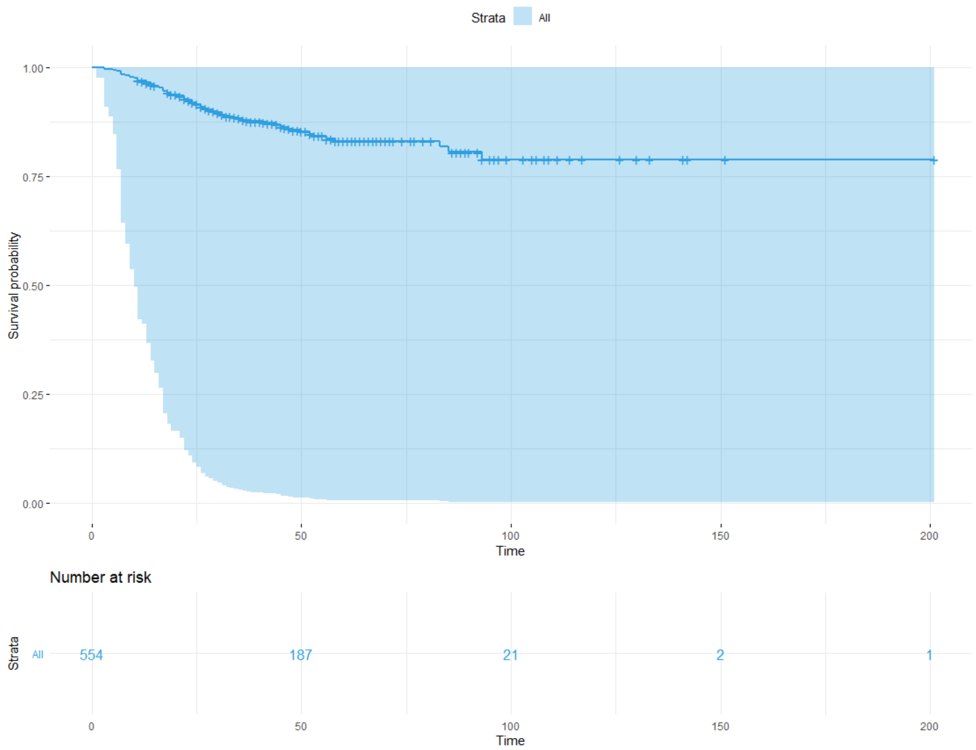

Figure 4

The plot shows that the survival probability decreases during the first 90 observations again and then stays slightly above 75% for the rest of the time.

The next step was the Schoenfeld residuals test and the Grambsch and Therneau test. The Schoenfeld residuals confirm the validity of Cox PH assumption. The number of variables that show some deviation from a straight line is the same as in previous test with six variables out of 19. To be sure we present the results of Grambsch and Therneau test. The test illustrates that the judicial mandate and restricted membership variables are violating the model due to their lower p-values. It seems that the test shows the evidence of non-proportionality. But the overall membership test is not significant. As a result, the Cox model remains as valid.

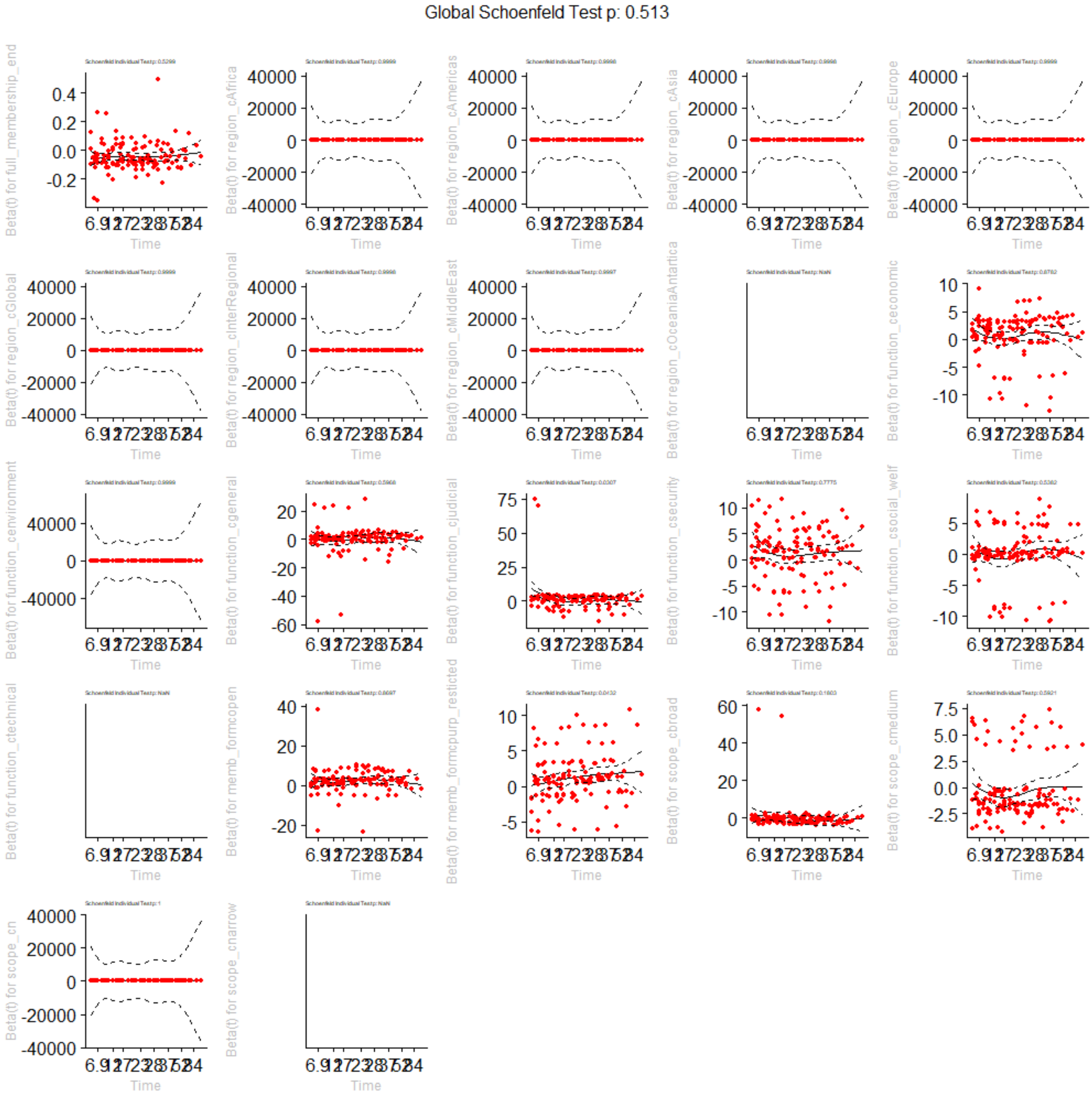

Figure 5

IV	rho	chisq	p
Membership	0.050643	0.394619	0.529881
Africa	-0.09061	1.47E-08	0.999903
Americas	-0.15459	3.59E-08	0.999849
Asia	-0.11722	5.33E-08	0.999816
Europe	-0.09951	1.18E-08	0.999913
Global	-0.05398	1.09E-08	0.999917
InterRegional	-0.19246	6.32E-08	0.999799
MiddleEast	-0.20869	1.66E-07	0.999675
OceaniaAntartica	NA	NA	NA
economic	0.012789	0.023476	0.878225
environment	0.287491	3.5E-08	0.999851
general	0.047376	0.279895	0.59677
judicial	-0.1764	4.668616	0.030719

security	-0.02478	0.079824	0.777536
social_welf	0.053972	0.378845	0.538222
technical	NA	NA	NA
Membership open	0.015534	0.026921	0.86967
Membership resticted	0.166843	4.085645	0.043249
Broad	-0.1235	1.795086	0.180308
Medium	0.047548	0.287153	0.592051
n	0.036958	1.73E-09	0.999967
Narrow	NA	NA	NA
GLOBAL	NA	21.12489	0.513032

In order to avoid mistakes in her model the author made additional calculations. The estimation of the baseline hazard is one example. The Cox proportional hazards model does not estimate the baseline hazard. That is why the author estimate a non-parametric possibility of the models by this way, when the baseline hazard estimator is calculated based on when analysed events occurred and the number of subjects at risk at these points in time. The first Model is the model "All death" and the second model is the model "Death - expiry, dissolution, desuetude". The plot reveals that the risk of dying increases with the "age" of an organisation.

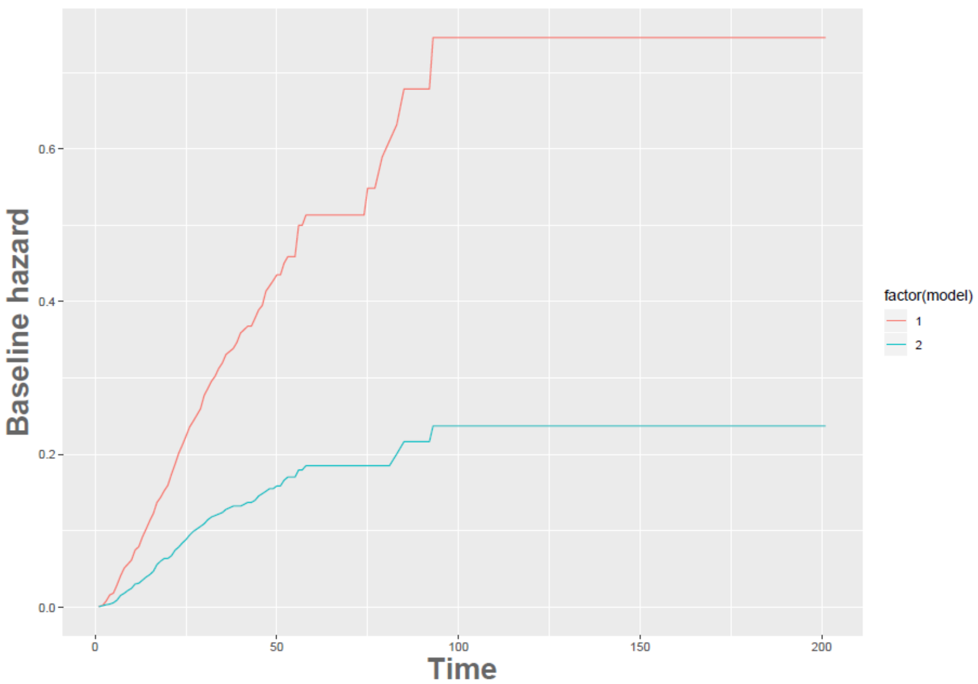

Figure 6

The survival probability for the first model looks a like:

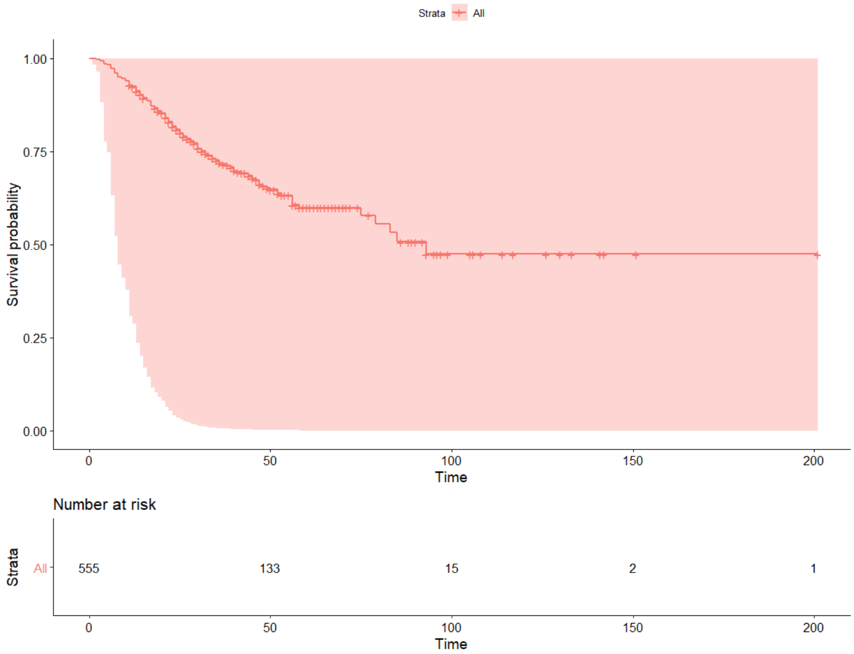

Figure 7

The survival rate stabilize itself at 50 percent by around the 100th observation, which equals approximately half of the overall timeline.
The survival probability for the second model shows that the death probability for organisation is much higher than in the first model but the trend is very similar.

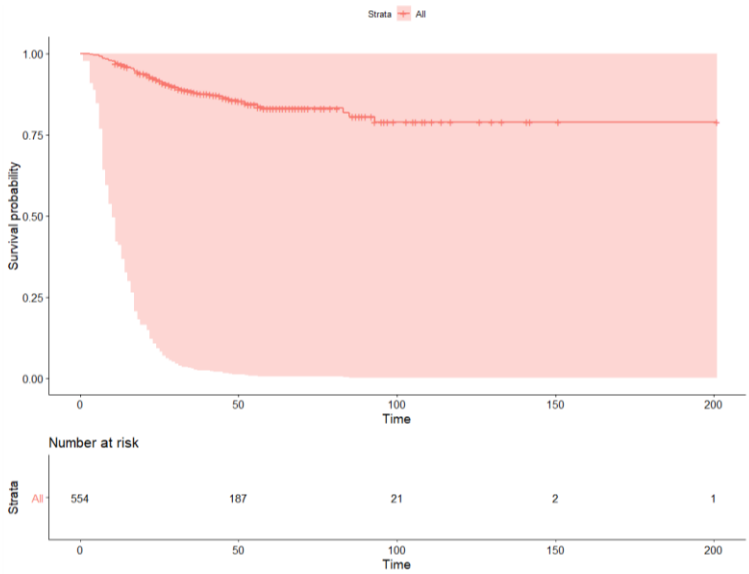

Figure 8

The author later uses the Weibull model with other covariates from this dataset. It is concerning that one covariate from dataset that was used in

Weibull model "members_censor" is not explained. It is difficult to guess what the aim for this model was. It probably was either the comparison with cox models or the creation of new model. The second variant is more possible because the new covariate WW was included. In the Weibull model all regions are excluded. As we can see the author ploted *dweibull* that gives the density and *pweibull* that gives the distribution function. It shows that the increased length of the life means the higher possibility of dying. At this point it comes clear that the World War is a variable that theoretically should have a lot of influence on the hazard rate. The same approach was written by the author. A very high mortality at the start of the observations and later a very slow process of dying can be observed.

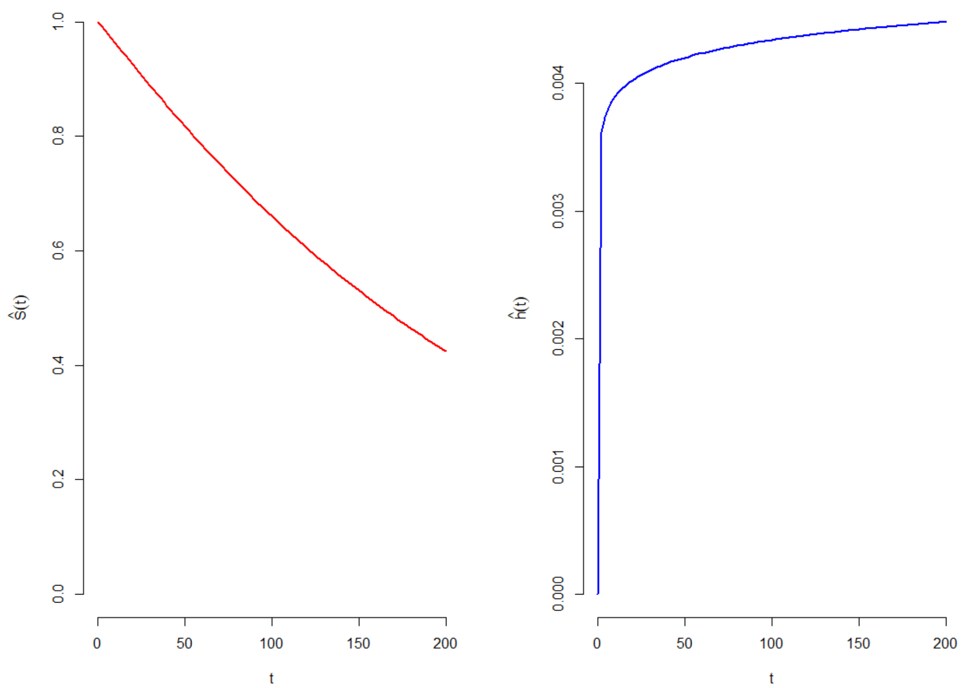

Figure 9

The last step by the author was the Kaplan-Meier estimation. The plot shows that during the first 50 events almost half of the organisations have died.

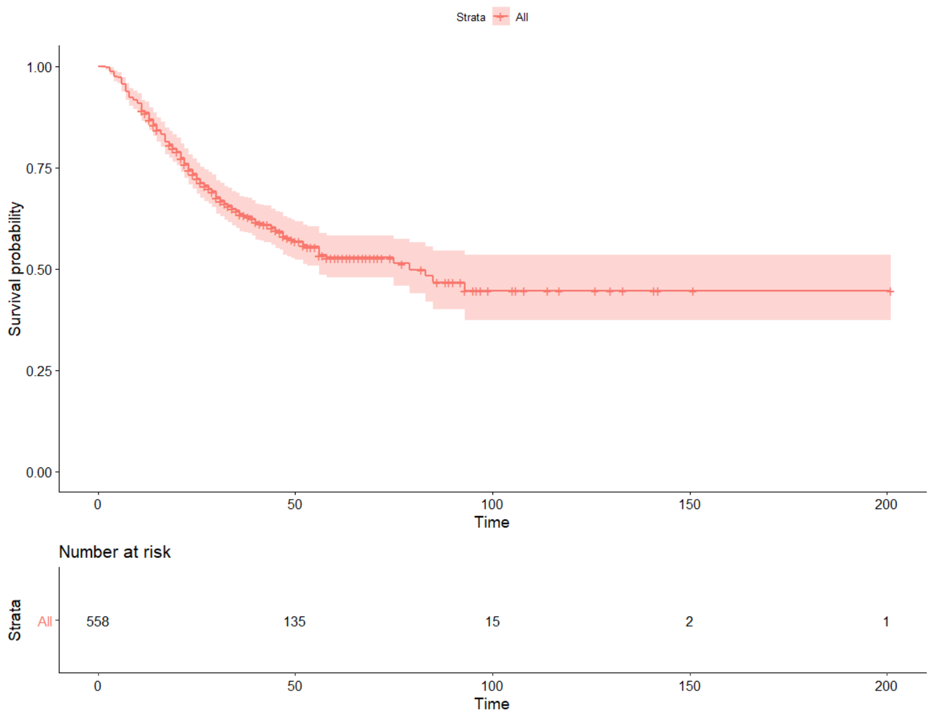

Figure 10

The last plot in the analysis presents us the mortality probability according to their regions.

The survival probability of organisations according to their membership form shows us that organisations which are purpose restricted have a very high chance to collapse. At the same time the absolute number of these organisations ich much smaller in comparison with the others.

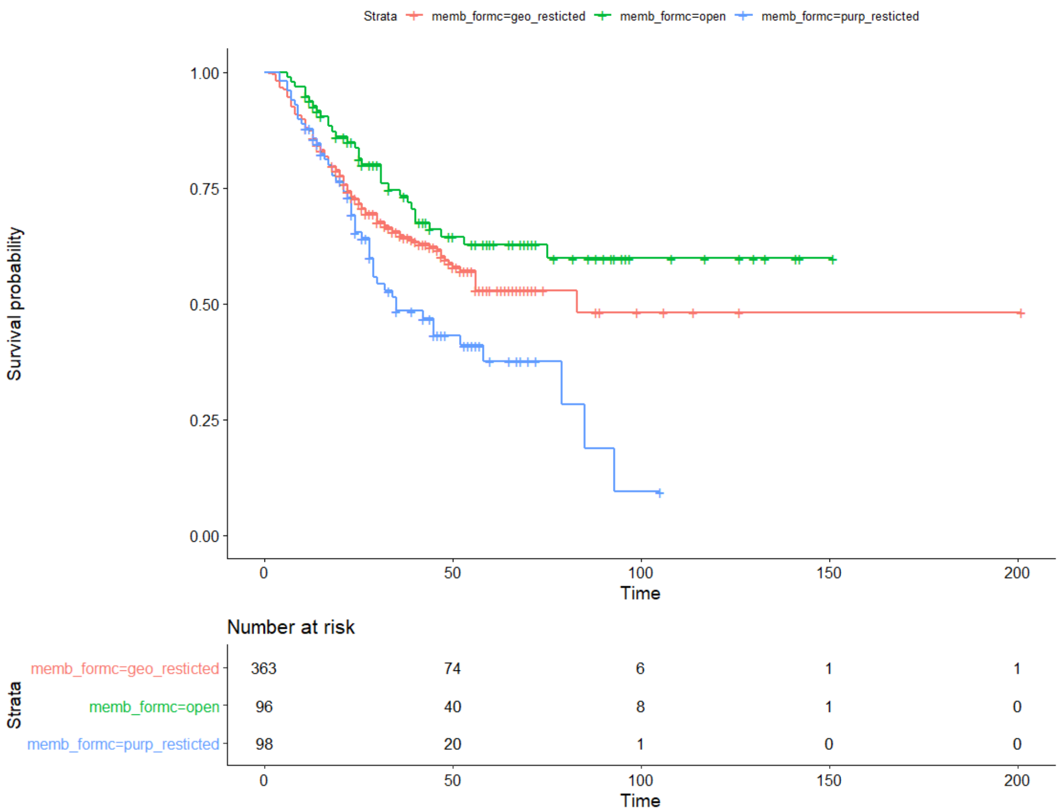

Figure 11

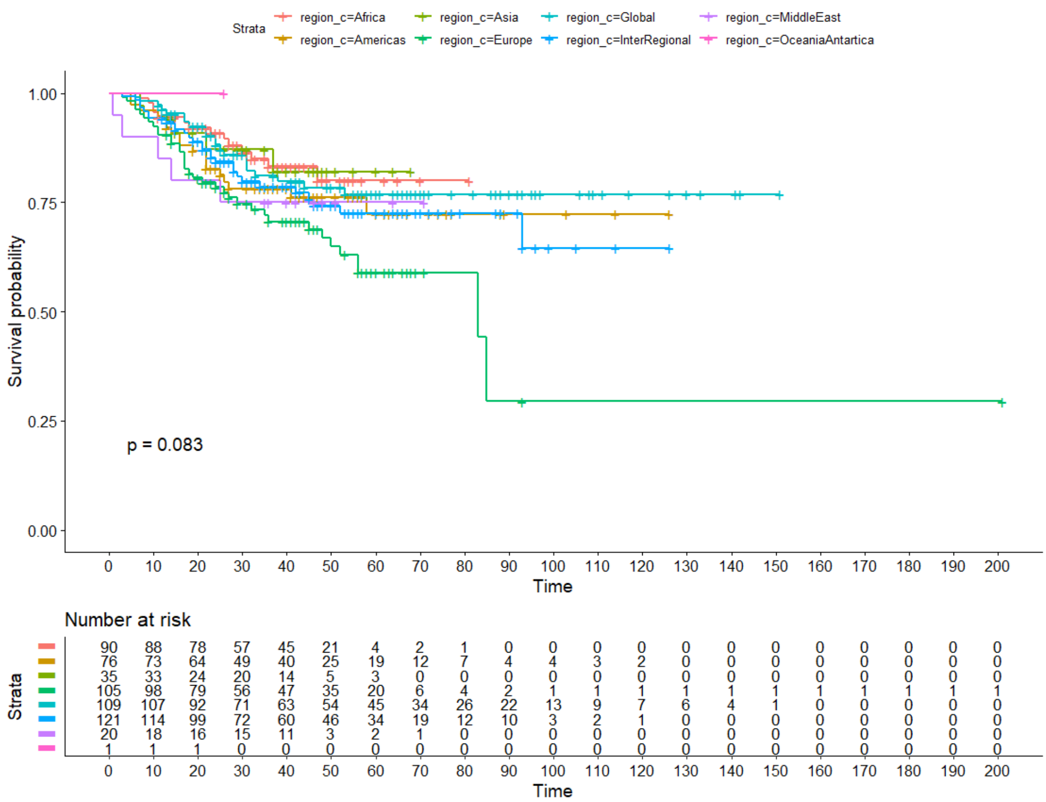

Figure 12

The results are significant due to the p-value of 0.083. The visual representation of the survival function shows what the probability of an event is at a certain time interval in a certain region. Organisations based in Europe have by far the highest mortality rate. The survivability probability drops until half of the observations. At that point it reaches only slightly above 25% in Europe, while still be around 60% for the next area (Interregional).

5. Summary

During my attempt to reproduce the research I encountered some problems that prevented me from getting the same results as the author. It was obvious that the material which was presented online for that purpose was not sufficient enough to recreate the full survival analysis and archiving the exact result. I can only assume that the paper and analysis is based on solid data and programming and just the instructions to reproduce the process were of a lower quality or not in its final state. After smaller adjustments to the code base I was able to get similar results which are still different in detail.
By testing for Non-Proportional Hazards (NPH) using the Schoenfeld residuals test and Grambsch and Therneau test methods I was able to solidify the results of my calculations and the model Ms. Eilstrup-Sangiovanni set up. It was not possible to establish nonproportionality in the model the researcher used.
The finding of the author is that IGO mortality is highly correlated with geopolitical conflict. My analysis supports this by proving that the World War is the covariate that have a lot of influence at the hazard rate. The next finding of author is that technical or scientific IGOs are significantly less likely to die than other types. I can not support this finding because in the given data has a too small number of technical and scientific organisations to archive a meaningful result using the same methods as the rest. As a result of this, there is no evidence if her conclusion is empirically strong.

6. Literature

Box-Steffensmeier, Janet M., Zorn, Christopher J. W. (1998): Duration Models and Proportional Hazards in Political Science, http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.475.4731&rep=rep1&type=pdf (last accessed on 30.09.2019)

Eilstrup-Sangiovanni, Mette (2018): Death of international organizations. The organizational ecology of intergovernmental organizations, 1815–2015, The Review of International Organizations, Volume 14, Issues 54, pp. 1-32

Keele, John Luke (2011): Replication data for: Proportionally Difficult: Testing for Nonproportional Hazards In Cox Models, https://doi.org/10.7910/DVN/VJAHRG, Harvard Dataverse, V1, (last accessed on 30.09.2019)

Keele, John Luke (2011): Testing for Nonproportional Hazards in Cox Models, Political Analysis, Issue 18, pp. 189–205

Park, Sunhee, Hendry David J. (2015): Reassessing Schoenfeld Residual Tests of Proportional Hazards in Political Science Event History Analyses, American Journal of Political Science, Volume 59, Issue 4, pp. 1072-1087

Therneau, Terry M., Patricia M. Grambsch (2000): Modeling survival data: Extending the Cox model. New York: Springer-Verlag

7. Apendix