Quantitative Methods in Political Science
University of Mannheim
Fall 2017
Gschwend, Neunhoeffer & Sternberg
Handed in on December 7th, 2017

Data Essay

Larissa Haas (1417669)

Descriptive Statistics

The variable, which is going to be dependent in our models to test the given hypotheses, counts the **number of speeches** of selected members of parliaments (MP) during an election period. The data was collected for Germany and the United Kingdom. When we look at the distribution of this (obviously) count data (Figure 1), we see its fast falling curve and its long tail. This pattern is typical for a so called Poisson Distribution or a Negative Binomial Distribution (we will test later, what specific kind of distribution we find here).

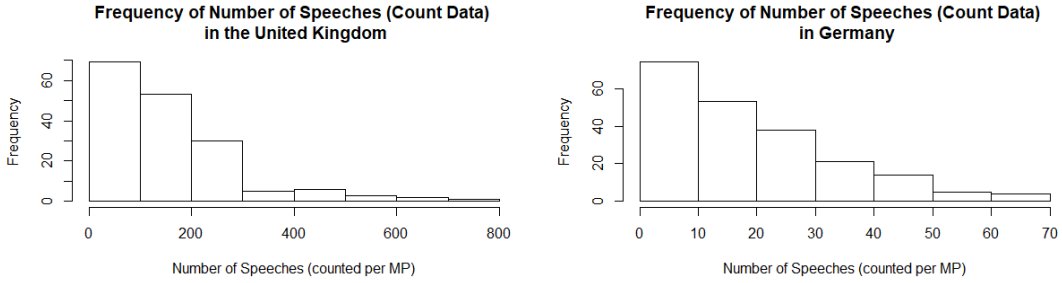

Figure 1: Comparison of Number of Speeches counted in the United Kingdom (2001-2005) and in Germany (2005-2009)

In this figure we also see immediately the different ranges of number of speeches in Germany and the UK: While in Germany the parliament member with the highest number of speeches talked 68 times, in UK we observed a maximum number of 711 speeches (see also the Appendix Overview: Provided Variables for a more details). Therefore, and from now on, we will investigate the variable dependencies for both countries separately.

Besides the dependent variable, both country data sets contain the variable **ideological distance**, which models the distances of the MPs to their respective party leaders. Unfortunately the scales of German and British data are again different so that we cannot compare possible results directly. While the German variable ranges between 0 and 0.833 (but is not bound to this value upwards), the British variable shows values from 0.333 up to 5.455.
To build some more fine grained and controlled models there are several other variables available (again more information in the Appendix: Overview: Provided Variables):

- **Party leader** is (besides ideological distance) the only independent variable that is available for both countries. It is a binary variable, showing 1 when the speaker is the leader of its party, showing 0 otherwise.

- **Conservative MP** is only available for the UK. It is a binary variable that indicates whether the observed speaker is a member of the conservative party or not. Unfortunately we cannot distinguish between the other two parties (Labour and Liberals), that are only marked with 0 in this variable.

- The **Committee** variable is only available for Germany and counts the number of committee assignments the observed speaker has during this electoral period. It's possible that highly engaged MPs are as well members of a lot of committees as giving a lot of speeches, therefore I will control for this "engagement" effect.

- The next very interesting variable is called **caolMPoutside** and indicates whether the observed MP is within the ideological range between the two party leaders of the current government (in this case CDU/CSU and SPD). This information is only given for MPs from CDU/CSU and SPD. If a speaker is quite distant to its party leader but within the range of the coalition it may be not that "rebellish" compared to someone who has the same distance but in the opposite direction.

The following variables I have created and modified based on other variables. I will not talk about the original ones because I will not use them in the models afterwards.

- I turned the categorical variable **party affiliation** into four party dummy variables. When all existing dummy party variables are equal 0, this indicates that the observed MP belongs to the party "Bündnis 90/Die Grünen", which is the only party without a dummy variable. It is important to mention that the distribution of speakers in this data set does not represent the actual proportion of seats in the parliament. While the proportion of CDU/CSU is roughly 7% lower than in the parliament, the other parties are slightly over-represented [1]. This distinction between the parties is important, especially for controlling while predicting the number of speeches, because the parties who are not in the government may have a different amount of speeches than the governing parties.

- Based on this thought and the before explained dummy variables I also created the binary **ingov** variable which indicates whether a speaker is member of the government parties or not.

- I used the existing variable **list candidate** to re-code it into a generic binary variable, because the original values of this variable where 1 or 2, but in the default style binary variables can only take the values 0 or 1, if the observed MP was voted into the parliament via the party list.

Model Selection

Underlying Distribution

First, we will test whether the pattern in the count data shows a poisson or a negative binomial distribution. They differ in the value of α, what is 0 in the poisson distribution. This difference would affect our estimates, because they are based on the assumption of an underlying distribution. The result of the Likelihood-Ratio-Test between the poisson model and the respective negative binomial model shows, that we can at least at a 95%-confidence be sure to reject the H0 that α equals 0. The following models will therefore be build upon the assumption of an underlying negative binomial distribution.

Basic Models

Table 1: Basic Models and first Expansion (created with stargazer [3])

	Dependent variable:			
	Number of Speeches			
	Basic Model	Basic Model	Expanded Model	Expanded Model
	Germany	United Kingdom	Germany	United Kingdom
Ideological Distance (to Party Leader)	−0.688*	0.038	−0.691*	0.041
	(0.362)	(0.060)	(0.358)	(0.060)
Party Leader (Binary)			0.420**	0.316
			(0.203)	(0.340)
Constant	3.060***	4.962***	3.014***	4.940***
	(0.087)	(0.136)	(0.088)	(0.138)
Observations	197	153	197	153
Log Likelihood	−774.302	−922.181	−771.950	−921.705
θ	1.542***	1.305***	1.580***	1.312***
	(0.163)	(0.140)	(0.168)	(0.141)
Akaike Inf. Crit.	1,552.604	1,848.362	1,549.899	1,849.409
RMSE	15.07193	129.8679	14.85278	129.3126
Note:				*p<0.1; **p<0.05; ***p<0.01

With the research hypothesis in mind, we start with a simple model that just uses the expected independent variable as explaining variable. We assume that the position of the party is the position of the party leader, therefore the ideological distance of the speaker to its party is modelled as the ideological distance to the respective party leader. As just mentioned above, there will be a separate model for each country, for that the scales of counts and ideological distances are not comparable.

In this basic model we can see that for Germany, the coefficient of ideological distance is negative (in accordance to our hypothesis 2) and the coefficient is significant on a 90% significance level. Looking at the UK we see a positive coefficient (as expected in hypothesis 1), but unfortunately the coefficient is not significant because the p-value is too high.

In a next step we want to expand this basic model with the remaining comparable variable party leader. It makes sense to include this variable, because a party leader may have a higher number of speeches in parliament, and simultaneously a very low ideological distance (measured to his/her own position, obviously). As we can see in Table 1, the variable ideological distance remains significant for Germany, and also the variable party leader has some impact on the number of speeches. The comparison of the AIC as well as the RMSE support the impression, that the Expanded Model works better for the German data. For the United Kingdom it seems that this additional variable has no additional impact, because it is not significant. This is also supported by the AIC values, because it rises from the Basic Model to the Expanded Model.

Deeper Modelling

United Kingdom

For the United Kingdom there is only one variable left that could be used for further model building. When we add this variable to the already expanded model (see Appendix United Kingdom: Further Models for detailed results), we observe that ideological distance is now significant, too, and also AIC and RMSE are reduced in comparison to the Expanded Model in table 1. When we exclude now the not significant variable party leader we remain with the (in this comparison) best model for predicting the number of speeches in the British parliament.

Germany

For Germany our possibilities to build further models are greater. Appendix Germany: Further Models shows an overview over the different calculated models. **Model 3** was calculated with all the available variables except for the party dummies and showed immediately an improvement as well for the AIC as for the RMSE, this measurement dropped by roughly 2.5 points. But in this model, compared to the previous one from table 1, party leader lost its significance. Ideological distance remained significant at an 0.1-level, while the new variables number of committee memberships and whether the speaker was member of a governing party are significant at a 0.01-level. This could support the suspicion that there is a general "engagement" level underlying, affecting both the number of speeches and the number of committee memberships in a positive manner. The coefficient calculated whether the MP is member of a governing party shows a negative sign, so it seems that the "smaller" parties (because we have a coalition between the two biggest parties at this point in time) are over-proportionally represented in speeches.

The **fourth model** adds the party dummy variables to the previosly described model, but this step worsens both the AIC and the RMSE. No party dummy is significant (CDU/CSU was excluded, because there would be a perfect collinearity with "member of government party" together with the SPD dummy). This can be seen also in the fact, that the signs

and significance levels of the coefficients of the other variables stay the same and the coefficients only change slightly.

In a **fifth model** we want to summarize only those variables that were found to be significant. Indeed, this model is the best of the models we calculated so far, measured by the RMSE as well as by the AIC. There would have been a model with an even lower RMSE (when excluding ideological distance), but this would increase the AIC a lot.

The last model, **model 6** respectively, was only build to be a good foundation for the later on simulation of party differences, therefore, besides the party dummies, only the previous significant variables were included, except for member of governing party, because this would lead to redundancy with the party dummies of SPD and CDU/CSU. This model is not as good as the best model (compared by AIC and RMSE), but it is interesting that SPD and CDU/CSU become significant on a 0.01-level, when the member of governing party variable was removed.

All in all, the best model for Germany achieves a Root-Mean-Squared-Error of roughly 12. This value indicates that we are able to predict the number of speeches a MP does during the observed period with an average error of +- 12. Compared to the range of values the dependent variable can have, this number is quite high. Therefore I would expect that there is still a better model for this data.

Quantities of Interest

The above calculated regression models give us a first impression what to expect from the data, but to really capture the meaning of this coefficients, we want to calculate and interpret some quantities of interest in the following section.

Simulated Differences between Germany and the United Kingdom

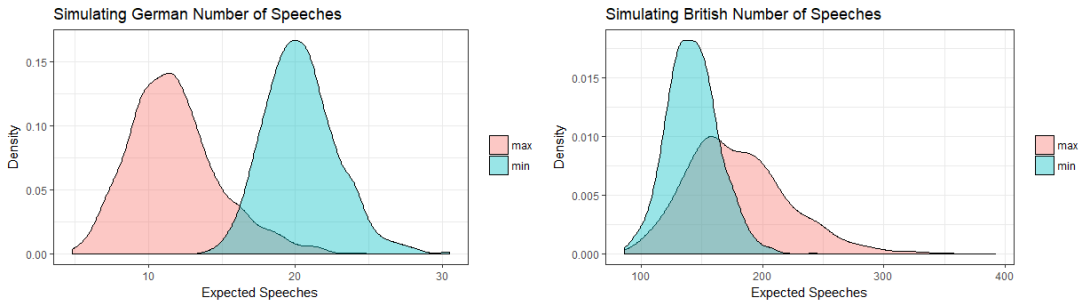

Figure 2: Differences between High and Low Ideological Distances, Comparison between Germany and the United Kingdom (generated with ggplot2 [2])

Figure 2 used the Expanded Models from table 1 to simulate outcomes with interesting variable values. For each of the two countries we used the highest and the lowest value for ideological distance and defined the variable party leader equals 0, because it was the most common value for this variable. The resulting outcomes are shown in red for the highest value and in blue for the lowest value.

By comparing this two figures, we see that the order of the two curves is swapped. This supports the impression we had from the coefficients from table 1, that the hypotheses are supported, at least in their general direction. But we can also see in this comparison, that the curves of the UK overlap more than for Germany, which is another indicator that the impact of ideological distance may be more meaningful for Germany than for the UK.

First Differences for Germany

Table 2: Comparison of four First Differences (simulated with German Data)

Interval	**y/n party leader**	**y/n inside coal. interval**	**y/n list candidate**	**y/n member of gov. party**
quants. 2.5%	−0.125	−0.410	−0.109	−1.158
quants. 50%	0.240	−0.148	0.117	−0.881
quants. 97.5%	0.553	0.099	0.367	−0.591

To support the results interpreted from the regression table, we applied some simulation to have a look at the expected values of number of speeches in different scenarios. These scenarios were built based on Model 3 and took the median values of the variables, except for those the first difference was calculated on.

As we can see (in table 2) our results from the coefficient interpretation were supported. The only First Difference with substantial meaning is whether the MP is member of a governing party or not, because the confidence interval does not include 0. But when we look at the size of the difference, it means 1 speech more or less per election period.

Differences between the German Parties

Because it was quite surprising that the parties didn't show any difference in the number of speeches, we simulated this difference, too, to get a better feeling of the data. In the figure below we see clearly the difference between government and opposition parties. This difference may result out of the unusual predominant size of the government (which controlled nearly 70% of the parliament seats [1]).

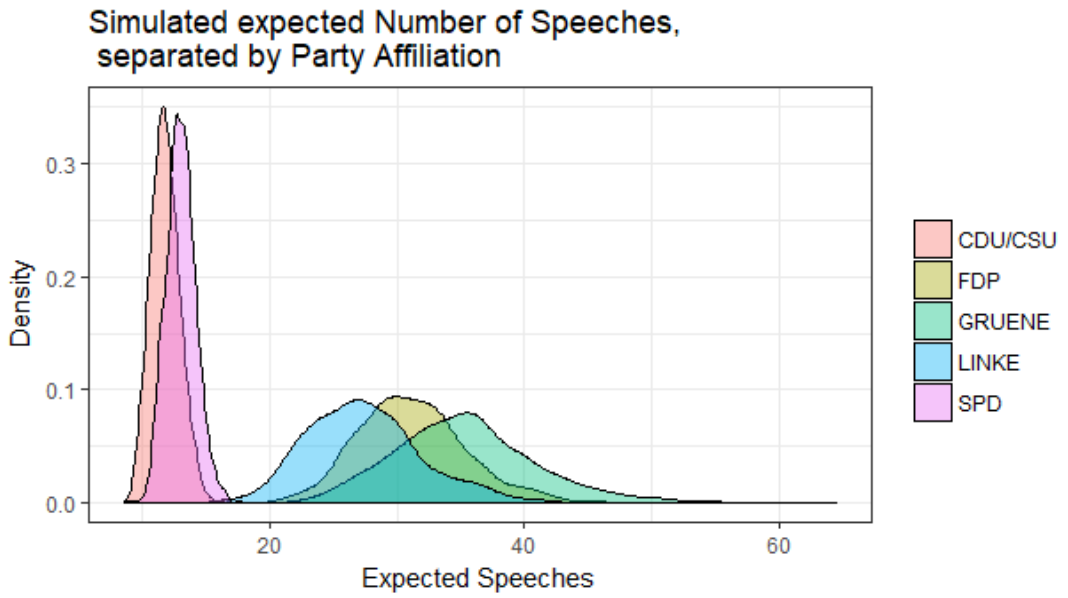

Figure 3: Simulation by Party Affiliation (with German Data) (generated with ggplot2 [2])

Conclusion

All in all we can say, that we can support both hypotheses at least in the direction they stated above. The strength may be worth to be discussed, as well as the importance of other variables, which were not available for the computation of models in this case.

It would be interesting, for example, if the appeared difference in government and opposition parties for Germany is also found for the UK, but unfortunately there was no separation between the Labour and the Liberal party. We saw for the German case, that this difference is highly significant, so maybe we could observe this in other countries as well.

Besides that there is still much work to do when we look at the sizes of effects (for example in table 2) or the size of the RMSE, which is still quite high compared to the range of actual values.

WC: 2176

References

[1] Bundeswahlleiter. Bundestagswahl 2005: Wahl zum 16. deutschen bundestag am 18. september 2005. *https://www.bundeswahlleiter.de/bundestagswahlen/2005.html*, 2005.

[2] Winston Chang Hadley Wickham (hadley@rstudio.com). ggplot2: Create elegant data visualisations using the grammar of graphics v.2.2.1.

[3] Marek Hlavac (Harvard University; hlavac@fas.harvard.edu). stargazer: beautiful latex, html and ascii tables from r statistical output v.5.2.

A Overview: Provided Variables

Table 3: German Parliament Data (2005-2009) (created with stargazer [3])

Statistic	N	Mean	St. Dev.	Min	Max
number_speeches	209	19.182	15.148	0	68
ideological_distance	197	0.175	0.167	0.000	0.833
committee	209	3.196	1.495	0	7
caolMPoutside	209	0.287	0.453	0	1
party_leader	209	0.091	0.288	0	1
FDP	209	0.134	0.341	0	1
CDU_CSU	209	0.287	0.453	0	1
SPD	209	0.368	0.484	0	1
Linke	209	0.100	0.301	0	1
ingov	209	0.656	0.476	0	1
list_candidate_new	209	0.593	0.492	0	1

Table 4: British Parliament Data (2001-2005) (created with stargazer [3])

Statistic	N	Mean	St. Dev.	Min	Max
number_speeches	169	155.657	134.959	0	711
ideological_distance	153	1.949	1.196	0.333	5.455
conservative_MP	169	0.331	0.472	0	1
party_leader	169	0.041	0.200	0	1

B United Kingdom: Further Models

Table 5: Further Models for the United Kingdom (created with stargazer [3])

	Dependent variable:	
	Number of Speeches	
	Full Model Model 3	Reduced (best) Model Model 4
Party Leader	0.434	
	(0.335)	
Ideological Distance	0.118*	0.106*
	(0.064)	(0.064)
Conservative MP	0.379**	0.340**
	(0.160)	(0.161)
Constant	4.643***	4.704***
	(0.171)	(0.169)
Observations	153	153
Log Likelihood	−919.180	−920.108
θ	1.351*** (0.146)	1.337*** (0.144)
Akaike Inf. Crit.	1,846.361	1,846.216
RMSE	128.7014	128.6885
Note:		*p<0.1; **p<0.05; ***p<0.01

C Germany: Further Models

Table 6: Further Models for Germany (created with stargazer [3])

	Dependent variable:			
	Number of Speeches			
	Model 3 Expanded Model	Model 4 Full Model	Model 5 Best Model	Model 6 Party Model
Party Leader (Binary)	0.232	0.253		
	(0.171)	(0.174)		
Ideological Distance (to Party Leader)	−0.524*	−0.571*	−0.385	−0.459
	(0.316)	(0.316)	(0.304)	(0.305)
List Candidate new (Binary)	0.118	0.102		
	(0.124)	(0.123)		
Number of Committees	0.129***	0.129***	0.118***	0.118***
	(0.033)	(0.033)	(0.033)	(0.033)
Outside of Coal. Interval (Binary)	0.153	0.147		
	(0.129)	(0.133)		
Member of gov. Party (Binary)	−0.881***	−1.022***	−0.918***	
	(0.142)	(0.197)	(0.105)	
FDP (Dummy)		−0.102		−0.124
		(0.194)		(0.196)
CDU_CSU (Dummy)				−1.092***
				(0.175)
SPD (Dummy)		0.043		−0.987***
		(0.127)		(0.170)
Linke (Dummy)		−0.279		−0.252
		(0.218)		(0.218)
Constant	2.951***	3.086***	3.128***	3.257***
	(0.199)	(0.238)	(0.148)	(0.198)
Observations	197	197	197	197
Log Likelihood	−732.853	−731.979	−734.806	−733.806
θ	2.456***	2.483***	2.403***	2.431***
	(0.292)	(0.297)	(0.285)	(0.289)
Akaike Inf. Crit.	1,479.705	1,483.958	1,477.611	1,481.612
RMSE	12.31825	12.32628	12.07485	12.09867
Note:				*p<0.1; **p<0.05; ***p<0.01