Microeconometrics Take-Home Assignment

Julia Knöbl, 1108789

Winter term 2015

1 Introduction

Only two American Presidents have ever been in an impeachment trial, Andrew Johnson (1986) and Bill Clinton (1992). This paper looks at the senators behaviour in the impeachment trial of Bill Clinton in February 1999, when the U.S. Senate voted on two articles whether to remove Clinton as president. Those articles were Perjury and Obstruction of Justice, both were based on his testimony in the trial for sexual harassment of Paula Jones, where Clinton denied his sexual relationship with Monica Lewinsky. Both votes fell short of the Constitutional two-thirds majority requirement to convict and remove an officeholder. The final vote was generally along party lines, with no Democrats voting guilty, and only a handful of Republicans voting not guilty.

There has been a huge political and social discussion about this topic, but such as discourse was rarely informed by systematic statistical analysis of the voting behaviour.

The aim of this paper is to analyse, interpret and predict the senators votes in each of the articles and also jointly. The paper is structured in the following way: The first part will describe the most important variables and state hypothesis about their influence on the voting behaviour. The second part features the empirical Analysis. I will go though the different models, explain and analyse them. Then I will predict part of the data by out-of-sample prediction and also take a closer look on the role of conservatism. The last section concludes.

1.1 Data description

The data set contains the following variables for each of the 100 senators (two senators for each state):

- Variables that have to be explained:
  - VoteI: Vote on Article I, Perjury: 1 if the senator voted for guilty
  - VoteII: Vote on Article II, Obstruction of Justice: 1 if the senator voted for guilty
  - NVote: Total number of votes for guilty on Article I and II (0, 1 or 2).
- Variables that can be used to explain the above:

- Party: 0 if senator is a Democrat, 1 if senator is a Republican
- Conserv: Senator's degree of ideological conservationism (0-100)
- ClintVote: Percent of the vote Clinton received in the 1996 Presidential election in each state
- year_end: The year each Senator's seat is up and he/she must run for re-election (or retire)
- first_term: First-term senator? 1 if it is the senators first term, i.e. he or she has the possibility of re-election.

In February 1999 the U.S. senate consisted of 45 democrats and 55 republican senators. All democrats and five voted for not guilty on the first article, and even more voted for not guilty on the second one. In table 1 one sees how the votes on the two articles are distributed. In figure 1 you see the number of "Guilty"-Votes on none, one or both articles.

Variable	**Mean**	**Std. Dev.**
VoteI	0.45	0.5
VoteII	0.5	0.503
N	100	

Table 1: Summary statistics Vote I & II

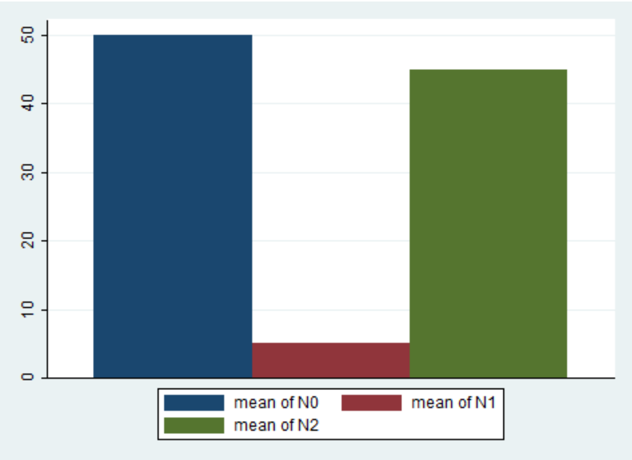


Figure 1: Number of "Guilty"- Votes on none, one or both articles

The senators conservativeness is highly correlated with the party they belong to, and plays therefore an important role in the analysis. The correlation between party an degree of conservatism is 0.9061, so those variables explain each other very well as one can see in figure 2. Hardly any senator whose conservatism is higher than 20 is a democrat. In figure 3 one sees that conservatism is highly correlated to the voting behaviour. Most senators voted either in none or in both cases guilty. One can see clearly that there is a certain degree of conservatism from which on the senator votes guilty in at least one case. There are five votes for "Guilty" in the case of Obstruction of Justice, but not in Perjury. The degree of conservatism does not seem to explain these cases very well.

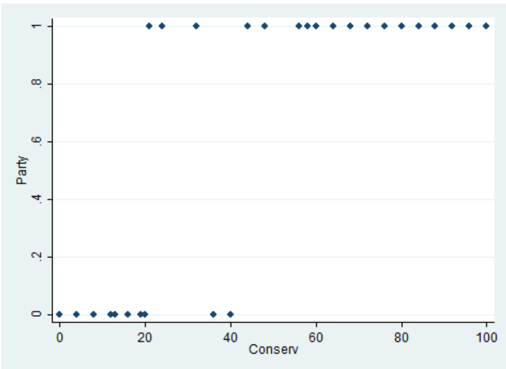


Figure 2: Party vs. degree of conservatism

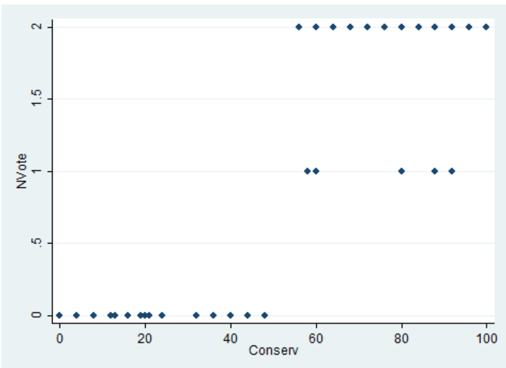


Figure 3: Number of votes vs. conservatism

Looking at the data suggest that *Party* and *Conserv* will play a big role modelling the voting outcomes. The share of votes for Clinton in the 1996 presidential elections will probably be important too. It seems a little as those three variables are all together highly correlated and will cause multi-collinear problems and proxy for each other. The correlation between *ClintVote*and *Party* has an absolute value of just 0.38, which is not very high. But the election of the senators differs from presidential election, so the differences might occur from that.

It seems reasonable that variable *first_term* has different effects for Republicans and Democrats. A senator in his or her first term might be more likely to vote along the party line.

As every second year a third of the Senate is replaced, different senators have shorter or longer until to their possible re-election. Should this variable be relevant it is expected to have, like *first_term*, different effects for Democrats and Republicans. A senator whose re-election will be very soon might be more careful to not disagree too much with the party line.

I also constructed several interaction variables to control for interaction between variables. The ones that proved to be relevant are *firstparty* and *sqrdConserv*. *firstparty* is the interaction between *first_term* and *Party*, so it indicates the Republicans in the first term as senator. One can argue, that these senators want to be re-elected and are therefore going to vote along the party-line. *sqrdConserv* is just the degree of conservatism squared.

2 Empirical Analysis

2.1 Vote I - Perjury

2.1.1 Choosing a Model

As the Variable Vote I is a binary variable it has to be estimated either with a probit or a logit model. The estimated coefficients are very similar in both cases and I made a decision on the basis of information criteria and the Pseudo R^2.

Table 2 shows a sample of different modelling steps. Columns (1) to (5) are estimated by logit and column (6) shows a probit model with the same inputs as in column (2).

The variable *Party* is not included in the model because it explains the negative outcomes of the vote on perjury perfectly (Remember all democrats voted for "Not Guilty"). This causes computational problems and Stata cannot produce an output.

	(1) VoteI	(2) VoteI	(3) VoteI	(4) VoteI	(5) VoteI	(6) VoteI
VoteI						
Conserv	0.524*	0.539*	0.525**	0.0993***	0.108***	0.285*
	(0.265)	(0.245)	(0.171)	(0.0228)	(0.0238)	(0.129)
sqrdConserv	-0.00305	-0.00315*	-0.00302**			-0.00166*
	(0.00167)	(0.00154)	(0.00106)			(0.000814)
ClintVote	-0.0318					
	(0.0637)					
year_end	-0.0645					
	(0.224)					
first_term	-12.25***	-13.25***		-14.67***		-3.189**
	(2.547)	(2.250)		(0.772)		(1.154)
firstparty	12.96***	13.90***		14.92***		3.497***
	(2.129)	(1.832)		(0.982)		(0.971)
_cons	110.8	-20.37*	-19.81**	-5.637***	-6.207***	-10.75*
	(447.6)	(9.391)	(6.418)	(1.213)	(1.275)	(4.941)
N	100	100	100	100	100	100
pseudo R^2	0.770	0.768	0.764	0.738	0.730	0.765
AIC	45.72	41.89	38.42	44.00	41.11	42.36
BIC	63.95	54.91	46.24	54.42	46.32	55.38

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2: Modelling Vote I

In Column (1) one sees that *ClintVote* and *year_end* show no significance. An interaction of those variables with *Party* is also not significant.

Comparing Pseudo-R^2 and the information criteria one can decide for a model. The model in Column (1) has the highest Pseudo- R^2 of 0.77, but this does not mean that it is the best model because a model with more variables is always more general and has therefore a better fit with the data. A rule of thumb for the information criteria (AIC and BIC) is "the smaller, the better". But in contrary to the Pseudo- R^2 they have a so-called "punishment-term" for additional variables. Adding an additional variable that does not improve the model results in a higher AIC and BIC and also in a higher R^2. While the AIC just adds two times the number of parameters, the BIC adds the number of parameters multiplied by the logarithm of the sample size, which results in a way higher punishment for additional variables when the sample size gets large.

The decision for a specific model is quite hard in this case. The differences between probit and logit models are marginal, but in each case the indicators are in favour of the logit model. This is also the reason why I put mainly logit estimates in the table, and only one probit to compare it. Comparing column (2) and (6) shows that all coefficients are on the same significance levels, but the R^2 is slightly higher in the logit case and the AIC and BIC are a little bit lower. Person Goodness-of-fit tests show that both distributions fit the data, logit fits just a little better than probit. For all of those reasons I decided for a logit model.

The next step was then to decide between the models in column (2) and (3). As the Person-goodness-of-fit just tests whether the data is really distributed according to the assumed distribution it does not help in this specific decision. The important factors are R^2 and the information criteria. Although R^2 is marginally higher in the model in column (2) the AIC and BIC are in favour of the model in column(3). So I decided for the model in column(3) although *first_term* and *firstparty* are significant, but they don't seem to add more relevant information, so I take the simpler model.

2.1.2 Interpretation and Model Diagnostics

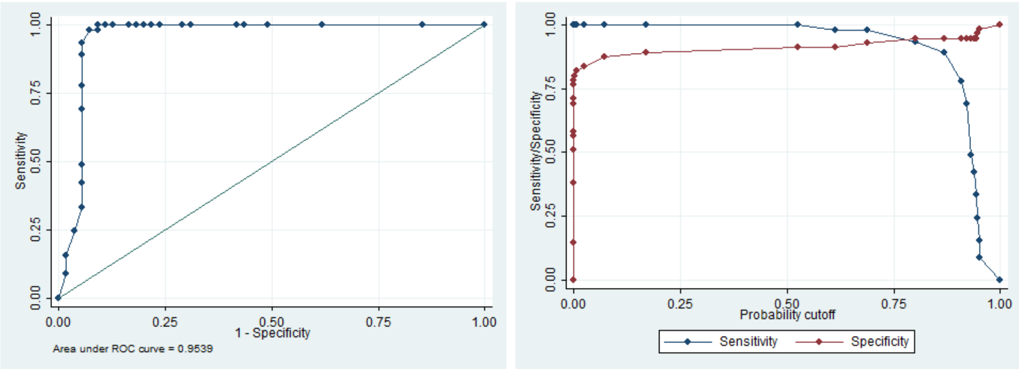
Choosing the logistic model, where the only inputs are the degree of conservatism and the squared form of it one sees that the more conservative a senator is the more likely he or she votes guilty in the perjury impeachment. The coefficients in the table are the changes in the log-odds, so when the degree of conservatism raises by one unit the logarithm of

the odds ratio $ln\frac{p_i}{1-p_i}$ changes by 0.25-0.003= 0.22. These are, however, not the marginal effects (ME). ME in binary response models are level dependent, i.e, they depend on all characteristics of an individual. The ME for the average senator for *Conserv* are 0.0014 and for *sqrdConserv* they are very close to 0 and negative.
The Pearson goodness of fit statistic indicates how far the model deviates from the data in each pattern. There are 28 covariate pattern, and the statistic has a value of 12.53. This is a very low value, and we see that the Null-Hypothesis of the data being logistic distributed cannot be rejected because the p-value is very high (0.98). So as I already mentioned above the underlying data fits the distribution I chose very well.
Using the model one can of course predict the probabilities of the outcome of the vote. In table 3 one sees the mean and Standard deviation of the predicted probabilities as well as the real outcomes.

Variable	**Mean**	**Std. Dev.**	**Min.**	**Max.**
phatI	0.45	0.451	0	0.951
VoteI	0.45	0.5	0	1
N		100		

Table 3: Predicted probabilities and real outcomes in Vote I

The mean is by construction the same, but as the the std. dev. is similar one sees that the prediction comes pretty close to reality. Also because one predicts a minimal probability of 0 and a maximal probability on 0.95. This is very close to the range of the real outcome. For predicting concrete outcomes, instead of just probabilities, one has to choose a threshold from which on the predicted outcome should be positive. As the data is roughly equally distributed (45/55) a threshold of 0.5 seems suitable. Using this threshold one predicts 50 ”Guilty” and 50 ”Not-guilty” Votes. All of the ”Not-Guilty” were predicted right, so just 5 votes were predicted wrong, those are the 5 republicans. In Figure 4(a) you see the resulting ROC curve, and in Figure 4(b) how Sensitivity and Specificity change with the cutoff-value.



(a) ROC Curve of Vote I (b) Sensitivity and Specificity in Vote I

Figure 4: ROC, Sensitivity and Specificity

Looking at Figure 4(b) suggests to take a very high cut-off value. This results in predicting 55 "Guilty" votes, so the correct number of votes, but increases the number of wrong predictions. There is a trade-off of what one aims for - predicting the correct outcome in terms of number of votes, or predicting the behaviour of certain senator correctly.

When analysing a model it is important to look at the residuals. In figure 5 you see that there are a few outliers but in general there is no structure visible.

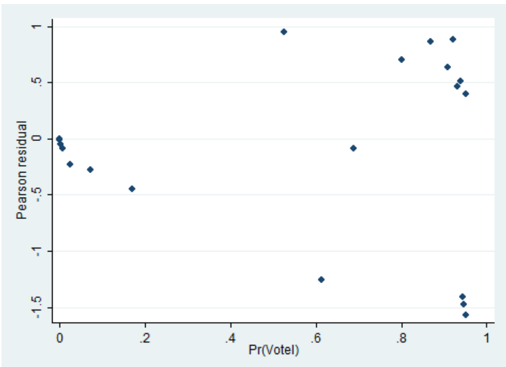
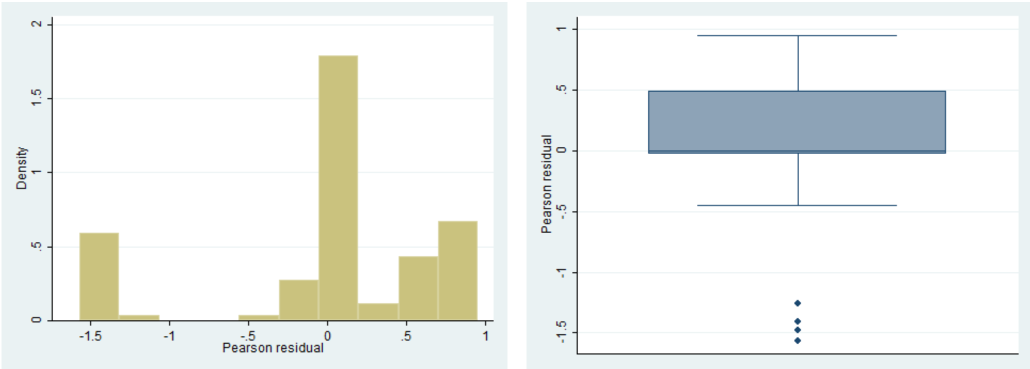


Figure 5: Residuals vs. Predicted probabilities

In figure 6 ones sees in a Histogram and in a box plot that the residuals are centred around 0, slightly skewed, and that there are some outliers. These are the ones of the five republicans voting "Not-Guilty".



(a) Histogram of residuals (b) Boxplot of residuals

Figure 6: Residual distributions

2.2 Vote II - Obstruction of Justice

2.2.1 Choosing a Model

Table 4 shows different modelling steps of the outcome of the second vote, the one on obstruction of justice. Columns (1) and (2) are estimated using a probit model, while column (3) is a logit model. None of the models includes the party of the senator or his or her degree of conservatism. The reason for this is the same as in the case of Vote I: *Party* explains almost all negative outcomes and a degree of conservatism above 48 predicts the data perfectly [1] and stata has computational difficulties.

Because *Conserv* is not included in the model any more *ClintVote* becomes significant[2]. The varible *year_end* is, as before, not significant at all, and therefore taken out of the model. The decision for a model is now between the models in columns (2) and (3). The only difference between them is the underlying distribution. A Parson-goodness-of-fit shows that both distributions fit the data(p-values are 0.977 in logit model and 0.98 in probit model). The deciding factors are again information criteria and the R^2. All of them are in favour of the probit model.

[1]The same is true for a level above 2304 of *sqrdConserv*

[2]*ClintVote*alone is also significant in the other regressions it just has very poor information criteria and a low R^2

	(1) VoteII	(2) VoteII	(3) VoteII
VoteII			
ClintVote	-0.120***	-0.120***	-0.200***
	(0.0299)	(0.0300)	(0.0559)
year_end	0.0357		
	(0.0992)		
first_term	-5.555***	-5.510***	-17.24***
	(0.284)	(0.263)	(0.618)
firstparty	7.306***	7.230***	20.09***
	(0.463)	(0.423)	(0.893)
_cons	-65.98	5.459***	9.161***
	(198.8)	(1.454)	(2.730)
N	100	100	100
pseudo R^2	0.430	0.429	0.425
AIC	89.02	87.16	87.70
BIC	102.0	97.58	98.12

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4: Modelling Vote II

2.2.2 Interpretation and Model Diagnostics

Model chosen has three different input variables. The negative signs of *ClintVote* and *first_term* indicate that the probability of voting guilty on Obstruction of Justice decreases with this variables. This also makes sense as in States where Clinton had a high share of votes also the senator (representing the population) might prefer Clinton. The interaction between the party a senator belongs to and whether it is his or her first period in the senate is negative, which is also intuitive. A republican senator that does not Vote along party-line is less likely to be re-elected.

The coefficients in table 4 show the changes in the z-score when the variable changes by one unit. The numbers itself can therefore not be interpreted well. Like in any binary response model the marginal effects depend on all characteristics of an individual. For the average senator the probability of voting guilty increases by 4.2 percentage points when Clinton has a by one percentage point higher share of the votes in his or her state. For a senator being in his or her first period in the senate the probability of voting guilty decreases by 95 percentage points. This seems very odd. On the contrary a republican senator in his or her first period is 98 percentage points more likely to vote guilty[3] (All other variables fixed at their average level). These numbers seem very unrealistic and might be an indicator that the model is not good.

In table 5 one sees how the predicted probabilities and the real outcome of the vote on Obstruction of Justice were. The mean is very closes together, again this is by construction of the prediction. But as in the previous vote the prediction range covers all outcomes, the highest probability predicted is 0.999 which is almost 1. This indicated that the model is not as bad, as the ME might suggest.

Variable	**Mean**	**Std. Dev.**	**Min.**	**Max.**
phatII	0.497	0.349	0	0.999
VoteII	0.5	0.503	0	1
N		100		

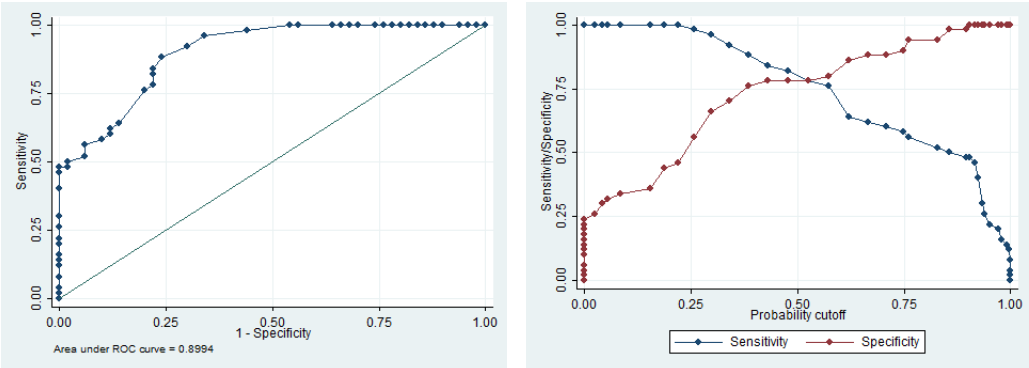
Table 5: Predicted probabilities and real outcomes in Vote II

Here, again for predicting concrete outcomes instead of probabilities the threshold of 0.5 seems suitable. 78% of the votes are correctly predicted. But the total number "Guilty" and "Not-Guilty" Votes is right, it predicts a 50/50 share. All in all 22 outcomes are predicted wrongly but the estimation meets the real outcome in numbers. Changing the cut-off threshold to 0.35 yields a higher hit-rate, i.e the behaviour of more senators is correctly predicted, but the predicted outcome is then 61 "Guilty"/39 "Not-Guilty" which is far off the real outcome.

In Figure 7(a) sees that the area under the ROC curve is smaller here, than previously. In

[3]ME are : 0.042 for ClintVote; 0.95 for first_term and 0.98 for firstparty

Figure7(b) one sees that the sum of of Sensitivity and Specificity is maximised at a cut-off value of slightly above 0.5.



(a) ROC Curve of Vote II (b) Sensitivity and Specificity in Vote II

Figure 7: ROC, Sensitivity and Specificity

In a probit model Pearson residuals cannot be estimated, therefore I take deviance residuals here. They are differently constructed that Pearson residuals, but naturally also centred around zero and graphically they should look similar. In Figure 8 one sees that residuals and predicted values are not perfectly independent. There is hardly any structure but one sees that residuals are very close to zero for predicted values close to 0 or one, and a more spread around the middle.

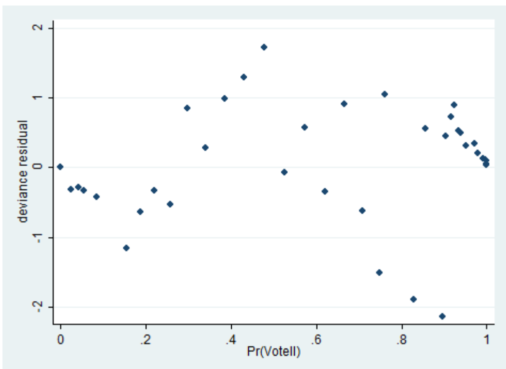
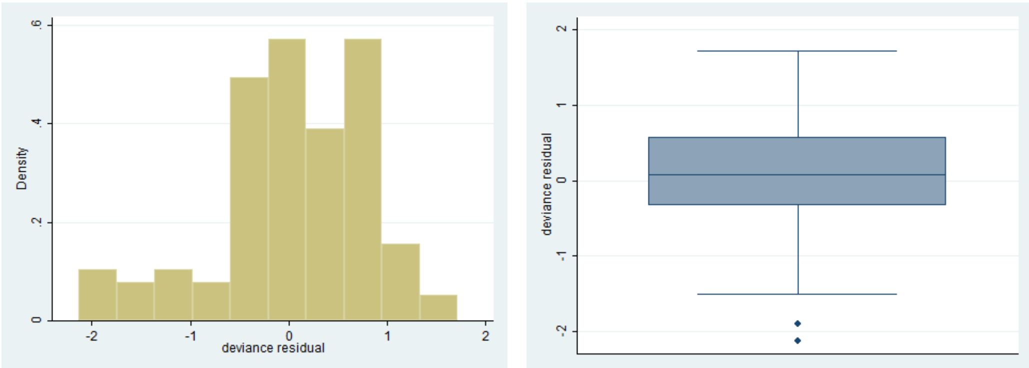


Figure 8: Residuals vs. Predicted probabilities

Figure 9 shows the distribution of the residuals. They are centred around zero, and slightly skewed. Most outliers are negative.



(a) Histogram of residuals (b) Boxplot of residals

Figure 9: Residual distribution

2.3 Joint Analysis

2.3.1 Choosing a Model

Modelling the outcome of the two (or more) binary response variables can be done either by a bivariate model, or by summing up the two variables and perform a ordered response model. In this case a bivariate model does not work because the Likelihood function is not globally concave any more and there is no (unique) maximum. Therefore I use the sum of the variables *Vote I* and *Vote II*,which is called *NVote*. *NVote* takes the values 0, 1 and 2. Clearly there is a natural ordering, so a orderd response model is suitable.
Table 6 shows different modelling steps for the joint outcome. Columns (1) - (3) are estimated by ordered probit and column (4) by ordered logit. Many variables could not be used because they cause predict too exact and don't allow Stata to produce outputs. In other cases, e.g. when using *firstparty* and *first_term* simultaneously Stata warns that the marginal effects are biased. Including those variables individually shows no significance. So as in the case of the Vote on Obstruction of Justice the decision is whether one claims the underlying distribution to be logistic or normal. As for any ordered response model one cannot use Pearson-goodness-of-fit, the decision has to be made, again, on the basis of information criteria and the Pseudo-R^2. This time all of them are in favour of the logistic

model.

	(1) NVote	(2) NVote	(3) NVote	(4) NVote
NVote				
Conserv	0.0697***	0.0691***	0.0693***	0.137***
	(0.0107)	(0.0114)	(0.0106)	(0.0297)
first_term	-0.249			
	(0.449)			
firstparty		0.0318		
		(0.529)		
cut1				
_cons	3.312***	3.398***	3.399***	6.675***
	(0.517)	(0.566)	(0.563)	(1.559)
cut2				
_cons	4.122***	4.189***	4.191***	8.212***
	(0.586)	(0.613)	(0.609)	(1.655)
N	100	100	100	100
pseudo R^2	0.735	0.733	0.733	0.738
AIC	53.34	53.64	51.64	50.89
BIC	63.76	64.06	59.46	58.71

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 6: Modelling the sum of the Votes

2.3.2 Interpretation and Model Diagnostics

Model diagnostics and analysis works a little different when dealing with multinomial response models. The sign of the coefficient, however, can still be interpreted as the direction of the change when the variable increases by one unit. The model I chose just has one input, namely the degree of conservatism. One has to be even more careful when interpreting the marginal effects.

The ME have to be computed individually for each outcome. Table 7 shows the Marginal effect of *Conserv* for all three possible outcomes of *NVote*. The ME can be interpreted in the following way: When the degree of conservationism increases by one unit for the average senator he or she will be by 3.38 percentage points less likely to vote "Not Guilty"

NVote	Marginal effect
0	-0.0338
1	0.016
2	0.0172

Table 7: Marginal effects of *Conserv* for average senator

in both cases, by 1.6 percentage points more likely to vote "Guilty" in one vote and "Not Guilty" in the other vote, and the probability to vote "Guilty" in both cases will rise by 1.7 percentage points.

Estimating the model, one does not only get one success-probability, but one for each category. Table 8 shows a summary statistic for the predicted probabilities and compares it to the outcomes for the sum of Votes. The probabilities for the categories *Nvote*=0 and *Nvote*=2 are very close to the real variables. Just for the few republican senators, that voted "Guilty" only in one Vote the probabilities are not covering the full range. The maximal probability predicted for voting "Guilty" in just one of the votes equals 36%. This results in the case that the outcome *NVote*=1 will never be predicted.

Variable	Mean	Std. Dev.	Min.	Max.
p0	0.502	0.471	0.001	0.999
N0	0.5	0.503	0	1
p1	0.057	0.094	0.001	0.363
N1	0.05	0.219	0	1
p2	0.441	0.451	0	0.996
N2	0.45	0.5	0	1
N		100		

Table 8: Predicted probabilitied and real outcomes for the sum of Votes

It is not possible with the methods we used in this course to estimate residuals or to perform a goodness-of-fit test. It is also hard to predict exact outcomes, and therefore one cannot easily observe sensitivity and specificity .

2.4 Predictive Power

I will now take a closer look on the predictive power of the models an the data. I therefore create a random variable, that is either 0 or 1. It takes the value 1 in 80% of the cases. Then I estimate the same models as above, but just using these 80% of the data and predict the probabilities of voting guilty in the different impeachments for all senators. Then I will compare the estimated probabilities. In table 9 one sees a summary statistic of the predicted probabilities from the out-of-sample-prediction, compared to the predicted probabilities in the full sample prediction and the real voting outcomes. The table just looks at the 20% of the sample that have not been used, when estimating the out-of sample prediction.

Variable	Mean	Std. Dev.	Min.	Max.
sample_phatI	0.295	0.437	0	0.963
phatI	0.291	0.431	0	0.951
VoteI	0.263	0.452	0	1
sample_phatII	0.494	0.337	0	0.971
phatII	0.455	0.321	0	0.934
VoteII	0.316	0.478	0	1
sample_p0	0.649	0.448	0	0.999
p0	0.657	0.444	0.001	0.999
N0	0.684	0.478	0	1
sample_p1	0.058	0.094	0.001	0.287
p1	0.056	0.087	0.001	0.258
N1	0.053	0.229	0	1
sample_p2	0.293	0.428	0	0.997
p2	0.287	0.42	0	0.996
N2	0.263	0.452	0	1

Table 9: Out-of-sample predicted probabilites, full sample predicted probabilies and real outcomes of the Votes for those senators, that have not been used for out-of-sample prediction

Figure 10 in the Appendix shows scatter plots of the out-of-sample predictions vs. full-sample-predictions for all models. For these plots all observations have been used.

One sees that the predicted probabilities differ, depending on what sample was used to estimate them, but they are still pretty close to the 45° line. All observation lying on the

45° line would indicate that the predicted probabilities are the same for both samples.

2.5 The Role of Degree of Conservatism

Conservatism is obviously the most relevant variable in this data set. It is the only significant variable for the sum of the votes. It is basically the only input for Vote I. And the outcome of the Vote on Obstruction of Justice is perfectly predicted for any value of *Conserv* greater than 48. As mentioned in the beginning, it is highly correlated to the party of a senator, every senator that has a higher degree of conservatism than 40 is a Republican, and every senator that is less conservative than 20 is a Democrat. The votes in the senate are way more influenced by the party-line than by the decision of the individuals. and even if they decide to vote according to their own opinion they risk not to be re-elected after their period in the senate is over. Also when they pursue other political goals they might not want to risk a disagreement with their party.

The Interesting question is how *Conserv* enters the regressions. In the vote on perjury it appears also in its squared form and improves the model. One could also try to let it enter the regressions in categories. Including conservatism in terms of categories would indicate the change in conservatism is just important if it crosses certain thresholds. Table 10 in the Appendix shows a sample of regression outputs using two different categorical variables of *Conserv* in the logit regression of VoteI and the sum of votes. The variable *catcon* consist of four different equally large categories, and the other variable *cat2con* consists of 3 categories. The first one for all observations where the degree of conservatism is smaller or equal 20, the second one for all between 20 and 40, and the third for all larger or equal to 40. The model in Column (2) of table 10 has better values for the information criteria and a higher R^2, but Stata warns that the standard errors might be biased. The decision between those models is trade-off on fitting the data better and correct inference. Depending on what is the aim one has to decide. As the model in column (1) still has a very good fit on the data it might be preferable to have correct inference.

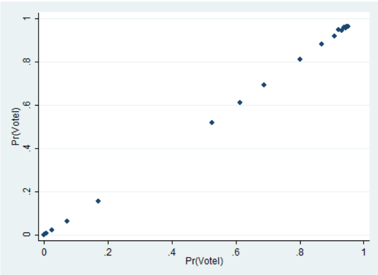
Using categorical variables in Vote I or Vote II causes a lot of perfect predictions and computational errors.

3 Conclusion

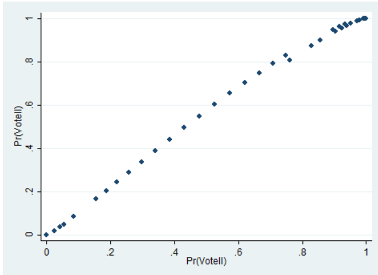
Throughout the paper it is clear that the outcome of the votes is mainly determined by the degree of conservatism and the party a senator belongs to. I analysed the impacts of various factor on the outcome of the votes and found that the more conservative a senator is the more likely he is going to vote guilty in both impeachments against Clinton. In general the voting behaviour of the senator can be very well explained using the models constructed in this paper. However, ten Republican senators voted not-guilty in the perjury impeachment, five of them also found him not-guilty in the Obstruction of Justice impeachment. Those five have very low conservatism levels compared to other average Republican, so *Conserv* can be a good explanation here. The other five (the ones that voted mixed) have, however, quite high levels of conservatism. In general there are always outliers that do not fit the model, and especially when dealing with human behaviour one can never catch all facets as some people might pursue different, individual goals or simply make mistakes in their choices (They might not get the question right, or press by accident a button they did not mean to press).
After estimating the different models I also took a closer look on how stable the results are across the cross-section of senators. Taking only a part of data and using this to predict the voting behaviour of the rest of the senators works quite well. This indicates that the the model is quite good on the one hand and also that the data is homogeneously distributed. I also took a closer look on different ways to include *Conserv* and saw that it might have some advantages to include it categorically but that for the sake of correct standard errors it is preferably to model it the way I did in the first part of the paper. All in all, from a statistical point of view one can say that the senators voting was pretty predictable.

4 Appendix

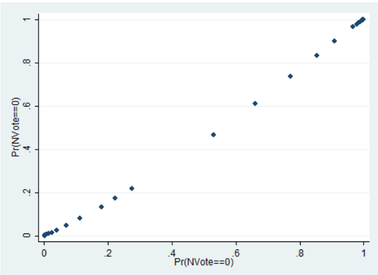
4.1 Out-of-Sample-Prediction



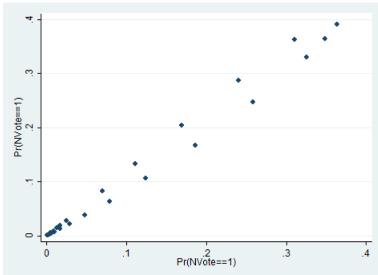
(a) Vote I



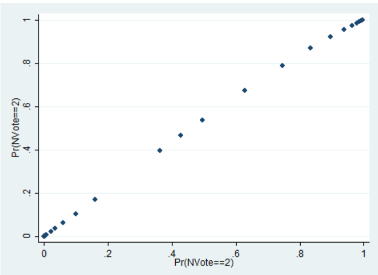
(b) VoteII



(c) Sum of votes, NVote=0



(d) Sum of votes, NVote=1



(e) Sum of votes, NVote=2

Figure 10: Out-of-sample vs. full-sample predictions

4.2 Categorical conservatism

	(1) NVote	(2) NVote	(3) NVote	(4) VoteI	(5) VoteI
main					
Conserv	0.137***			0.525**	
	(0.0297)			(0.171)	
25.catcon		0			0
		(.)			(.)
50.catcon		0.000000128			0
		(41208.7)			(.)
75.catcon		43.06			-0.389
		(14768.8)			(0.968)
100.catcon		43.45			0
		(14768.8)			(.)
1.cat2con			0		
			(.)		
2.cat2con			-5.47e-10		
			(5142.7)		
3.cat2con			20.91		
			(1516.5)		
sqrdConserv				-0.00302**	
				(0.00106)	
_cons				-19.81**	2.335***
				(6.418)	(0.605)
cut1					
_cons	6.675***	22.92	18.39		
	(1.559)	(14275.1)	(1516.5)		
cut2					
_cons	8.212***	41.11	19.30		
	(1.655)	(14768.8)	(1516.5)		
N	100	100	100	100	50
pseudo R^2	0.738	0.811	0.643	0.764	0.005
AIC	50.89	42.35	69.03	38.42	36.35
BIC	58.71	55.38	79.45	46.24	40.17

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 10: regression outputs using categorical variables