Causal Inference and Machine Learning Homework 3

Andrew Bennett

March 4, 2024

1 Exercise 1

1.1 Estimate (1) using sex, age, educ and work and both Logit and Probit models. Interpret the odds ratio.

Logit Model

Dep. Variable:	is_voting	**No. Observations:**	2822
Model:	Logit	**Df Residuals:**	2817
Method:	MLE	**Df Model:**	4
Date:	Mon, 04 Mar 2024	**Pseudo R-squ.:**	0.05321
Time:	18:39:32	**Log-Likelihood:**	-478.74
converged:	True	**LL-Null:**	-505.64
Covariance Type:	nonrobust	**LLR p-value:**	5.770e-11

	coef	**std err**	**z**	**P> |z|**	**[0.025**	**0.975]**
const	1.2111	0.470	2.576	0.010	0.290	2.133
sex	-0.2838	0.190	-1.496	0.135	-0.656	0.088
age	0.0136	0.006	2.429	0.015	0.003	0.025
educ	0.5605	0.094	5.992	0.000	0.377	0.744
work	-0.1838	0.077	-2.372	0.018	-0.336	-0.032

Probit Model

Dep. Variable:	is_voting	**No. Observations:**	2822
Model:	Probit	**Df Residuals:**	2817
Method:	MLE	**Df Model:**	4
Date:	Mon, 04 Mar 2024	**Pseudo R-squ.:**	0.05280
Time:	18:41:18	**Log-Likelihood:**	-478.94
converged:	True	**LL-Null:**	-505.64
Covariance Type:	nonrobust	**LLR p-value:**	7.027e-11

	coef	**std err**	**z**	**P> |z|**	**[0.025**	**0.975]**
const	0.8722	0.219	3.992	0.000	0.444	1.301
sex	-0.1229	0.087	-1.413	0.158	-0.293	0.048
age	0.0064	0.003	2.328	0.020	0.001	0.012
educ	0.2472	0.041	6.071	0.000	0.167	0.327
work	-0.0913	0.036	-2.547	0.011	-0.162	-0.021

The sex odds ratio is approximately 0.753. This means that the odds of voting for females over males decrease by about 24.7The odds ratio is approximately 1.014. For each additional year of age, the odds of voting increase by about 1.4%. The odds ratio is approximately 1.752. With each additional unit of education, the odds of voting increase by about 75.2%. The odds ratio is approximately 0.832. For each unit increase in the work variable, the odds of voting decrease by about 16.8%.

1.2 Provide marginal effects for sex and age variables at the mean. Calculate the probability of voting for a male 26-year-old working individual with an intermediate level of education.

Dep. Variable:	is_voting					
Method:	dydx	**dy/dx**	**std err**	**z**	**P> |z|**	**[0.025**	**0.975]**
At:	mean					
sex	-0.0096	0.006	-1.505	0.132	-0.022	0.003
age	0.0005	0.000	2.444	0.015	9.11e-05	0.001
educ	0.0190	0.003	6.863	0.000	0.014	0.024
work	-0.0062	0.003	-2.383	0.017	-0.011	-0.001

In the marginal effects table, the first column is the marginal effect which shows the amount of change in the probability of voting for a one-unit change in the independent variable. With sex at a value of -0.0096 that means a 1 unit increase in sex will decrease the probability of voting by 0.0096. In other words, if it is a female the probability of voting will decrease by .96%. Similarly a 1 unit increase in age increases voting probability by 0.05%.

Dep. Variable:	is_voting					
Method:	dydx	**dy/dx**	**std err**	**z**	**P> |z|**	**[0.025**	**0.975]**
At:	mean					
sex	-0.0098	0.007	-1.417	0.157	-0.023	0.004
age	0.0005	0.000	2.341	0.019	8.25e-05	0.001
educ	0.0196	0.003	6.502	0.000	0.014	0.026
work	-0.0072	0.003	-2.564	0.010	-0.013	-0.002

With sex at a value of -0.0098 that means a 1 unit increase in sex will decrease the probability of voting by 0.0098. This means the probit model is slightly more sensitive to changes in sex than the logit model. Similarly a 1 unit increase in age increases voting probability by 0.05% which is about the same sensitivity. Overall, the probit model is slightly more sensitive to changes in the independent variables than the logit model.

Now at a specified interval.

To calculate the marginal effects at a specific value of the independent variables we would mathematically calculate the marginal effects using the formula

$$marginaleffects = model.coeffs * model.cdf(X) * (1 - cdf(X)) * X$$

where X is the value of the independent variables at which we want to calculate the marginal effects.

Logit model probability of voting for a Full-time employed male, aged 26 with an intermediate level of education: 0.9553205244353641

Probit model probability of voting for a Full-time employed male, aged 26 with an intermediate level of education: 0.9543561587438019

The logit model predicts a higher probability of voting than the probit model, this may be due to the fact that the probit model is more sensitive to changes in the independent variables than the logit model.

1.3 Re-estimate the Logit model without pre-programmed functions (except the NewtonRhapson). You can program either in Julia/Matlab/Python/R or in Stata using the ml command.

To create a logit model, we first create a logistic transformation function to transform the independent variables into probabilities. Next, we calculate the gradient of the logistic transformation function to find the maximum likelihood estimates of the coefficients. Then, we use a hessian operation to find the curvature of the logistic transformation function which allows us to find the steepness of the curve.

Finally, we create the coefficients by dividing the gradient by the hessian. Which is essentially the slope of the curve. Each iteration will have a new set of beta coefficients which will multiply with the independent variables to create a new set of probabilities. Each iteration will continue until the difference between the new and previous beta coefficients is less than a certain threshold.

2 Exercise: Multinomial Models

2.1 Consider only the three largest parties: CDU/CSU, SPD and Die Grüne. (You can drop individuals who have not voted or voted for another party.) Using Model (2) with number of categories m = 3 and multinomial logit, estimate the impact of the same variables as in the previous exercise on the probability of choosing particular party. Interpret the estimates.

Dep. Variable:	pv01	**No. Observations:**	1831
Model:	MNLogit	**Df Residuals:**	1821
Method:	MLE	**Df Model:**	8
Date:	Mon, 04 Mar 2024	**Pseudo R-squ.:**	0.03407
Time:	16:46:47	**Log-Likelihood:**	-1903.3
converged:	True	**LL-Null:**	-1970.5
Covariance Type:	nonrobust	**LLR p-value:**	3.666e-25

pv01=2	**coef**	**std err**	**z**	**P> |z|**	**[0.025**	**0.975]**
const	-0.9244	0.341	-2.709	0.007	-1.593	-0.255
sex	0.0719	0.112	0.642	0.521	-0.147	0.291
age	0.0037	0.004	0.994	0.320	-0.004	0.011
educ	0.0239	0.049	0.483	0.629	-0.073	0.121
work	0.0674	0.047	1.433	0.152	-0.025	0.160
pv01=4	**coef**	**std err**	**z**	**P> |z|**	**[0.025**	**0.975]**
const	-1.9668	0.368	-5.341	0.000	-2.689	-1.245
sex	0.3995	0.120	3.331	0.001	0.164	0.635
age	-0.0050	0.004	-1.320	0.187	-0.013	0.002
educ	0.3765	0.052	7.185	0.000	0.274	0.479
work	-0.1144	0.049	-2.344	0.019	-0.210	-0.019

All values provided in the tables are compared with the reference category which is the CDU/CSU party. Sex and education are the highest predictors of choosing The Greens party over the reference party. The same independent variables have much smaller magnitudes in the SPD party, which indicates that the independent variables are better predictors of The Greens vs SPD. This is corroborated by the P-values of the coefficients for SPD vs The Greens, which indicate that the independent vars for SPD are not very significant.

2.2 Compute marginal effects of the estimates for female for the median voter.

Dep. Variable:	pv01						
Method:	dydx	**pv01=1**	**dy/dx**	**std err**	**z**	**P> |z|**	**[0.025**	**0.975]**
At:	median						
	sex	-0.0490	0.024	-2.023	0.043	-0.097	-0.002
	age	-9.38e-05	0.001	-0.119	0.906	-0.002	0.001
	educ	-0.0394	0.011	-3.661	0.000	-0.061	-0.018
	work	0.0004	0.010	0.039	0.969	-0.019	0.020
	pv01=2	**dy/dx**	**std err**	**z**	**P> |z|**	**[0.025**	**0.975]**
	sex	-0.0087	0.023	-0.384	0.701	-0.053	0.036
	age	0.0011	0.001	1.486	0.137	-0.000	0.003
	educ	-0.0177	0.010	-1.745	0.081	-0.038	0.002
	work	0.0215	0.009	2.320	0.020	0.003	0.040

pv01=4	dy/dx	std err	z	P> |z|	[0.025	0.975]
sex	0.0578	0.015	3.759	0.000	0.028	0.088
age	-0.0010	0.001	-1.837	0.066	-0.002	6.76e-05
educ	0.0571	0.007	8.428	0.000	0.044	0.070
work	-0.0219	0.007	-3.037	0.002	-0.036	-0.008

2.3 For which party would you expect the largest gain in voting share as the population becomes older on average?

I would expect the SPD party (pv01=2) to increase the most given an aging population. This is due to the fact that in the marginal effects table, there is a specified 0.0011 increase in voting probability given every additional year of age. This is significantly higher than the other marginal effects for age which are in fact negatively affected by age increases.

2.4 Re-estimate the previous model in Julia/Matlab/Python/R without pre-programmed functions or in Stata using the ml command.

Essentially, this is the same process as the logit model; however, in my Python implementation a few things changed. First y becomes dummied to create a column of 1s and 0s for each class (in this case 3). Then we make a zeroed out array of beta coefficients. Then we take the linear combination of the covariates and the parameters. In my implementation, I then use a softmax function to convert the linear combinations into probabilities. Afterwards, depending on the implementation, I calculate a cost function, gradient. Finally, I compute the updated weights depending on the learning rate and gradient. This occurs iteratively until the loss function is minimized or the epoch limit is reached.

	CDU-CSU	SPD	The Greens
const	0.332	-0.113	-0.219
sex	-0.104	-0.186	0.291
age	0.083	-0.007	-0.076
educ	-0.268	-0.339	0.606
work	0.205	0.314	-0.519

Table 1: Coefficients for late outcomes

However, this model does not produce the same results as the Stata model, so I also implemented this in Stata. The stata program takes three inputs: log likelihood variable, and 2 independent variables. Predicted probabilities are stored for each potential outcome. These probabilities are determined using the logistic function for each outcome, based on the linear predictors xb1 and xb2. Subsequently, the log-likelihood variable is calculated utilizing the predicted probabilities alongside and the actual dependent variables.

	Coefficient	Std. err.	z	P> |z|	[95% conf.	interval]
SPD						
sex	0.0719	0.1119	0.64	0.521	-0.1475	0.2913
age	0.0037	0.0037	0.99	0.320	-0.0036	0.0109
educ	0.0239	0.0495	0.48	0.629	-0.0730	0.1208
work	0.0674	0.0470	1.43	0.152	-0.0248	0.1596
_cons	-0.9244	0.3413	-2.71	0.007	-1.5933	-0.2555
Grüne						
sex	0.3995	0.1199	3.33	0.001	0.1645	0.6346
age	-0.0050	0.0038	-1.32	0.187	-0.0125	0.0024
educ	0.3765	0.0524	7.19	0.000	0.2738	0.4791
work	-0.1144	0.0488	-2.34	0.019	-0.2101	-0.0187
_cons	-1.9668	0.3683	-5.34	0.000	-2.6886	-1.2451

Table 2: Coefficients for SPD and Grüne

2.4.1 Include alternative varying variables into Model (2). cu-assess, spd-assess and gre-assess measure how closely the party's position on the left-to-right spectrum aligns with the respondents own position.

After a long fight with these Python conditional logit model packages: torch-choice, pylogit, statsmodel.ConditionalLogit, and biogeme. I was not able to reproduce comparable Stata results and at this point I switched to the dark side. The correct Stata results are included on the following page. The coefficient of the alternative varying variable is -.55. This means that for a one unit increase in the party alignment, indicates decrease in probability of .55 in choosing the alternatives.

Table 3: Coefficients for Alternative-specific conditional logit

	Coefficient	Std. err.	z	P> |z|	[95% conf. interval]	
parties						
distance	-0.5489	0.0292	-18.79	0.000	-0.6062	-0.4916
CDU_CSU	(base alternative)					
SPD						
sex (FEMALE)	0.0410	0.1346	0.30	0.761	-0.2228	0.3048
age	0.0026	0.0043	0.60	0.549	-0.0058	0.0110
educ (LOWEST LEVEL)	0.2626	0.7544	0.35	0.728	-1.2161	1.7413
educ (INTERMEDIARY LEVEL)	0.1937	0.7522	0.26	0.797	-1.2806	1.6680
educ (QUALI.UNIV.APPL.SCI.)	0.2048	0.7746	0.26	0.791	-1.3134	1.7231
educ (QUALI.FOR UNIVERSITY)	-0.0081	0.7530	-0.01	0.991	-1.4839	1.4677
educ (OTHER SCHOOL CERTIF.)	-12.1970	706.3033	-0.02	0.986	-1396.5260	1372.1320
educ (STILL AT SCHOOL)	0.8217	1.0599	0.78	0.438	-1.2557	2.8990
work (EMPLOYED, PART-TIME)	-0.0274	0.2330	-0.12	0.906	-0.4840	0.4292
work (LESS THAN PART-TIME)	0.8702	0.3789	2.30	0.022	0.1277	1.6127
work (NOT WORKING)	0.1554	0.1645	0.94	0.345	-0.1671	0.4779
_cons	-0.9259	0.7687	-1.20	0.228	-2.4326	0.5808
GREENS						
sex (FEMALE)	0.1830	0.1451	1.26	0.207	-0.1014	0.4674
age	-0.0062	0.0045	-1.36	0.174	-0.0151	0.0027
educ (LOWEST LEVEL)	-0.5692	0.7851	-0.72	0.468	-2.1080	0.9696
educ (INTERMEDIARY LEVEL)	0.0028	0.7742	0.00	0.997	-1.5146	1.5203
educ (QUALI.UNIV.APPL.SCI.)	0.2802	0.7954	0.35	0.725	-1.2788	1.8392
educ (QUALI.FOR UNIVERSITY)	0.3603	0.7719	0.47	0.641	-1.1526	1.8732
educ (OTHER SCHOOL CERTIF.)	0.2970	2.0153	0.15	0.883	-3.6528	4.2469
educ (STILL AT SCHOOL)	0.3651	1.1458	0.32	0.750	-1.8806	2.6108
work (EMPLOYED, PART-TIME)	0.0866	0.2192	0.39	0.693	-0.3430	0.5162
work (LESS THAN PART-TIME)	0.8494	0.3843	2.21	0.027	0.0963	1.6026
work (NOT WORKING)	-0.4449	0.1758	-2.53	0.011	-0.7895	-0.1002
_cons	-0.3477	0.7885	-0.44	0.659	-1.8931	1.1978

2.5 Estimate a Multinomial Logit model. Include the regressors pa01 (respondent's own position on the left-right spectrum), sex, age, educ and work. Briefly comment on the coefficients (4-5 sentences).

Multinomial Logistic Regression Results

Variable	Category	Coefficient	Std. err.	z	P> |z|	[95% conf.	interval]
CDU_CSU							
SPD							
	pa01	-0.185	0.026	-7.03	0.000	-0.236	-0.133
	FEMALE	-0.009	0.119	-0.07	0.941	-0.242	0.224
	age	0.006	0.004	1.52	0.129	-0.002	0.014
	LOWEST LEVEL	0.598	0.599	1.00	0.318	-0.576	1.772
	INTERMEDIARY LEVEL	0.521	0.597	0.87	0.383	-0.649	1.691
	QUALI.UNIV.APPL.SCI.	0.644	0.621	1.04	0.300	-0.573	1.861
	QUALI.FOR UNIVERSITY	0.520	0.598	0.87	0.385	-0.653	1.693
	OTHER SCHOOL CERTIF.	-14.115	699.955	-0.02	0.984	-1386.001	1357.771
	STILL AT SCHOOL	1.397	0.892	1.57	0.117	-0.351	3.145
	EMPLOYED, PART-TIME	-0.033	0.208	-0.16	0.874	-0.441	0.375
	LESS THAN PART-TIME	0.811	0.330	2.46	0.014	0.164	1.457
	NOT WORKING	0.095	0.147	0.65	0.517	-0.193	0.384
	$_cons$	-0.455	0.630	-0.72	0.470	-1.690	0.780
FDP							
	pa01	0.050	0.040	1.26	0.209	-0.028	0.128
	FEMALE	-0.132	0.151	-0.87	0.385	-0.428	0.165
	age	-0.000	0.005	-0.09	0.929	-0.010	0.009
	LOWEST LEVEL	0.448	0.793	0.56	0.572	-1.107	2.003
	INTERMEDIARY LEVEL	0.511	0.789	0.65	0.517	-1.035	2.056
	QUALI.UNIV.APPL.SCI.	1.089	0.805	1.35	0.176	-0.490	2.667
	QUALI.FOR UNIVERSITY	0.619	0.789	0.78	0.433	-0.927	2.165
	OTHER SCHOOL CERTIF.	0.955	1.399	0.68	0.495	-1.786	3.696
	STILL AT SCHOOL	0.387	1.384	0.28	0.780	-2.326	3.100
	EMPLOYED, PART-TIME	-0.337	0.261	-1.29	0.197	-0.849	0.175
	LESS THAN PART-TIME	1.156	0.341	3.39	0.001	0.488	1.823
	NOT WORKING	-0.671	0.190	-3.54	0.000	-1.043	-0.300
	$_cons$	-1.577	0.840	-1.88	0.060	-3.223	0.069
THE_GREENS							
	pa01	-0.174	0.028	-6.31	0.000	-0.228	-0.120
	FEMALE	0.249	0.127	1.96	0.050	-0.000	0.498
	age	-0.004	0.004	-0.96	0.338	-0.012	0.004
	LOWEST LEVEL	-0.180	0.678	-0.27	0.790	-1.510	1.149
	INTERMEDIARY LEVEL	0.343	0.668	0.51	0.608	-0.967	1.653
	QUALI.UNIV.APPL.SCI.	0.820	0.686	1.20	0.232	-0.525	2.166
	QUALI.FOR UNIVERSITY	1.043	0.666	1.57	0.117	-0.261	2.348
	OTHER SCHOOL CERTIF.	-0.826	1.458	-0.57	0.571	-3.684	2.032
	STILL AT SCHOOL	1.047	0.984	1.06	0.287	-0.882	2.975
	EMPLOYED, PART-TIME	0.143	0.192	0.75	0.456	-0.233	0.519
	LESS THAN PART-TIME	0.770	0.322	2.40	0.017	0.140	1.401
	NOT WORKING	-0.386	0.154	-2.50	0.013	-0.688	-0.083
	$_cons$	-0.098	0.697	-0.14	0.888	-1.464	1.267
THE_LEFT							
	pa01	-0.265	0.028	-9.41	0.000	-0.320	-0.209
	FEMALE	-0.095	0.153	-0.63	0.531	-0.394	0.203
	age	-0.007	0.005	-1.45	0.147	-0.016	0.002
	LOWEST LEVEL	-0.720	0.627	-1.15	0.251	-1.949	0.510
	INTERMEDIARY LEVEL	-0.128	0.613	-0.21	0.834	-1.330	1.073

	QUALI.UNIV.APPL.SCI.	-0.108	0.650	-0.17	0.868	-1.382	1.165
	QUALI.FOR UNIVERSITY	0.166	0.610	0.27	0.785	-1.029	1.362
	OTHER SCHOOL CERTIF.	-1.831	1.561	-1.17	0.241	-4.889	1.228
	STILL AT SCHOOL	0.137	1.000	0.14	0.891	-1.824	2.097
	EMPLOYED, PART-TIME	-0.323	0.277	-1.16	0.245	-0.866	0.221
	LESS THAN PART-TIME	1.141	0.356	3.21	0.001	0.444	1.839
	NOT WORKING	0.168	0.179	0.94	0.346	-0.182	0.519
	$_cons$	0.520	0.645	0.81	0.419	-0.743	1.784
AFD							
	pa01	0.436	0.047	9.27	0.000	0.344	0.528
	FEMALE	-0.276	0.169	-1.63	0.104	-0.608	0.056
	age	-0.016	0.005	-3.13	0.002	-0.027	-0.006
	LOWEST LEVEL	1.402	1.065	1.32	0.188	-0.685	3.489
	INTERMEDIARY LEVEL	1.581	1.061	1.49	0.136	-0.499	3.661
	QUALI.UNIV.APPL.SCI.	1.196	1.087	1.10	0.271	-0.935	3.327
	QUALI.FOR UNIVERSITY	0.652	1.070	0.61	0.542	-1.445	2.749
	OTHER SCHOOL CERTIF.	3.169	1.389	2.28	0.022	0.448	5.891
	STILL AT SCHOOL	1.333	1.393	0.96	0.338	-1.397	4.063
	EMPLOYED, PART-TIME	-0.308	0.303	-1.02	0.309	-0.902	0.286
	LESS THAN PART-TIME	0.809	0.443	1.83	0.068	-0.059	1.677
	NOT WORKING	-0.222	0.197	-1.13	0.258	-0.608	0.163
	$_cons$	-3.869	1.115	-3.47	0.001	-6.055	-1.684

One notable coefficient of interest is the female demographic in the Green and AFD parties. In the Greens, females are more represented and positively correlated with voting, whereas AFD has strong negative correlation. Additionally, education levels are interesting across the parties. FDP and AFD have generally stronger correlation with voting across all levels of education. However, education in the SPD party is also unusually negatively correlated. Specifically, school certifications are extremely negatively correlated with SPD party voting. In the AFD party, where female voters are less likely to vote, a positive correlation of certificate education voters, is found.

2.6 Estimate a Nested Logit, with voters choosing between left and right in the first level, and then choosing between parties in each branch.

article longtable booktabs

Table 5: SPD Regression Results

	Variable	**Coefficient**	**Std. err.**	**z**	**P¿—z—**	**[95% Conf. Interval]**
2***sex**	FEMALE	-0.0576316	94.45797	-0.00	1.000	-185.1918 to 185.0766
	age	0.0081437	4.218832	0.00	0.998	-8.260615 to 8.276902
7***educ**	LOWEST LEVEL	0.6789934	-	-	-	-
	INTERMEDIARY LEVEL	0.2633187	238.6081	0.00	0.999	-467.4 to 467.9266
	QUALI.UNIV.APPL.SCI.	0.1787069	80.03004	0.00	0.998	-156.6773 to 157.0347
	QUALI.FOR UNIVERSITY	-0.079457	225.3046	-0.00	1.000	-441.6684 to 441.5094
	OTHER SCHOOL CERTIF.	-37.71708	6.81e+07	-0.00	1.000	-1.33e+08 to 1.33e+08
	STILL AT SCHOOL	0.5563524	166.2829	0.00	0.997	-325.3522 to 326.4649
3***work**	EMPLOYED, PART-TIME	-0.0182825	-	-	-	-
	LESS THAN PART-TIME	-0.1222906	167.9967	-0.00	0.999	-329.3897 to 329.1451
	NOT WORKING	0.1501312	39.59536	0.00	0.997	-77.45535 to 77.75561
_cons		-1.452334	164.9812	-0.01	0.993	-324.8096 to 321.9049

article booktabs

In order to have a good nested logit model, a few things are key. Firstly is the nested structure. The nested structure should be oriented such that the nests have similar characteristics. The initial split was done based on Left and Right party lines, and the other independent variables look well distributed among the classes.

Table 6: FDP Regression Results

	Variable	Coefficient	Std. err.	z	P¿—z—	[95% Conf. Interval]
2***sex**	FEMALE	-0.1429811	39.26923	-0.00	0.997	-77.10926 to 76.8233
	age	0.0000365	0.0115597	0.00	0.997	-0.02262 to 0.0226931
7***educ**	LOWEST LEVEL	0.4648729	127.6776	0.00	0.997	-249.7786 to 250.7084
	INTERMEDIARY LEVEL	0.5279594	145.0035	0.00	0.997	-283.6736 to 284.7296
	QUALI.UNIV.APPL.SCI.	1.214095	333.4448	0.00	0.997	-652.3256 to 654.7538
	QUALI.FOR UNIVERSITY	0.6345372	174.2739	0.00	0.997	-340.9361 to 342.2051
	OTHER SCHOOL CERTIF.	0.8201027	225.2416	0.00	0.997	-440.6453 to 442.2855
	STILL AT SCHOOL	0.7235613	198.7276	0.00	0.997	-388.7755 to 390.2226
3***work**	EMPLOYED, PART-TIME	-0.4003481	109.9535	-0.00	0.997	-215.9052 to 215.1045
	LESS THAN PART-TIME	1.294616	355.5582	0.00	0.997	-695.5867 to 698.1759
	NOT WORKING	-0.7827317	214.9724	-0.00	0.997	-422.1209 to 420.5554
_cons		-1.476238	405.4404	-0.00	0.997	-796.1248 to 793.1723

Table 7: GREENS Regression Results

	Variable	Coefficient	Std. err.	z	P¿—z—	[95% Conf. Interval]
2***sex**	FEMALE	0.1836712	.	.	.	.
	age	-0.0026337	.	.	.	.
7***educ**	LOWEST LEVEL	-0.1338158	318.1748	-0.00	1.000	-623.7451 to 623.4774
	INTERMEDIARY LEVEL	0.0978913	173.8525	0.00	1.000	-340.6467 to 340.8425
	QUALI.UNIV.APPL.SCI.	0.3831401	.	.	.	.
	QUALI.FOR UNIVERSITY	0.4961037	.	.	.	.
	OTHER SCHOOL CERTIF.	1.238095	90.94429	0.01	0.989	-177.0094 to 179.4856
	STILL AT SCHOOL	0.1315739	.	.	.	.
3***work**	EMPLOYED, PART-TIME	0.2011023	85.87837	0.00	0.998	-168.1174 to 168.5196
	LESS THAN PART-TIME	-0.131133	171.458	-0.00	0.999	-336.1826 to 335.9203
	NOT WORKING	-0.3426886	232.509	-0.00	0.999	-456.052 to 455.3666
_cons		-1.030878	.	.	.	.

Table 8: LEFT Regression Results

	Variable	Coefficient	Std. err.	z	P¿—z—	[95% Conf. Interval]
2***sex**	FEMALE	-0.134383	124.5022	-0.00	0.999	-244.1543 to 243.8855
	age	-0.0064935	1.510905	-0.00	0.997	-2.967813 to 2.954826
6***educ**	LOWEST LEVEL	-0.6215317	509.0905	-0.00	0.999	-998.4206 to 997.1775
	INTERMEDIARY LEVEL	-0.3462282	.	.	.	.
	QUALI.UNIV.APPL.SCI.	-0.5802096	377.1038	-0.00	0.999	-739.6901 to 738.5297
	QUALI.FOR UNIVERSITY	-0.4034092	352.115	-0.00	0.999	-690.536 to 689.7292
	OTHER SCHOOL CERTIF.	1.470377	.	.	.	.
	STILL AT SCHOOL	-0.7253417	335.4413	-0.00	0.998	-658.1783 to 656.7276
3***work**	EMPLOYED, PART-TIME	-0.2694314	98.31261	-0.00	0.998	-192.9586 to 192.4197
	LESS THAN PART-TIME	0.3068741	.	.	.	.
	NOT WORKING	0.2512805	.	.	.	.
_cons		-0.7713792	101.5843	-0.01	0.994	-199.873 to 198.3302

Table 9: AFD Regression Results

	Variable	Coefficient	Std. err.	z	P¿—z—	[95% Conf. Interval]
2***sex**	FEMALE	-0.5406655	148.4905	-0.00	0.997	-291.5766 to 290.4953
	age	-0.0150671	4.138097	-0.00	0.997	-8.125588 to 8.095454
6***educ**	LOWEST LEVEL	1.63573	449.2443	0.00	0.997	-878.867 to 882.1384
	INTERMEDIARY LEVEL	1.775797	487.7128	0.00	0.997	-954.1237 to 957.6753
	QUALI.UNIV.APPL.SCI.	1.32735	364.5504	0.00	0.997	-713.1783 to 715.833
	QUALI.FOR UNIVERSITY	0.5984508	164.3651	0.00	0.997	-321.5513 to 322.7482
	OTHER SCHOOL CERTIF.	2.958238	812.4626	0.00	0.997	-1589.439 to 1595.356
	STILL AT SCHOOL	1.550988	425.9718	0.00	0.997	-833.3384 to 836.4404
3***work**	EMPLOYED, PART-TIME	-0.3353103	92.09146	-0.00	0.997	-180.8313 to 180.1606
	LESS THAN PART-TIME	0.7870437	216.1571	0.00	0.997	-422.8731 to 424.4472
	NOT WORKING	-0.2759964	75.80097	-0.00	0.997	-148.8432 to 148.2912
_cons		-1.634732	448.9704	-0.00	0.997	-881.6005 to 878.3311

In comparison with point 3, the SPD party still is more positively correlated with an older population than the other parties. In fact, the coefficient for age is actually higher in the nested logit model signifying that the Die Grune, CDU-CSU, and SPD parties have on average older populations than the other parties. The predictions below show that CDU-CSU are the highest average predicted probabilities. This could be due to increased data size for these two parties. Regardless, this model is not optimal among 6 alternatives, given the low probabilities; however, could be improved with additional regressors.

Table 10: Summary of Pr(parties alternatives)

Parties	Mean	Std. dev.	Freq.
CDU/CSU	0.6	0.10591335	1,320
SPD	0.42117289	0.14199889	1,313
FDP	0.21212121	0.07990778	1,320
GREENS	0.37014471	0.12987541	1,313
LEFT	0.20868241	0.0533295	1,313
AFD	0.18787879	0.07826761	1,320

Below are the results for the averagge marginal effects for 1 year of age difference. As seen in point 3 SPD has the highest sensitivity to the age variable at .2% change in probability for every 1 year increase in age. Notably, FDP very insensitive to age, suggesting they have a wide variety of aged voters. CDU/CSU has the second highest positive correlation with age, which suggests that the older demographics of Germany tend to vote for the majority parties.

Table 11: Summary of dagedgreen

parties	Mean	Std. dev.	Freq.
CDU/CSU	0.00141581	0.00046034	1,320
SPD	0.00253235	0.00030742	1,313
FDP	0.00052415	0.00026706	1,320
GREENS	-0.00110739	0.00023993	1,313
LEFT	-0.00142505	0.00031148	1,313
AFD	-0.00193985	0.00061305	1,320

2.7 Estimate a few different versions of Multinomial and Nested Logits (around 5-6 versions).

article booktabs threeparttable

Table 12: Condensed Voting Intention: Federal Election

	CDU_CSU	SPD	FDP	THE_GREENS	THE_LEFT	AFD
Racism	Ref.	0.903**	0.961	0.933	0.945	1.315***
		(0.028)	(0.043)	(0.034)	(0.039)	(0.076)
Jews Influence	Ref.	1.045*	1.040	1.040	1.012	1.080*
		(0.022)	(0.031)	(0.027)	(0.029)	(0.035)
Female	Ref.	1.100	0.887	1.406**	1.020	0.649**
		(0.130)	(0.134)	(0.178)	(0.153)	(0.107)
Age	Ref.	1.007	1.000	0.996	0.992	0.983***
		(0.004)	(0.005)	(0.004)	(0.005)	(0.005)
Lowest Level	Ref.	1.905	1.483	0.857	0.578	4.818
		(1.143)	(1.176)	(0.581)	(0.357)	(5.151)
Not Working	Ref.	1.135	0.512***	0.699*	1.227	0.749
		(0.166)	(0.097)	(0.108)	(0.215)	(0.145)

The AFD party is slightly more racist than other parties. Additionally, from previous analysis this party is most likely to have certificate educated peoples, and the least likely to attract women voters.

Table 13: Condensed Analysis of Voting Intention: Federal Election

	CDU_CSU	SPD	FDP	THE_GREENS	THE_LEFT	AFD
Old States	Ref.	1.041	1.106**	1.016	0.973	1.058
		(0.026)	(0.043)	(0.027)	(0.029)	(0.041)
New States	Ref.	0.998	1.014	1.002	1.141**	1.220***
		(0.029)	(0.042)	(0.031)	(0.051)	(0.073)
Female	Ref.	1.070	0.864	1.378*	1.003	0.604**
		(0.126)	(0.130)	(0.173)	(0.150)	(0.099)
Age	Ref.	1.005	0.998	0.995	0.990*	0.983***
		(0.004)	(0.005)	(0.004)	(0.005)	(0.005)
Not Working	Ref.	1.150	0.532***	0.699*	1.243	0.810
		(0.169)	(0.101)	(0.108)	(0.219)	(0.155)

The interesting metrics to me here are the SPD and LEFT coefficients for support of New Federal States vs Old Federal States. This data signals that the Left party may be more likely to contain voters who support the New Federal States and the SPD party is more likely to support the Old Federal States.

Table 14: Condensed Analysis of Voting Intention: Federal Election

	CDU_CSU	SPD	FDP	THE_GREENS	THE_LEFT	AFD
Secret Ballot Meaning	Ref.	0.999	0.935*	1.007	0.957	0.982
		(0.024)	(0.026)	(0.028)	(0.028)	(0.030)
UN Security Council Knowledge	Ref.	1.017	1.032	1.025	1.052**	1.030
		(0.013)	(0.017)	(0.014)	(0.018)	(0.017)
Solidarity Surcharge	Ref.	0.962*	1.043	1.021	1.026	0.964
		(0.018)	(0.030)	(0.023)	(0.028)	(0.024)
Female	Ref.	1.080	0.922	1.458**	1.109	0.629**
		(0.130)	(0.141)	(0.187)	(0.168)	(0.106)
Age	Ref.	1.006	0.998	0.994	0.990*	0.987**
		(0.004)	(0.005)	(0.004)	(0.005)	(0.005)
Not Working	Ref.	1.092	0.512***	0.705*	1.241	0.745
		(0.161)	(0.098)	(0.109)	(0.219)	(0.143)

This particular regression did not tell much about the differences between parties. The Greens are most likely to know about secret ballots.

Table 15: Condensed Summary of Voting Intention: Federal Election

	CDU_CSU	SPD	FDP	THE_GREENS	THE_LEFT	AFD
Acting President of EC	Ref.	0.985	1.041*	1.019	1.019	0.985
		(0.012)	(0.018)	(0.014)	(0.017)	(0.016)
Knows Chancellor of FRG	Ref.	1.039	0.991	1.044	1.044	0.991
		(0.021)	(0.025)	(0.025)	(0.030)	(0.025)
Female	Ref.	1.078	0.925	1.470**	1.080	0.585**
		(0.129)	(0.142)	(0.189)	(0.164)	(0.098)
Age	Ref.	1.006	0.998	0.994	0.990*	0.987**
		(0.004)	(0.005)	(0.004)	(0.005)	(0.005)
Less Than Part-Time	Ref.	2.388**	3.251***	2.415**	3.661***	1.830
		(0.785)	(1.109)	(0.774)	(1.284)	(0.795)
Not Working	Ref.	1.119	0.515***	0.700*	1.239	0.760
		(0.164)	(0.098)	(0.108)	(0.217)	(0.144)

Again we see mostly consistent metrics across the board. However, the measures of German political knowledge here are slightly lower for party AFD. SPD is less highly correlated with knowledge of the acting president of the EC, which may be due to some age related features that are unexplored. Still it is most notable that the AFD party is least likely to know about the current state of government politics.

Table 16: Condensed Summary of Voting Intention: Federal Election

	CDU_CSU	SPD	FDP	THE_GREENS	THE_LEFT	AFD
UN Security Council Seat	Ref.	1.017	1.027*	1.025	1.049**	1.029
		(0.013)	(0.016)	(0.014)	(0.017)	(0.017)
Solidarity Surcharge	Ref.	0.962*	1.026	1.022	1.018	0.960
		(0.017)	(0.028)	(0.023)	(0.027)	(0.023)
Female	Ref.	1.081	0.925	1.457**	1.112	0.629**
		(0.130)	(0.142)	(0.187)	(0.169)	(0.106)
Age	Ref.	1.006	0.999	0.994	0.990*	0.987**
		(0.004)	(0.005)	(0.004)	(0.005)	(0.005)
Less Than Part-Time	Ref.	2.358**	3.168***	2.360**	3.590***	1.862
		(0.775)	(1.078)	(0.755)	(1.258)	(0.808)
Not Working	Ref.	1.093	0.516***	0.706*	1.244	0.745
		(0.161)	(0.099)	(0.109)	(0.220)	(0.143)

In this table, testing the knowledge of global affairs and German economic policy, the AFD and SPD respondents both are least likely to know about economic policy. However, there no significant insights to be gained from knowledge of the UN Security Council Seat.

The most indicative features of the AFD party from the analysis henceforth conducted is as follows: racist, less knowledge of European Commission acting president, less knowledge of current German Chancellor, and less knowledge of German economic policy and less likely to be women.

Table 17: Party Vote Counts

Party	Votes
CDU-CSU	792
SPD	553
THE GREENS	486
FDP	280
THE LEFT	274
AFD	248

Table 18: Summary Statistics for Correct Predictions

Variable	Observations	Mean	Std. Dev.	Min	Max
correct_pred	2,633	0.3137	0.4641	0	1

Table 19: Accuracy by Category

Category	Accuracy
CDU-CSU	0.74873737
SPD	0.10488246
FDP	0.00000000
THE GREENS	0.33744856
THE LEFT	0.00729927
AFD	0.18951613

The overall accuracy of the model is fairly low at 31%. The model predicts very well the accuracy of the CDU-CSU party. This is likely due to the fact that there is a large class imbalance in the dataset. The model does however, do relatively well at predicting the AFD class. This is comparitive to other class prediction accuracies. This is the intended outcome given the party had the least representation in the data and the independent variables identified were directly isolating the AFD party.