

Assignment 4 DCM - Voting Biden or Trump in the 2024 US Presidential Elections

Marc Sparhuber

Table of contents

Introduction	2
Data and Methods	2
Results	4
Conclusions	7

Introduction

The following analysis aims to answer the question whether, from an electoral/political standpoint (not a leadership or competence standpoint), concerns about Biden's age are warranted. It is expected that among other variables, age will play a crucial role in predicting the probability of choosing a candidate for the next election in the US in the fall of 2024. The main dependent variable is the choice within a choice set of two fictive candidates with discretely varying attributes. The main independent variable is age measured in years, while other variables included in the analyses are Education and Profession, as these vary between the two likely candidates in the 2024 US election, with Trump having received a degree from an Ivy League university and Biden from a state university, as well as, Biden having been a Lawyer and Trump as Business Owner before their political careers. Other variables such as Religion and income vary between them and would have been interesting to include but were decided against due to Trump's beliefs not clearly fitting into any of the levels of the Religion variable and Income, as I argue it is not feasible to expect that decisions in a voting process would be made based on knowing how much a candidate has earned or, more importantly, being able to find out precisely how much wealth they possess. In the case of Trump's confession, subjectivity of respondents disallows taking it into account as I pose that despite his official, yet vague outing as a "confessionless Christian" will be either disregarded or misconstrued in a confirmatory manner to fit the respondents beliefs about Trump's beliefs. Regardless, it is expected that being a Lawyer may represent greater lawfulness, therefore predicting a higher share of choices and Ivy League education being preferred over other tertiary education, as it may be perceived as a measure of intelligence. Finally, age is expected to predict a greater likelihood of being chosen but decline at some point around 65 (arbitrarily chosen), which would result in Biden being perceived from an electoral standpoint as the less favored candidate, as he will be nearing 82 around the time of the 2024 US presidential election and Trump sporting a sprightly 78 years.

Data and Methods

The data to predict the results of the 2024 presidential elections, however, are from 2012 and were initially collected for a study by Hainmueller et al., 2014 with the goal of demonstrating a new technique for causal inference from survey experiment data. The data are non-representative and gathered from MTurk, yet it can be assumed that the criteria for what constitutes a desirable presidential candidate might not have changed much. The variables used in this analysis remain mostly unchanged factor variables (Profession and Education), while Age (factor variable with 6 discrete values - units in years) is supplemented by another variable squaring

it, aptly named Age Squared. Interestingly, the median age of the chosen alternatives is lower than those of the unchosen ones, showing that lower age seems to be more desirable in a presidential candidate, though the interquartile range is the same for both chosen and unchosen, indicating that too young a candidate, i.e., 36 is perhaps considered not experienced enough. While gender (and race, though not shown) are divided quite equally among chosen and unchosen alternatives, social desirability during response to the survey experiment may have been a biasing factor here. For Education, as can be seen in Table 1, generally speaking the higher the education, the likelier the candidate was picked. The picture is a little less clear with regards to profession but generally speaking it seems that prestige seems to win out, leaving car dealers and farmers lagging behind the other professions in percent chosen.

Three models are constructed. The first contains age and is meant to test the main hypothesis that older candidates may be deemed less qualified by the voting age public. The second model adds a squared term for age, looking out for non-linear relationships between age and candidate choice. The third model, as can be seen in Table 2, adds the variables which vary among Biden and Trump and they can be clearly identified on (in retrospect - I'm running out of time - I should've added all variables to the model regardless, as although Trump and Biden might not differ in certain variables, adding more variables may have improved the model fit). Model 3 includes the other two previously mentioned factor variables Education and Profession with their values "No BA" and "Car dealer" serving as reference categories, as can be seen in Table 2.

Table 1: Voting for Trump or Biden in the 2024 US Election: Descriptives Divided by Choice

Characteristic	Unchosen, N = 1,733	Chosen, N = 1,733
Age (in years)	60 (45, 68)	52 (45, 68)
Gender		
Female	854 (49%)	857 (49%)
Male	879 (51%)	876 (51%)
Education		
Baptist college	274 (16%)	255 (15%)
Community college	273 (16%)	264 (15%)
Ivy League university	243 (14%)	397 (23%)
No BA	379 (22%)	195 (11%)
Small college	282 (16%)	306 (18%)
State university	282 (16%)	316 (18%)
Profession		
Business owner	250 (14%)	330 (19%)
Car dealer	385 (22%)	190 (11%)
Doctor	276 (16%)	327 (19%)
Farmer	289 (17%)	266 (15%)
High school teacher	278 (16%)	313 (18%)
Lawyer	255 (15%)	307 (18%)

¹ Median (IQR); n (%)

Note: Data from Hainmueller et al., 2014 (collected July 2012)

Results

Table 2 presents the results of the three models. In Model 1, the significant Age coefficient of 0.98 indicates that respondents are 2% less likely to choose a candidate that is one ordinal step of the age variable higher. That this effect is likely non-linear is not only shown by inspecting the descriptive data of the chosen alternatives where 52 was the most chosen age category with the lower bound 36 being second most chosen and 75, the upper bound being the least chosen but also by the inclusion of the squared age term in Model 2. The log odds of the age squared term being at 1 and the one of age at 1.10 indicates that though the probability of choosing someone with a greater age initially increases, it then decreases later on. Model 3 adds Education and Profession, while keeping the squared age term. Relevant to our hypotheses it can be seen that Ivy League and State university education are most desirable, being 3.13 and 2.25 times as likely to be chosen with the same profession and age, compared to someone without a bachelor's degree, respectively. In regards to profession, compared to car dealers, business owners

and lawyers were 2.6 and 2.3 times as likely to be chosen, all else equal. All terms were significant.

Table 2: Voting for Trump or Biden in the 2024 US Election: Conditional Logistic Regression Models

Covariate	Model 1			Model 2			Model 3		
	Odds	95% CI	p-value	Odds	95% CI	p-value	Odds	95% CI	p-value
Age (in years)	0.98	0.98, 0.99	<0.001	1.10	1.05, 1.16	<0.001	1.10	1.05, 1.16	<0.001
Age Squared				1.00	1.00, 1.00	<0.001	1.00	1.00, 1.00	<0.001
Education									
No BA							—	—	
Baptist college							1.83	1.41, 2.38	<0.001
Community college							1.77	1.38, 2.27	<0.001
Ivy League university							3.13	2.43, 4.03	<0.001
Small college							2.06	1.60, 2.66	<0.001
State university							2.25	1.75, 2.89	<0.001
Profession									
Car dealer							—	—	
Business owner							2.60	2.01, 3.35	<0.001
Doctor							2.42	1.89, 3.11	<0.001
Farmer							1.80	1.40, 2.31	<0.001
High school teacher							2.17	1.69, 2.80	<0.001
Lawyer							2.27	1.77, 2.91	<0.001

¹ CI = Confidence Interval

Note: Data from Hainmueller et al., 2014 (collected July 2012). The units for the age term is years. The reference category for Education is No Bachelor's Degree and Car Dealer for Profession.

The inclusion of the squared age term in the third model is justified not only by the given rationale but also the log likelihood, AIC and BIC being lower than that of Model 1, as can be seen in Table 3. Likewise, and unsurprisingly Model 3 sports the best model fit of these models. Inspecting the likelihood ratio test it can similarly be seen that there is a greater jump of model fit from Model 2 to 3 than 1 to 2. Crucially, this does not mean that Model 3 is well specified, as there are plenty of interesting variables that should have been included here and the share of correct predictions are not calculated.

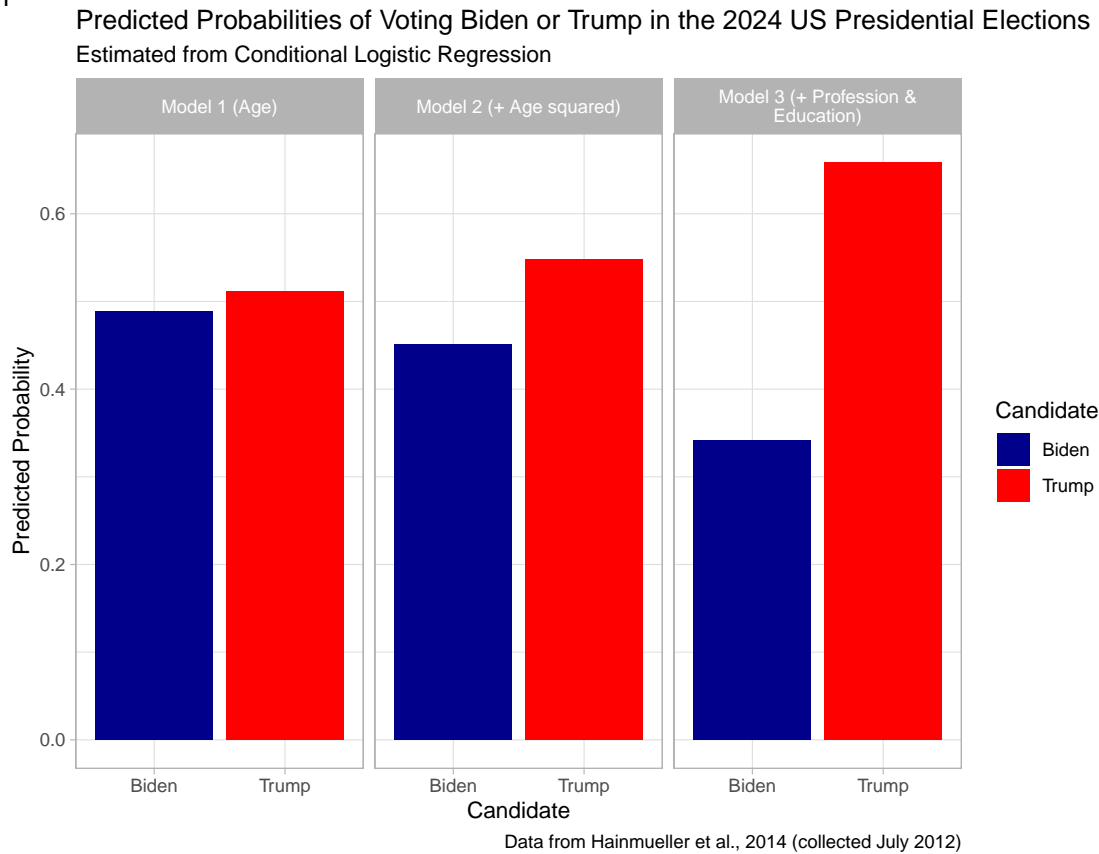
Table 3: Voting Trump or Biden in the 2024 US Presidential Elections: Model Fit Statistics

	Model 1	Model 2	Model 3
Log Likelihood	-1183.11	-1172.74	-1084.6
AIC	2368.23	2349.49	2193.19
BIC	2373.69	2360.41	2258.69
Likelihood Ratio		20.74	176.3
Likelihood Ratio (Df)		1	10
Likelihood Ratio (p-value)		5.27e-06	1.38e-32

Note: The Likelihood ratio is always calculated with the nested model to the left. Data from Hainmueller et al., 2014 (collected July 2012)

Figure 1 shows the out of sample predicted probabilities for voting Trump or Biden for each model. Specifying the corresponding values for the two candidates in each model - ages in model 1, adding age squared in model 2 and education and profession in model 3 - and then predicting the probabilities leads to the following results. While Model 1 only captures a linear effect of age this puts Trump slightly ahead of Biden in predicted probability. Adding a non-linear age effect, however, increases his lead in predicted probability by quite a bit. This corresponds to the results from Table 2 and a look at the distribution of the ages of chosen candidates decreasing noticeably with age, making even a relatively small age difference between Biden and Trump of 4 years result in an approximately 10 percent point lead for the latter. Model 3, which adds profession and education increases this lead, as would be expected by the results in Table 2, in which the combination encapsulating characteristics of Trump had higher log odds associated with them than Bidens, compared to the reference categories. This model predicts a 65% predicted probability of voting Trump.

Fig. 1



Conclusions

It was assumed that there would be a drop off around 65 when increasing age would not benefit the candidate anymore. This was not the case and indeed the drop off came after age 52 instead. This is also due to the categorical nature of the age variable which was, however, largely treated as continuous in order to make out of sample predictions. This definitely leads to some imperfections in the predictions but is a constraint of conditional logistical regression models based on survey experiments with categorical choices instead of random continuous age variables (which could and should've easily been implemented were the data made for the context it is used in here). Further, the variables taken into account in Model 3 are far from comprehensive and the choice of included predictors can justly be criticized (as I did above already). All in all, the way the Models were specified here, it is predicted that Trump will be voted into office again in 2024 so I should've really changed my models lol.