The Impact of Postgraduate Education Attainment on Legislators' Progressive Voting Patterns

Andrew Chang | SID: 21760758
Undergraduate Honors Thesis
Fall 2013 Term
Advisor: Prof. Enrico Moretti

Department of Economics
University of California, Berkeley

Abstract

This paper seeks to answer the question of whether legislators with post-graduate levels of education tend to vote more progressively. Our dataset is constructed by taking members of the U.S. House of Representatives between 1999 and 2005 (from the 106th to the 109th Congress) as data points, examining their voting scores established by the ADA (Americans for Democratic Action) as proxies for their propensity to vote liberally. The ADA compiles these scores annually, taking a set of 20 policies and assigning each legislator a score equal to the percentage of their votes, within those 20 policies, that correspond to a politically "progressive" vote. Having controlled for age, gender, and district-related factors, I conclude that my results from an OLS regression corroborate a positive correlative relationship between post-graduate degrees and higher liberal voting.

Acknowledgements: I would like to thank Professor Enrico Moretti for guiding me through each step of this thesis, and being helpful and encouraging even when significant obstacles arose. This thesis would never have been possible otherwise, and I am extremely appreciative for his patience and time in this endeavor.

I. Introduction

In 2008, Ebonya Washington wrote a paper titled *Female Socialization: How Daughters Affect Their Legislator Fathers' Voting on Women's Issues*, exploring the influence of parenthood on legislators' feminist sympathies. Washington found that, ceteris paribus, politicians who were fathers to female children voted more liberally on legislation concerning women's issues (Washington 2008). Importantly, her paper concludes that personal ideology, irrespective of voter preferences, can hold a significant amount of weight in determining how a legislator chooses to vote.

However, personal ideology is only explored in one dimension in Washington's NBER paper, and invites the question of which other factors might contribute to a given politician's political sympathies. In this paper, I analyze another potential dimension of political ideology: the relationship between a politician's level of education and his/her political beliefs (ostensibly manifest in their voting patterns). In an era of political polarization, votes on certain key pieces of legislation can be identified as fundamentally conservative or liberal based on the rhetoric surrounding the issue. Beyond party affiliation, which has a strong effect on the way legislators vote, the other explanatory factors for how representatives choose to vote are often opaque and minimally understood. This paper attempts to identify a positive causal effect of postgraduate-level education on liberal voting.

There is a general myth that exists which posits that more "intelligent" people tend to express liberal or progressive beliefs comparatively more than those considered to be "less intelligent." Satoshi Kanazawa, in the March 2010 publication of *Social Psychology Quarterly*, lends

credence to this thesis by demonstrating that more intelligent individuals (categorized by IQ) increasingly express the belief that the government should actively engage as a purveyor of economic and social justice (Kanazawa 2010). Kanazawa's work bears relevance to this thesis in a very major respect, but only goes as far as to say that the *population* of a country has a higher propensity to represent a liberal ideology when it is measured to be considered more intelligent. *Legislators*, however, are qualitatively dissimilar to people in the general population for a myriad of reasons – to begin with, they are constrained by their party ideologies, their voter base, and concerns for re-election. In this paper, I adapt Kanazawa's supposition to a framework analyzing the behavior of elected representatives, controlling for these considerations which may distort their behavior.

It is critical to note that in this paper, the question of "intelligence" is replaced with the question of whether legislators have attained a postgraduate degree. This is intentionally established for two specific reasons. The first reason concerns the fact that information on legislators' IQ is neither readily nor publicly available. There exists an extremely high likelihood that a subset of congresspeople either a) did not take the same IQ test as the other congresspeople, b) took the same IQ test as the other congresspeople, but at a different time in their life, c) did not take the IQ test at all, ever, or d) took the IQ test, but chose not to disclose their score. The second reason is comparatively more important, and central to this thesis: that a postgraduate degree entails much more than raw intelligence measured by a test. Rather, it combines many different components such as exposure to topical political literature, further engagement with differently-minded peers, and continued development of critical thinking skills. These are topics of analysis

enabled only by looking at a Boolean "postgraduate degree obtained" regressor rather an "intelligence" regressor.

Our general model in this paper takes members of the House of Representatives between 1999 and 2005 as data points. In addition to other variables such as age, gender, and party affiliation, our major concern in this thesis is the dummy independent variable for postgraduate education, and the dependent variable measured as a given legislator's ADA score. The ADA score is a rating assigned by the Americans for Democratic Action to individual policymakers annually. It is calculated by taking a yearly set of 20 representative policies, and looking at the policymaker's vote on each of the 20 policies. For each vote in line with the progressive vote, the legislator is assigned 5 points; for each conservative vote or abstention, the legislator is assigned 0 points. In essence, the ADA score is a proxy for the legislator's propensity to vote progressively in a given year.

The goal of this paper is to capture of the effect of postgraduate education – independent of voter preferences – on ADA score increases. Insofar as this thesis intends to go beyond merely describing the data, my main challenge was proving causality between having a postgraduate degree and voting progressively. The ideal method of proving this causality would be to instigate a controlled experiment whereby I could administer the experience of having a postgraduate-level education to a treatment group of policymakers, and compare their ADA scores in a subsequent year to a control group of policymakers without postgraduate degrees. Realistically, it is impossible to disentangle policymakers with postgraduate educations from such factors as

the education levels of their electorates. These factors are almost certainly associated with legislators' ADA scores, problematizing our results.

In response to this issue, we explicitly address the question of district preferences through a district fixed-effects model. It might be true that districts with better-educated voters might ultimately choose to elect better-educated representatives, and an argument could be made that these districts would then compel their representatives to vote more liberally. However, our fixed-effects model absorbs districting as a categorical factor in way that allows our regression to operate as if each district was specified by a dummy variable. This fixes district-specific effects to the coefficients of our dummy variables and significantly mitigates the problem of entanglement. My findings are consistent with the hypothesis that there exists a positive causal relationship between having a postgraduate degree and voting more liberally.

In writing this thesis, I imagine that there are additional applicable conclusions that can be drawn by looking only at certain subsets of my sample. For example, it is widely understood that contemporary U.S. politics is characterized by contestation and extreme bipolarity in ideologies (Dixit and Weibull 2007). By looking at Republican members of the House of Representatives and whether they are more likely to vote for progressive policies given higher levels of education, we can also examine whether certain policymakers are more likely to defect from the stronghold of their party ideology. Although this paper is somewhat silent on whether this represents pragmatism or simply an ideological prioritization, a correlative relationship entails far-reaching implications in reducing political friction and inefficacy.

This paper also remains ultimately agnostic as to how a postgraduate degree might compel a policymaker to vote more progressively. Perhaps, as Maranto and Woessner suggest, higher education is dominated by professors whose politics are well to the left of the American political center (Maranto and Woessner 2012). Under this interpretation, we would imagine that policymakers are compelled by their professors to adopt marginally more left-leaning ideologies before they adopt their roles as elected representatives. Another more partisan explanation might be that progressive policies are more empirically justified, and higher rates of legal education bring policymakers to vote in line with historical and statistical "truths." These are but a few of many causal explanations that might illuminate the phenomenon explored in the paper.

IIa. Data

To construct my data set, I sourced the names of legislators in the House of Representatives from the Washington Post U.S. Congress Votes Database. For the purposes of this paper, I take the 106^{th}, 107^{th}, 108^{th}, and 109^{th} Congresses for examination, which yielded 1765 data points. Importantly, these data points are not necessarily representative of 1765 different people. For example, a representative in one Congress and the same representative in the next consecutive Congress still represent two discrete data points. There are in total 587 unique members of the House of Representatives that make up the 1765 observations.

The data for ADA scores was taken from the website of the Americans for Democratic Action: a political organization which tracks progressive issues in legislative policy and assigns an "ADA" rating to each legislator in a given session of Congress according to how they vote on a set of 20 identified "key" bills. According to the ADA, each member receives 5 points for a vote with the

ADA, and 0 points for a vote opposing the ADA. Absent or abstaining members of Congress also receive 0 points.

I found the education levels of each representative on my own. There were available biographies for a vast majority of the legislators on the Washington Post U.S. Congress Votes Database, but a small minority of the entries did not disclose legislators' degrees. In these cases, I consulted internet resources to code each legislator's educational attainment as analyzable data. The same can be said of each legislator's age and gender. I also determined each legislator's regional affiliation by assigning each district to one of the nine categories explicated by the U.S. Census Bureau by hand. In summary, I collected the data on education, gender, region, and age myself through a synthesis of available online biographies. Additional information on data sourcing methodology can be found in the variable descriptions in section IIIa.

In this paper, I take the ADA scores from years 1999, 2001, 2003, and 2005, which are the first years of the 106^{th}, 107^{th}, 108^{th}, and 109^{th} Congresses, to define my individual data points. This paper only examines the first years of each Congress because the personnel composition of each Congress remains relatively static between years 1 and 2. In applying the yearly ADA scores to the observations, there were 35 instances in which the ADA score was missing or inapplicable. Within these situations, there were several instances where the Washington Post Database listed the policymaker as belonging to a political party, and the ADA listed the policymaker as belonging to the opposite party. I omitted these 35 data points from the analysis to preserve the purity of the data, which left a sample size of 1730 for the proceeding econometric analysis.

There were two discretionary decisions involved in selecting the proxy for educational attainment. First, I made the choice to measure education by looking at postgraduate degrees (MA, JD, PhD, MD, DDS, DPhil, LLB) rather than undergraduate study. The justification for this choice is relatively simple – out of 843 Democrat data points, 836 (or 99.17%) had undergraduate degrees, and out of 922 Republican data points, 912 (or 98.92%) had undergraduate degrees. Given this fact, I am left with the choice of either making the assumption that graduate degrees actually represent a significant qualitative difference from having only an undergraduate degree (which is intuitively defensible and likely true), or taking an extremely small and unrepresentative sample as a control group for analysis. Second, I elected to represent a postgraduate education as a binary variable rather than a sliding scale. This assumption narrows the scope of the paper for the sake of simplicity. Perhaps a follow-up question to the correlation explored in this paper would determine the marginal influence of each additional degree beyond undergraduate study.

In summary, 64.0% of legislators had one or more postgraduate degrees; this figure is an amalgam of 70.7% of Democratic legislators and 57.9% of Republican legislators. Within the realm of personal characteristics, 80.4% of Democrats were male, and 91.0% of Republicans were male – with the average age of Democrats at 66.268 and the average age of Republicans at 64.925. There is presumably no propensity for the age difference to foster any significant differential in education opportunity between the two groups. The two regions with the highest number of House Democrats were the Pacific and Mid-Atlantic regions, with 20.9% and 16.7% of the Democrats studied coming from these two respective regions. In contrast, the two regions with the highest number of House Republicans were the South Atlantic and East North Central

regions, with 20.8% and 17.1% of Republicans studied coming from these two respective regions.

IIb. Data Weaknesses

One conceivable weakness of this data set was the high rate of incumbency throughout the 4 sessions of Congress covered by the timespan between 1999 and 2005. As I allude to previously, while there were 1765 unique observations, there were really only 587 unique legislators. This meant that at the district level, we frequently find one policymaker occupying all four years in this study. An ideal data set would be conversely characterized – if each district elected a different policymaker in each of the 106^{th}, 107^{th}, 108^{th}, and 109^{th} Congresses, then we would have a better understanding of the effect of postgraduate education attainment irrespective of selection at the district level. That is – we would be more certain that individual districts do not just have a preference for liberal policies and more educated representatives.

In addition, ADA scores are imperfect indicators of a politician's propensity to vote liberally. There are multiple dimensions to this problem. Firstly, ADA scores amalgamate many different types of progressivity into a single metric, which fails to speak to the differences between economic progressiveness, social progressiveness, and political progressiveness. To my knowledge at the time of writing, the ADA does not establish a predetermined percentage-allocation for the distribution of economic, social, and political bills included in the 20 used to calculate ADA scores. This effectively means that variation from year to year may exist solely as a function of the fact that one year the ADA rating might be derived from primarily social legislation, whereas the next year it might be derived from economic issues.

Aside from the distribution of social, economic, and political pieces of legislation, there is also the question of the degree to which a vote is considered to be "progressive." The ADA treats all progressive votes as equal at 5 points. However, even within social issues there are conceivably pieces of legislation that are more politically charged than others – and therefore represent a greater amount of "progressivity" with a yes vote. For example, one might imagine that the 2001 resolution to approve of a workplace safety rule is less relevantly "progressive" than the 1999 resolution to provide abortion assistance to minors. However, both result in the same point value.

On the other hand, it is ultimately necessary to quantify progressivity to have a workable regression model. It is critical to understand that the process of discretizing qualitative variables inherently entails some form of flattening – and in that vein, ADA scores represent our best possible choice for evaluating the effect of postgraduate education attainment on liberal voting. However, if I were to imagine the ideal data set, I would have ADA scores divided into categories based on the content of the legislation from which the ADA score is derived. Individual regressions for each category would be more specific and insightful than a single regression for a legislative cornucopia.

IIIa. Econometrics

Using ADA scores as the dependent variable, I estimate a regression of the following form:

$$Y = \alpha + \beta_1 GRADUATEEDUCATION + \beta_2 MALE + \beta_3 DEMOCRAT + \beta_4 AGE +$$
$$\beta_5 NEWENGLAND + \beta_6 MIDATLANTIC + \beta_7 EASTNORTHCENTRAL +$$

β_8WESTNORTHCENTRL + β_9SOUTHATLANTIC + β_{10}EASTSOUTHCENTRAL + β_{11}WESTSOUTHCENTRAL + β_{12}MOUNTAIN + ε

GRADUATEEDUCATION refers to a binary variable describing whether a given legislator has attained a level of education beyond undergraduate study, taking on a value of 1 if the individual has done so and a value of 0 otherwise. The information on whether legislators in the 106th to 109th Congresses have attained postgraduate degrees is taken from the Washington Post U.S. Congress Votes Database, which has compiled biographies and voting patterns of both the House and the Senate from 1991 onwards. The associated β_1 coefficient indicates that all else equal, attaining a postgraduate education (GRADUATEEDUCATION = 1) correlates with a legislator's ADA score being higher than that of a legislator without a postgraduate education (GRADUATEEDUCATION = 0) by β_1 points.

MALE refers to a binary variable describing whether a given legislator identified as male during his or her time in office, taking on a value of 1 if the individual is male, and 0 if the individual is female. The name of each representative is publicly available in the *Congressional Directory*, and the Washington Post U.S. Congress Votes Database provides portrait pictures of a majority of House representatives. If the gender was not immediately ascertainable by name (as in cases of ambiguity) and a portrait was not available, I consulted Internet resources. The associated β_2 coefficient indicates that all else equal, identifying as male (MALE = 1) correlates with an ADA score increase of β_2 compared to those identifying as female (MALE = 0).

DEMOCRAT refers to a binary variable describing whether a given legislator was a registered Democrat during his or her time in office, taking on a value of 1 if the individual was a registered Democrat and a value of 0 if the individual was a registered Republican. The associated β_3 coefficient indicates that all else equal, a Democrat party affiliation (DEMOCRAT = 1) correlates with an ADA score increase of β_3 compared to those with a Republican party affiliation (DEMOCRAT = 0).

AGE refers to the numerical age of a given legislator, in years. This number is calculated as of December 3, 2013 and is taken by subtracting the birth year of the legislator from the year 2013 as a rough calculation of his/her age. A simplifying assumption is made here that the legislator is not born after December 3 in a calendar year – which would make him or her a year younger than calculated – but in the interest of time it was determined that this assumption would not affect the results, as there is no realistic reason to assume that legislators born after December would exhibit any cohesive set of preferences that would move the data in a particular direction. The associated β_4 coefficient indicates that all else equal, and increase in age by 1 year correlates to an ADA score increase of β_4.

The remaining independent variables – NEWENGLAND, MIDATLANTIC, EASTNORTHCENTRAL, WESTNORTHCENTRAL, SOUTHATLANTIC, EASTSOUTHCENTRAL, WESTSOUTHCENTRAL, and MOUNTAIN – refer to the geographic regions of the U.S. from which the legislator's state may reside. These regional classifications, taken from the U.S. Census Bureau, are all binary variables that seek to capture a regional tendency for legislators to vote a certain way. The results of the overall regression

indicate that this is important – in 8 of the 8 regions a coefficient was found to be statistically significant at the $p < 0.05$ level. Importantly, a ninth PACIFIC region was omitted as a reference group. The coefficients ($\beta_5 - \beta_{12}$) are each associated with a specific region. If a legislator represents a region, the associated β coefficient to that region correlates to his/her ADA score increase/decrease stemming from regional effects.

IIIb. Results

Our results from this regression are summarized in the rightmost column in Table 2. From the leftmost column to the rightmost column, each regression includes an additional variable which clarifies and corrects biases inherent in its previous omission. This continues up until we reach the regression in the form isolated in part IIIa.

The most basic regression in the first column only considers postgraduate education attainment as an independent variable, resulting in a beta coefficient of 13.05 and a standard error of 2.02. This beta coefficient would suggest that ceteris paribus, a postgraduate education corresponds to a 13.05 point increase in ADA scores, which is absurdly large and almost certainly a product of an upward bias due to omitted variables. The R^2 value for this regression line is 0.02, confirming that this minimal regression is ill-equipped to explain why legislators might vote in a way as to garner higher ADA scores. Certainly, there are other factors that might lead someone to vote liberally. We move from left to right in Table 2, then, in order to add the omitted variables that are showing up in the beta value. For instance, since being male correlates negatively with ADA scores, the exclusion of MALE leads to a smaller estimation of GRADUATEEDUCATION's

beta value because the effect of being male shows up in GRADUATEEDUCATION's beta value. As we continue rightward, we correct for the biases of each omitted variable.

The final results (Table 3) are a more detailed profile of the rightmost regression in Table 2. These results are comparatively more reasonable. This regression indicates that having a postgraduate level of education correlates to a 3.67 point increase in ADA scores, and the MALE beta value only corresponds with a 6.04 point decrease in ADA scores. In addition, the AGE value is no longer statistically significant at the $p < 0.05$ level. The New England and Mid-Atlantic regions correlated positive with ADA scores, whereas the rest of the regions correlated negatively. Importantly, our R^2 value changes to 0.92, which signifies a significantly more fitting regression line for this data.

Even so, there is still a problem with this regression, which is that the standard errors are biased downwards. Our data set treats all observations as independent, but there is a strong argument in favor of the idea that they are not. For example, three observations consisting of a single legislator in three different years are absolutely not independent from each other. Even different Congressmen from the same district share the same electorate (with some minor modification over time), and insofar as the electorate ultimately selects the legislator, this correlation is problematic. This is not captured by our regression model.

My goal, in adjusting for this, is to account for the fact that we have an effective number of observations greater than the number of districts but smaller than our total dataset in its entirety. Given this fact, I chose to do a clustered regression in STATA, as opposed to a standard OLS in

GRETL, to find unbiased standard errors in our sample. I chose to cluster by district, with each district representing a cluster of four observations, which are the four legislators which represented the given district over the 106th to 109th Congresses. My choice to cluster by district precluded the opportunity to cluster by individual names. However, I found this to be the correct decision because clustering by individual name would not capture the aforementioned "same-electorate" effect that would bias standard errors downward. Additionally, clustering by district captures the same effect as clustering by individual name, because very infrequently did legislators switch districts.

The results (Table 4) can be compared against the previous results (Table 3). Our beta coefficients remain the same, as they should – but the corrected standard errors in Table 4 are significantly higher across the board. For GRADUATEEDUCATION, the standard error is 0.61 for the non-clustered regression, and 0.85 for the clustered regression – an increase of almost 38%. We see a similar upward progression in the standard errors for our other independent variables – MALE from 0.85 to 1.09, DEMOCRAT from 0.61 to 0.98, and all regions as well. The standard error is slightly more pessimistic under the clustered regression.

There is still an issue; this does not control for the fact that each district individually effects the propensity for its legislator to vote liberally. Each representative in a given Congress is necessarily affected by the demographic he or she represents – thus, the cluster regression is still inadequate in the sense that it does not incorporate the impact of these individual effects on our calculation of GRADUATEEDUCATION's beta coefficient. Therefore, we turn to a district fixed-effects model using the "areg" function in STATA to functionally regress ADA scores

against postgraduate education level, gender, age, party affiliation, and a dummy variable for each district. By absorbing the effects of districting in our "areg" model, we are able to extract a value of beta for GRADUATEEDUCATION that is theoretically independent of the ADA score fluctuations caused by district-specific effects.

The results (Table 5) indicate that in our model incorporating fixed effects for districts, a postgraduate education correlates to a 3.85 point increase in ADA scores, compared to a 3.67 point increase in our previous model. Surprisingly, our standard error is significantly higher, from 0.61 in our non-clustered regression and 0.85 in our clustered regression to 1.06 in this regression. Since district fixed effects subsume effects by region – insofar as regional assignments are determined as a function of districts – the eight dummy variables for regional effects were not included. A larger change occurred in the beta coefficient for DEMOCRAT, decreasing from 74.68 to 69.62. This indicates that the omission of district fixed effects biased our results upward for that independent variable.

We might also be interested in the interaction between postgraduate education attainment and other independent variables. For example, in our current regression, we understand ADA scores to be a function of both postgraduate education and gender (among other factors), and find the effect attributable to having a postgraduate education and the effect attributable to being a certain gender separately. However, there is no way to identify any unique effect that may arise from the combination of being both male and having a postgraduate degree in our current model. Thus, we can estimate a regression of the following form (Table 6):

$$Y = \alpha + \beta_1 GRADUATEEDUCATION + \beta_2 MALE + \beta_3 DEMOCRAT + \beta_4 AGE + \beta_5 NEWENGLAND + \beta_6 MIDATLANTIC + \beta_7 EASTNORTHCENTRAL + \beta_8 WESTNORTHCENTRL + \beta_9 SOUTHATLANTIC + \beta_{10} EASTSOUTHCENTRAL + \beta_{11} WESTSOUTHCENTRAL + \beta_{12} MOUNTAIN + \beta_{13} MALEINTERACTION + \varepsilon$$

where MALEINTERACTION represents the interaction between being male and having a postgraduate education. In other words, MALEINTERACTION is a dummy that takes on a value of 1 if and only if MALE = 1 and GRADUATEEDUCATION = 1. The beta coefficient represents the unique numerical change in ADA score when, if all else is equal, a legislator is both male and has a postgraduate degree. As we can see in Table 10, this value is equal to .71, but is not statistically significant with a p value exceeding 0.7. This makes intuitive sense; I would not hypothesize that there is any reason why being male would change the nature of how a postgraduate education informs legislative decision-making. Interactions for age similarly fail to be statistically significant.

Of interest, however, is the question: does the correlation between postgraduate-level education and higher ADA scores relate more strongly to Republicans than Democrats, or vice versa? To test whether the results are different for Democrats and Republicans, I also ran two additional regressions: one where data points are limited to only House Democrats (n =826) and one where data points are limited to only House Republicans (n=904). The Democrat regression had a beta coefficient for GRADUATEEDUCATION at 4.27, whereas the GRADUATEEDUCATION beta coefficient for House Republicans was only 2.43 (see tables 7 and 8). This works similarly to an interaction – by dividing the sample into the two party affiliations, we can examine the

difference in GRADUATEEDUCATION betas between the two to conjecture about the unique effect of having a postgraduate education and being a Democrat, versus having a postgraduate education and being a Republican.

This is an interesting finding, because the ADA scores are capped at 100. This means that if a postgraduate education were to spur a Democrat representative to be progressive at a level beyond an ADA score of 100, our model would not capture that effect. Put differently, there are a significant number of Democrats (compared to 0 Republicans) whose ADA scores are 100, but some may have voted more progressively than others in other pieces of legislation not included in the 20 used to calculate ADA scores. Republicans, on the other hand, are not bounded by the data set, as even Republicans with the highest non-outlier ADA scores did not exceed around 55 points. One might imagine that the explanation to this phenomenon exists in the different ways education is perceived in the value systems of Democrats and Republicans. Perhaps the Democratic party is more likely to celebrate academia, whereas the Republican party would be more likely to value industry experience, leading Democrats to express the effects of their educational attainment more so than Republicans. Even so, the effect of a postgraduate education continues to be larger for Democrats than Republicans by about 76%. Given our constraints, I expect that this 76% figure is conservative, and has the propensity to be larger.

IIIc. Limitations

There is a possibility that our model runs into issues with omitted variable bias – a case where an omitted independent variable is correlated with both the dependent variable and one or more of our independent variables. Let us imagine a formula for omitted variable bias as the following:

$\beta_m - \beta_a = \rho_{xz}\,(\sigma_z/\sigma_x)$

where β_e is equal to our estimated beta, β_a is equal to the theoretical "true" beta that we seek to estimate, x is our independent variable, z is an omitted variable, ρ_{xz} is equal to corr(x,z), and σ_z and σ_x represent the variances of z and x, respectively. The distance between our estimated beta, β_e, and the "true" beta, β_a is the bias caused by the omitted variable, $\rho_{xz}\,(\sigma_z/\sigma_x)$. In the formula, we imagine this bias to be a function of both the sign and the strength of the correlation between x and z.

An example of omitted variable bias might exist in the fact that we never consider the effect of net worth. We can imagine that a policymaker's net worth is positively correlated with higher education – insofar as obtaining a graduate degree requires a significant up-front cost, legislators with higher incomes may have had a history of affluence enabling them to pursue a postgraduate degree whereas those with a lower net worth may never have had the chance to do so. Even so, I would not estimate the correlation to be very strong, since a legislator's present net worth can be significantly different from the legislator's net worth at the time of choosing whether or not to pursue a postgraduate degree. Applying our formula to this instance, we would have a beta value for GRADUATEEDUCATION that is biased upwards to a small extent – since what we wanted to measure (β_a) is augmented by the positive bias ($\rho_{xz}\,(\sigma_z/\sigma_x)$) in our measured beta ($\beta_m$).

We might also consider two examples of potential omitted variables which would bias our beta coefficient for GRADUATEEDUCATION downwards rather than upwards: minority status and

military service. If we make the assumption that there are racial patterns in postgraduate education (which is augmented by the fact that a significant portion of sampled policymakers went to school when the civil rights era had yet to gain momentum), our beta value might be underestimated. Even if we assume that the correlation between being a minority and having a postgraduate degree is weakly negative, we still run into omitted variable bias because our β_m is equal to the sum of β_a and the negative expression $\rho_{xz}(\sigma_z/\sigma_x)$. I would assert that military service could also be negatively correlated with postgraduate education attainment, since they represent divergent choices for people in their twenties and thirties – this would also lead us to underestimate the beta coefficient for GRADUATEEDUCATION.

A key difficulty in this regression might be the fact that different geographic locations certainly have different characteristics relative to levels of education. It may certainly be possible that a given district full of highly-educated academics might seek to elect a legislator who has obtained a postgraduate degree, who would then be pushed by his or her constituents to vote liberally on the policies from which his or her ADA score is derived. However, the correlation between highly educated voters and highly educated legislators might not be as strong as imagined; the Republican with the highest ADA score, for example, represents the Inland Empire area of California: an area where 35% of adults did not graduate from high school. Political science literature suggests that voters make their decisions based on three categories: the value image (ethics and morals), the trajectory image (campaign platform), and the strategic image (how the policymaker achieves goals) (Levine 2005). I do not see education levels are particularly salient to any three of these categories – and it is very reasonable to believe that in many campaigns, candidates do not publicly disclose their education levels to garner votes. Education levels may

factor into some compartment of a candidate's overall image, but the effect seems to be demonstrably minor.

An additional issue might exist in the fact that this paper only looks at the first year of each Congress from 1999-2005. This means that the 106th Congress, which has two sets of ADA scores for 1999 and 2000, is represented only by the behavior of its legislators in 1999 in our data set. The same goes for the 107th Congress, which is represented by legislators in 2001 but not 2002, and so on. This invokes the implicit assumption that Congresspeople are not spurred to vote in a specific way in the second year of each two-year Congress. An argument might be made that legislators who are up for re-election would vote more in line with their parties in the second year of each Congress, dragging ADA scores for Republicans downwards and Democrats upwards. However, insofar as there exist no significant differences between average ADA scores in both Democrats' and Republicans' first and second years in each Congress, I don't believe that this issue would confound our results in any essential way.

Furthermore, I did not look at the influence of majority politics in each Congress – that is, I assumed that the propensities of legislators to vote a certain way under a Democratic-majority House would not be fundamentally dissimilar to how they would vote under a Republican-majority House. This is not, however, an issue – the data set was selected such that all of the Congresses considered had a Republican majority. There is only one aberration to this selection process: the 107th Senate had, at more than one time, a Democratic majority. However, we only consider policymakers from the House of Representatives. While there might be a reason to

think that there is political spillover from the Senate to the House, the two houses have historically counterbalanced each other and acted independently.

There is also another potential downward bias in the beta value for GRADUATEEDUCATION. Recall that GRADUATEEDUCATION is only a *binary* variable for whether a legislator has attained a postgraduate degree. There is no differentiation between different types of postgraduate degrees within our model; for example, if a legislator has an M.D, it shows up as a "1" for our dummy variable in the exact same way a legislator with a J.D. would show up as a "1" for our dummy variable. However, a J.D. and an M.D. are not equally relevant to a politician's political leanings. It is debatably true that there will be a smaller effect of an M.D. degree compared to a JD degree on compelling a policymaker to vote more liberally, because issues of law are more relevant to policymaking than biology is. Thus, in our regression, we have a subset of legislators who are labeled as having postgraduate-level educations, but who will likely act less liberally than other policymakers who got their postgraduate degrees in a field more relevant to policymaking.

IV. Conclusion

In this thesis, I substantiate the hypothesis that legislators' postgraduate education attainment is a statistically significant factor correlating to higher levels of progressive voting. By looking at four different meetings of the House of Representatives, I discern 1730 functional data points which collectively point to this conclusion. An original regression indicates that having a postgraduate degree correlates to a 16.19 point increase in ADA scores, but after including party affiliation – which would otherwise incur tremendous omitted variable bias – the corrected

regression indicated that the postgraduate education would correlate to a 3.67 point increase in ADA scores.

One major challenge of this thesis was overcoming the inherent limitations to the data set – that is, the fact that our data points are not independent of each other – and we adjust our findings accordingly through a clustered regression. This does not change the 3.67 beta, but changes the standard errors from .61 to .85. Additionally, we want to include the effect of districts on ADA scores, but cannot include over 400 individual dummy variables – in order to correct for this, we use a district fixed-effect model to absorb district effects as a categorical variable. I also performed a within-group analysis for both Democrats and Republicans, and found that the effect of having a postgraduate education on ADA levels was much higher for Democrats than Republicans. I speculate that this disparity is actually higher, due to the fact that many Democrat scores are clustered at the 100-point ceiling.

Previously, we established that there are many explanations which may describe this phenomenon. The first is Maranto and Woessner's assertion that higher education is dominated by left-leaning academics; another explanation might be that higher education invites policymakers to consider issues outside of their party leanings. However, given that our data supports the fact that Democrats with postgraduate educations voted more progressively – in line with their party ideology – my findings support the comparative validity of Maranto and Woessner's claim above the latter theory.

Interpreting these econometric results for the purpose of creating a better political system is tricky. This thesis makes no normative assumption that voting progressively in key legislative debates is an inherently good thing – on the other hand, it is reasonable to assume that eliminating pure ideological voting would ameliorate friction in the political system. Our goal is not to get more legislators to vote liberally on issues, but rather to have issues considered based on their value to the U.S. in aggregate. It is not prudent, then, to prescribe a solution seeking to increase postgraduate education levels in elected representatives based on the results of this analysis. However, this information might be useful as a means of reducing information asymmetry in voters.

Works Cited

1. Dixit, Avinash, and Jorgen Weibull. "Political Polarization." *National Academy of Sciences* 104.18 (2007): 7351-356. Web. 9 Dec. 2013.

2. Kanazawa, Satoshi. "Why Liberals and Atheists Are More Intelligent." *Social Psychology Quarterly* 73.1 (2010): 33-57. Print.

3. Levine, K. J. "Voter Decision Making: The Tensions of Personal Identity, Personal Ethics, and Personal Benefit." *American Behavioral Scientist* 49.1 (2005): 63-77. Print.

4. Maranto, Robert, and Matthew Woessner. "Diversifying the Academy: How Conservative Academics Can Thrive in Liberal Academia." *PS: Political Science and Politics* 45.3 (2012): 469-74. Print.

5. Washington, Ebonya L. "Female Socialization: How Daughters Affect Their Legislator Fathers' Voting on Women's Issues." *American Economic Review* 98.1 (2008): 311-32. Prin

Table 1: Data Means and Summary

Variable	Total Legislators n=1730	Democratic Legislators n=826	Republican Legislators n=904
Independent Variables			
Level of Education			
Percentage with Post-Graduate Degree	0.640	0.707	0.579
Characteristics			
Male	0.859	0.804	0.910
Age (Mean)	65.567	66.268	64.925
Regional Representation			
New England	0.052	0.085	0.021
Mid-Atlantic	0.144	0.167	0.124
East North Central	0.163	0.153	0.171
West North Central	0.071	0.059	0.081
South Atlantic	0.178	0.146	0.208
East South Central	0.061	0.049	0.072
West South Central	0.112	0.101	0.121
Mountain	0.059	0.031	0.085
Pacific	0.161	0.209	0.117

Table 2: Models, incl. Increasing Independent Variables

Beta Coefficients and Standard Errors

	education	education, gender	education, gender, age	education, gender, age, region	education, gender, age, region, party
beta coefficient	education: 13.05*** male: - age: - region: -	education: 14.65*** male: -25.04*** age: - region: -	education: 14.75*** male: -24.53*** age: 0.30*** region: -	education: 16.19*** male: -22.87*** age: 0.25*** newengland: 24.09*** midatlantic: 1.14 eastnorthcentral: -13.80*** westnorthcentral: -18.51*** southatlantic: -19.35*** eastsouthcentral: -27.68*** westsouthcentral: -19.28*** mountain: -28.52***	education: 3.67*** male: -6.04*** age: 0.06* newengland: 9.03*** midatlantic: 2.88 eastnorthcentral: -2.43** westnorthcentral: -3.44*** southatlantic: -3.53*** eastsouthcentral: -11.60*** westsouthcentral: -9.50*** mountain: -3.67*** democrat: 74.68
standard error	education: 2.02 male: - age: - region: -	education: 1.99 male: 2.75 age: - region: -	education: 1.98 male: 2.75 age: 0.10 region: -	education: 1.90 male: 2.64 age: 0.09 newengland: 4.60 midatlantic: 3.29 eastnorthcentral: 3.18 westnorthcentral: 4.07 southatlantic: 3.12 eastsouthcentral: 4.31 westsouthcentral: 3.56 mountain: 4.31	education: 0.61 male: 0.85 age: 0.03 newengland: 1.47 midatlantic: 1.05 eastnorthcentral: 1.02 westnorthcentral: 1.31 southatlantic: 1.00 eastsouthcentral: 1.38 westsouthcentral: 1.14 mountain: 1.39 democrat: 0.61

note:

*** = significant at $p < .01$

** = significant at $p < .05$

* = significant at $p < .1$

Table 3: All Factors

Model 4: OLS, using observations 1-1765 (n = 1730)
Missing or incomplete observations dropped: 35
Dependent variable: ADA

	Coefficient	*Std. Error*	*t-ratio*	*p-value*	
const	13.5241	2.26545	5.9697	<0.00001	***
male	-6.04063	0.854965	-7.0654	<0.00001	***
democrat	74.6761	0.606861	123.0531	<0.00001	***
graduate	3.66756	0.614618	5.9672	<0.00001	***
age	0.0575487	0.0295897	1.9449	0.05195	*
newengland	9.03355	1.47258	6.1345	<0.00001	***
midatlantic	2.88016	1.04885	2.7460	0.00610	***
eastnorthcentral	-2.42746	1.01855	-2.3832	0.01727	**
westnorthcentral	-3.43952	1.30631	-2.6330	0.00854	***
southatlantic	-3.53041	1.00355	-3.5179	0.00045	***
eastsouthcentral	-11.6046	1.38209	-8.3964	<0.00001	***
westsouthcentral	-9.49967	1.14015	-8.3320	<0.00001	***
mountain	-3.66572	1.39031	-2.6366	0.00845	***

Mean dependent var	47.73873	S.D. dependent var	40.86686
Sum squared resid	243856.3	S.E. of regression	11.91741

Table 4: All Factors, Clustered

Linear regression

Number of obs	= 1730
F(12, 446)	= 1138.78
Prob > F	= 0.0000
R-squared	= 0.9156
Root MSE	= 11.917

(Std. Err. adjusted for 447 clusters in statedistrict)

ada	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
DEMOCRAT	74.6761	.9762957	76.49	0.000	72.75739	76.59481
MALE	-6.040633	1.094911	-5.52	0.000	-8.192459	-3.888807
GRADUATE	3.667555	.8538681	4.30	0.000	1.989451	5.34566
age	.0575487	.045654	1.26	0.208	-.0321749	.1472723
NEWENGLAND	9.033547	1.801211	5.02	0.000	5.493631	12.57346
MIDATLANTIC	2.880157	1.426666	2.02	0.044	.076334	5.68398
EASTNORTHCE~L	-2.427457	1.326126	-1.83	0.068	-5.033689	.1787754
WESTNORTHCE~L	-3.439522	2.210203	-1.56	0.120	-7.783227	.9041829
SOUTHATLANTIC	-3.530412	1.407508	-2.51	0.012	-6.296584	-.7642396
EASTSOUTHCE~L	-11.60456	2.281022	-5.09	0.000	-16.08745	-7.121676
WESTSOUTHCE~L	-9.499672	1.589302	-5.98	0.000	-12.62312	-6.376221
MOUNTAIN	-3.665723	1.402686	-2.61	0.009	-6.422419	-.9090275
_cons	13.52413	3.29965	4.10	0.000	7.039334	20.00892

Table 5: All Factors, with District Fixed Effects

Linear regression, absorbing indicators

Number of obs	= 1730
F(4, 1279)	= 920.89
Prob > F	= 0.0000
R-squared	= 0.9607
Adj R-squared	= 0.9469
Root MSE	= 9.4200

ada	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
DEMOCRAT	69.61922	1.176258	59.19	0.000	67.31161	71.92683
MALE	-4.929784	1.466597	-3.36	0.001	-7.806984	-2.052585
GRADUATE	3.854187	1.056803	3.65	0.000	1.78093	5.927445
age	.133522	.0450576	2.96	0.003	.045127	.221917
_cons	7.509736	3.40359	2.21	0.028	.8325031	14.18697
statedistr~t	F(446, 1279) =		4.248	0.000	(447 categories)	

Table 6: All Factors, Clustered, with Gender Interaction

Linear regression

Number of obs	= 1730
F(13, 446)	= 1055.67
Prob > F	= 0.0000
R-squared	= 0.9156
Root MSE	= 11.92

(Std. Err. adjusted for 447 clusters in statedistrict)

ada	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
GRADUATE	3.066948	1.815993	1.69	0.092	-.5020184	6.635915
MALE	-6.43554	1.414471	-4.55	0.000	-9.215396	-3.655685
DEMOCRAT	74.67841	.9760672	76.51	0.000	72.76015	76.59667
age	.0568205	.0459047	1.24	0.216	-.0333959	.1470369
NEWENGLAND	9.005071	1.801844	5.00	0.000	5.463911	12.54623
MIDATLANTIC	2.899724	1.422205	2.04	0.042	.1046693	5.69478
EASTNORTHCENTRAL	-2.408393	1.325035	-1.82	0.070	-5.012481	.1956948
WESTNORTHCENTRAL	-3.449831	2.210506	-1.56	0.119	-7.794133	.8944712
SOUTHATLANTIC	-3.512475	1.403298	-2.50	0.013	-6.270373	-.7545784
EASTSOUTHCENTRAL	-11.63739	2.291405	-5.08	0.000	-16.14068	-7.134096
WESTSOUTHCENTRAL	-9.481625	1.590909	-5.96	0.000	-12.60823	-6.355015
MOUNTAIN	-3.655015	1.406341	-2.60	0.010	-6.418893	-.8911367
maleinteraction	.7093401	1.999494	0.35	0.723	-3.220261	4.638941
_cons	13.88631	3.480558	3.99	0.000	7.045981	20.72664

Table 7: Democrats, All Factors

Model: OLS, using observations 1-843 (n = 826)
Missing or incomplete observations dropped: 17
Dependent variable: ADA

	Coefficient	*Std. Error*	*t-ratio*	*p-value*	
const	99.9188	3.43499	29.0885	<0.00001	***
newengland	4.13883	1.78804	2.3147	0.02088	**
midatlantic	-0.848721	1.43656	-0.5908	0.55482	
eastnorthcentral	-2.89758	1.43563	-2.0183	0.04388	**
westnorthcentral	-8.17689	2.01577	-4.0565	0.00005	***
southatlantic	-5.50725	1.48213	-3.7158	0.00022	***
eastsouthcentral	-20.7583	2.19976	-9.4366	<0.00001	***
westsouthcentral	-12.4833	1.66099	-7.5156	<0.00001	***
mountain	-1.89693	2.58343	-0.7343	0.46300	
male	-5.94883	1.13145	-5.2577	<0.00001	***
graduate	4.2773	0.966695	4.4247	0.00001	***
age	-0.0914192	0.0455282	-2.0080	0.04498	**

Mean dependent var	88.25303	S.D. dependent var	13.80637
Sum squared resid	121362.3	S.E. of regression	12.21039

Table 8: Republicans, All Factors

Model: OLS, using observations 1-922 (n = 904)
Missing or incomplete observations dropped: 18
Dependent variable: ADA

	Coefficient	*Std. Error*	*t-ratio*	*p-value*	
const	-0.486184	2.96011	-0.1642	0.86958	
graduate	2.43075	0.767157	3.1685	0.00158	***
male	-3.60816	1.31667	-2.7404	0.00626	***
newengland	20.9115	2.76791	7.5550	<0.00001	***
midatlantic	8.6309	1.50725	5.7262	<0.00001	***
eastnorthcentral	-0.0916304	1.42038	-0.0645	0.94858	
westnorthcentral	2.76864	1.69548	1.6330	0.10283	
southatlantic	-0.0133894	1.36553	-0.0098	0.99218	
eastsouthcentral	-3.75343	1.75471	-2.1391	0.03270	**
westsouthcentral	-5.18792	1.53144	-3.3876	0.00074	***
mountain	-1.21576	1.66111	-0.7319	0.46442	
age	0.190115	0.0374559	5.0757	<0.00001	***

Mean dependent var	10.72013	S.D. dependent var	12.25975
Sum squared resid	109317.0	S.E. of regression	11.07035