Who Becomes a Member of Congress? Evidence From De-Anonymized Census Data*

Daniel M. Thompson,† *Stanford University*

James J. Feigenbaum,‡ *Boston University and NBER*

Andrew B. Hall,§ *Stanford University*

Jesse Yoder,¶ *Stanford University*

February 25, 2019

Abstract

What types of citizens become politicians? We link former members of the U.S. Congress to the de-anonymized 1940 U.S. Census, allowing us to observe their education, earnings, and family backgrounds along with those of over 40 million similarly aged individuals. Future members between the ages of 18 and 40 in 1940 reported substantially higher earnings and education levels than non-politicians, both in raw comparisons and in matched comparisons to individuals sharing the same age, race, gender, geographical location, socioeconomic background, and occupation. Parents of future members earned more and had much higher levels of education than the general population. The results suggest that the American electoral system is best characterized as an exclusive meritocracy with regards to earning ability and education, attracting individuals with higher-than-average outside options but restricting the political opportunities of those from less privileged backgrounds.

*The protocol for this study was approved by Stanford's IRB (protocol #42719) as well as by NBER. For research assistance, the authors thank Qianmin Hu, Elise Kostial, and Anna Nakai. For helpful discussion, the authors thank Matt Gentzkow, Justin Grimmer, Greg Martin, Jaakko Meriläinen, the students of Econ 220 in the Fall 2018-2019 quarter, and the members of Stanford's Democracy Policy Lab.

†Ph.D. Candidate, Department of Political Science. dthomp@stanford.edu.

‡Assistant Professor, Department of Economics. jamesf@bu.edu.

§Corresponding author. Associate Professor, Department of Political Science. andrewbhall@stanford.edu.

¶Ph.D. Candidate, Department of Political Science. yoderj@stanford.edu.

1 Introduction

In representative democracies, elected politicians craft policy on behalf of their constituents. As a result, voter welfare depends on how competent representatives are. Models of electoral accountability with adverse selection emphasize this problem theoretically (e.g. Banks and Sundaram 1998; Besley 2004; Fearon 1999), and empirical research shows that politician characteristics matter for policy outcomes in practice (Anzia and Berry 2011; Chattopadhyay and Duflo 2004; Jones and Olken 2005).[1] Because highly skilled individuals in free-market societies can often earn more in the private sector, scholars and political thinkers have often wondered if democracies can attract them to political office.[2] Caselli and Morelli (2004: abstract) describes the idea succinctly: "Low-quality citizens have a 'comparative advantage' in pursuing elective office, because their market wages are lower than those of high-quality citizens."[3] This is a particular concern in the United States, where income inequality is high (Piketty and Saez 2003), social mobility is uneven (Chetty, Hendren, Kline, Saez, and Turner 2014), and members of Congress are systematically wealthier than the general public (Carnes 2013, 2018; Eggers and Klašnja 2018).[4] To what extent does the U.S. system select individuals who possess unusual private-sector talent for political office, and to what extent does it simply advantage those from privileged backgrounds regardless of their abilities?

Answering this question requires separating members' built-in socioeconomic advantages from their earning ability and education, something that is not possible with existing data that mainly captures members' accumulated wealth rather than their earnings prior to

[1] Citizen-candidate models (Besley and Coate 1997; Osborne and Slivinski 1996) focus on a related idea, that representatives are of distinct ideological types which affect policy outcomes.

[2] Douglas Adams captures this idea most vividly in *The Restaurant at the End of the Universe*: "To summarize: it is a well-known fact that those people who must want to rule people are, ipso facto, those least suited to do it."

[3] For related work, see Besley (2004). For an overview on political selection, see Dal Bó and Finan (n.d.).

[4] An important literature studying political ambition, and in particular the gender gap in political ambition, also documents income differences between politicians and non-politicians. See for example Lawless and Fox (2005) and Lawless and Fox (2015). For some possible consequences of the inequality in political representation, see Gilens (2012) and related literature.

becoming politicians, and that offers only partial glimpses into their socioeconomic backgrounds. To make progress, we take advantage of more than 40 million individual records from the de-anonymized 1940 U.S. Census. We link future members of Congress who were between the ages of 18 and 40 in 1940 to their Census records, and we compare their labor income (or earnings as we will often refer to it), education, and demographic characteristics to comparable citizens who do not become elected politicians, building on previous work about Sweden (Dal Bó et al. 2017).[5] Our sample spans members of Congress elected as early as the 1940s and serving as late as the 2000s, and is weighted towards the 1950s and 1960s.

We find that future members of Congress earned significantly more in 1940 than the general public, and roughly half of this overall earnings gap persists when we compare future members of Congress to non-politicians of the same occupation, education, socioeconomic background—as measured through parental occupation scores that we explain below—age, place of birth, and geographical location. Future members of Congress also report substantially more years of education, both compared to the general population and compared to individuals with similar socioeconomic backgrounds, consistent with existing work that suggests that democracies select more educated leaders (Besley and Reynal-Querol 2011). To account for family background another way, we also compare men who go on to serve in Congress to their brothers who do not enter politics, finding a much smaller though still positive gap in earnings but a relatively large gap in education. Together, these results suggest that those who seek out and win political office in America are, on average, those with better opportunities in the private market than comparable non-politicians, contrary to the idea that politics will mainly attract those with worse private-sector options.

We also find that access to political office is severely unequal in the United States, consistent with previous work. First, using the 1910, 1920, and 1930 Censuses, we locate the parents of future members of Congress who are adults in the 1940 Census, allowing us to observe parental education and occupation; and second, we find the parents of future

[5]See also Meriläinen (2018).

members of Congress who are still children living at home in 1940, allowing us to measure parents' earnings, education, and occupation directly (the 1910, 1920, and 1930 Censuses do not provide earnings). Future members of Congress grew up in households with average earnings more than twice as large as the population average, and their parents were more than four times as likely as the general population to hold college degrees. However, across all levels of parental earnings, we continue to find a pronounced earnings gap between future members of Congress and other, comparable individuals.

In the final part of the paper, we investigate some of the factors that might help explain why democracies can attract highly educated, high-earning individuals. We find that patterns of positive selection are stronger for the U.S. Senate than for the U.S. House, suggesting that the value of holding office—in terms of its prestige, its effect on future private market earnings, and the length of its term—may be important in attracting high-earning types to office. Although estimates are imprecise, we do not find much evidence that selection effects are stronger outside of the one-party south, suggesting that general-election competition may not be an important mechanism.

Our work is inspired by Dal Bó et al. (2017), which uses remarkable individual-level data on Swedish politicians and non-politicians and finds that the Swedish system is an *inclusive meritocracy*, attracting competent types to political office while offering widespread opportunities to pursue politics for people of varying socioeconomic backgrounds. Sweden is a unique country with a small, ethnically homogeneous population of roughly 10 million, a relatively even distribution of income, strong civic norms with high political participation, and a strong party system that empowers elite control over candidate recruitment. The United States is a large country with a heterogeneous population of roughly 325 million, weak civic norms with low levels of political participation, high income inequality, and a remarkably flaccid party system in which party elites have no direct levers to reward the entry of favored candidates. These differences make the United States a useful alternative test bed for beginning to accumulate knowledge about political selection across contexts.

Our results suggest that the United States political system is best conceived of as an *exclusive meritocracy*, in the framework of Dal Bó et al. (2017). Unlike Sweden, in America access to political office is heavily skewed, with members of Congress disproportionately coming from highly educated and high-earning families. Like Sweden, though, we find that those with more lucrative options in the private sector are more likely to enter politics in America, on average. This pattern of exclusive rather than inclusive meritocracy makes sense given that in America, unlike in Sweden, education and earnings ability is highly correlated with parental education and income (Bolotnyy and Bratu 2018; Chetty, Hendren, Kline, and Saez 2014), meaning that a political meritocracy on the basis of earnings ability will translate into inequality in access to political office.[6]

Drawing normative conclusions from our results requires caution, because it is unclear whether it is good or bad for society for higher earners to hold office. Are people who are better at earning money are also better at being politicians? Meriläinen (2018) presents evidence that higher earners in Finland are more efficient politicians, reducing budget gaps without harming spending programs. Existing research also suggests that more educated leaders may produce better political outcomes (Besley, Montalvo, and Reynal-Querol 2011). On the other hand, Carnes and Lupu (2015) challenges the link between higher education and political competence. And more broadly, there are reasons to worry that venal types who successfully secure higher earnings in the private sector will be more corrupt, rather than more competent, in office. For example, in *Dangers of a Salaried Bureaucracy* Benjamin Franklin feared that salaries for political offices would attract "men of strong passions and indefatigable activity in their selfish pursuits," who would be characterized by "ambition and avarice."

Even if higher earners are more competent politicians, this benefit would have to be weighed against the societal costs of having a disproportionately wealthy political class.

[6]Specifically, Bolotnyy and Bratu (2018) finds that the slope of the line relating parental income rank to child income rank is 0.182, while Chetty, Hendren, Kline, and Saez (2014) finds that this same slope is 0.341 for the United States.

These costs include a lack of descriptive representation and a loss of ideological congruence, because individuals from wealthier backgrounds may hold systematically different political views (e.g., Carreri and Teso 2017; Eggers and Klašnja 2018; Gilens and Page 2014).

Separate from questions of competence, models of accountability typically predict more effective accountability—and therefore higher voter welfare—when the benefits of office are high (Besley and Coate 1997; Besley 2004; Osborne and Slivinski 1996). Because our results imply that the benefits of office are sufficiently high to attract high earners to political office, they suggest that electoral accountability could be high in the United States—or, if it is low, that it is not because of a failure to kindle a desire for reelection among incumbents.

From a positive perspective, the similarity of our political selection results to those found in Sweden suggests a durable ability for stable, wealthy democratic systems to attract high earners to office, a pattern which is relevant for theories of political selection and electoral accountability.

2 Linking Legislators to the 1940 Census

To study the earning ability and socioeconomic backgrounds of legislators, we need to link members of Congress to their records in the 1940 Federal Census. In addition, we find members' siblings and parents by tracing members back to earlier Censuses. Because there are no unique identification numbers that link individuals across Censuses—or to our Congressional biographical data—we rely instead on a fuzzy matching process, making use of both manual matching and algorithmic matching.

2.1 Record Linking Procedures

In this section, we describe the linking procedures we use.

Biographical Information on Members of Congress

To link legislators to the 1940 U.S. Census, we begin by collecting data on the name, birth year, and birth place of every member of the U.S. House and Senate from the Biographical Directory of the U.S. Congress.[7] We match legislators to the 1940 Census because it is the most recent U.S. Census with personally identifiable information.[8] For the main analysis, we link future members of Congress who were between the ages of 18 and 40 in 1940 to their Census records, so we can compare their education and private-sector wages prior to entering Congress to comparable citizens who do not go on to become elected politicians.

Obtaining Potential Matches in 1940 Census

Using the biographical information on members of Congress, we next collect a set of possible matches in the 1940 Census for each legislator. We pull every record from the 1940 U.S. Census where a legislator matches to the Census record exactly on birth state, where the birth year in the Census is within two years of the legislator's birth year in the Congressional biographies, and where both the last name and first name are sufficiently similar using approximate string matching.[9] This outputs a list of Census records that are potential matches for each future legislator.

Manual Disambiguation of Potential Matches

For each future member of Congress we then manually search the list of potential matches and identify the correct Census record.[10] For our comparisons, the main source of bias would

[7] `http://bioguide.congress.gov/biosearch/biosearch.asp`

[8] Pesonally identifiable information on the Federal Census is restricted for 72 years after the enumeration of the Census: `https://www.Census.gov/history/www/genealogy/decennial_Census_records/the_72_year_rule_1.html`.

[9] For the approximate string matching, we use the Jaro-Winkler distance metric with a penalty parameter $p = 0.1$, implemented by the **stringdist** package in R. Jaro-Winkler string distances are common in record linkage procedures. The distances grow with character edits and transpositions and put more weight on character matching in the beginning of a string. We keep Census records where the last name and first name distances are both less than 0.3.

[10] Recent work in economics and political science on record linkage has focused on an Expectation-Maximization (EM) algorithm approach (e.g., Abramitzky, Mill, and Pérez 2018; Enamorado, Fifield,

be from the risk of false positives in the linking process, where a legislator is mistakenly linked to the Census record of a different individual. The number of future members of Congress is many orders of magnitude smaller than the number of non-members, so falsely linking the Census record of an individual from the general population as a future member of Congress will bias our estimates much more than vice versa. Therefore, our strategy is to minimize the potential for false positives in the linking (in the Appendix, we discuss potential biases associated with false negatives in our linking process). The Census includes a number of other pieces of information about each record that helps us to disambiguate between potential matches. First, the 1940 Census name entries often include middle initial or middle name, which we can cross-reference with the middle name in the legislator's biography. Second, the Census includes an occupation description as of 1940. Third, the Census includes a city code for where the individual resided in 1940, which we can occasionally compare with the member's biographical information.[11] Using this information, along with a manual comparison of name and birth year, we select the Census record that correctly matches to the future member of Congress, if there is one.[12] In Section A.1 of the Appendix, we provide some suggestive evidence that our match improves when we use our manual method.

Finding Parents and Siblings in Earlier Censuses

We also want to collect data on the parents and siblings of the future members of Congress we found as adults in 1940. Because few members are living with these relatives when we observe

and Imai n.d.). While this method is very promising for linking contemporary datasets with some measurement error in names, birthdays or addresses, the historical record linkage presents additional complications. First, our set of matching variables are more limited. We observe not date of birth in the Census but reported age (in years) and we do not have information on street address in the legislator's biography. Second, because Census data is spoken, then enumerated in cursive, and then transcribed nearly a century later, we suspect the error rates are much higher in all variables. Finally, while EM algorithms have low rates of false positives in historical data, they do so at the cost of very low match rates, failing to recover many links that manual review would code as links: `https://ranabr.people.stanford.edu/sites/g/files/sbiybj5391/f/wehc_slides_2018.pdf`.

[11] Census city codes are listed at `https://usa.ipums.org/usa-action/variables/CITY#codes_section`.

[12] To illustrate why manual matching is necessary, we find Lyndon B. Johnson, who was born in 1908, in the 1940 Census with his first name transcribed as "Lymdan B" and the birth year 1909. His occupation is listed as "Congressman," so we know it is the correct match. Of course, Lyndon B. Johnson does not ultimately enter into our analysis because he was already a member of Congress by 1940.

them in the 1940 Census, this requires us to link backwards to the 1910, 1920, and 1930 Censuses. We start by locating the future members of Congress in these previous Censuses. Because this task multiplies the number of matches we must locate, we are no longer able to do it manually. Therefore, we follow the method described in Feigenbaum (2018), training an algorithm to make links based on a sample of manually linked observations. The procedure is as follows. First, we start with the future member of Congress in 1940. We search for the set of possible matches for that member in the 1910, 1920, and 1930 Censuses.[13] Second, we manually evaluate and identify links for a sample of congress members. Third, we train a record linkage algorithm using the manually constructed data. The algorithm uses various features of a potential match—including string distances in names, year of birth differences, agreement on specific characters, phonetic name differences, and name commonness—to generate a match score. When a human is record linking, these features are used implicitly to parse matches and non-matches; the algorithm makes these implicit weights explicit and allows us to apply the linkage at scale. Fourth, we identify the matches to each Census—if a sufficiently confident match exists—for each legislator.

Once we have located a future member of congress in a previous Census, we then collect data on his or her parents and siblings. In the historical federal Censuses, households are recorded together. Thus, once we observe a legislator in 1910, we can collect data on his or her parents including name and occupation. Though earnings and education are not collected in the Federal Census until 1940, we make use of occupation score, a variable economic historians and other scholars have relied on to map disparate occupations to a continuous measure of socioeconomic status. Occupation score is based on the median earnings in each occupation in the 1950 Census and constructed by IPUMS. We can also collect information on the legislator's siblings, including first and last names and state and year of birth. We

[13]Similar to our Congress to 1940 link, this set of possible matches is anyone born in the same state, within 3 years of the 1940 year of birth, and with Jaro-Winkler string distance in first and last name of less than 0.3. The bounds on year of birth are slightly larger because there is more age and year of birth misreporting in the Census than in the Congressional data.

then use this information to link the brothers ahead to the 1940 Census.[14] We use the same machine learning approach to the Census linking described in the previous paragraph, this time searching and training data with reference to the siblings. Linking brothers to the 1940 Census enables us to observe the brothers' earnings, education, and occupation, exactly the data we collected from the 1940 Census about the legislators themselves. We consider potential biases from error in these sibling links in the estimation section below.

2.2 Dataset of Future Politicians and the General Population

Before proceeding to our analyses, we provide a brief overviewing of the final linked dataset. The dataset contains information on 494 adults in 1940 with non-missing earnings who go on to be members of Congress, 359 children in 1940 who go on to be members of Congress, and another roughly 45 million men and women of similar ages from whom we draw our various comparison groups.

Congressional Time Period Covered by the Data

Our dataset is a snapshot of individuals in 1940. Of the future members of Congress in the data, some are middle-aged adults in 1940 and are elected into office in the 1940s; others are young adults who will work in the private sector for decades before seeking office. In thinking through the subsequent analyses, it is useful to understand which congressional terms the results generally reflect. Figure A.2 presents a count of the number of members of Congress in our sample, by the years in which they entered office. The first year anyone in our sample serves in office is 1941, because we restrict to only adults who are not already members of Congress in 1940. The bulk of the adults in 1940 who go on to serve in Congress enter Congress in the 1940s, 1950s, and 1960s, with smaller numbers entering as late as the 1970s and 1980s. Since these are the individuals whose earnings we are able to investigate,

[14] We focus on brothers because sisters are likely to change their names if and when they marry, making linking on names impossible and highly biased.

the reader can think of the electoral selection analyses below as speaking primarily to the mid-20th century Congress.

Among the future members of Congress who are children in 1940, the bulk enter Congress in the 1970s and early 1980s. Because we are able to use both youths and adults in 1940 to study the socioeconomic background of members of Congress, the reader can think of the socioeconomic background analyses as covering a wider timeframe, spanning from 1940 through the 1980s, roughly.

Occupations in 1940

To give a flavor of what the data looks like, Table A.1 presents the top five most common occupations in 1940 for future members of Congress vs. the general population.[15] In both cases, the very most common occupation is "Nonclassifiable," a catch-all category. For future members of Congress, the next most common occupation is Lawyer, covering nearly 25% of the future members in our data. The next three for future members of Congress are Unemployed (which may include time spent as a student), Proprietor/Manager, and Clerical. For non-MCs, the remainder of the top 5 after Nonclassifiable is Unemployed, Laborer, Farmer, and Machine Operator.

3 Characterizing Political Selection in Congress

In this section, we use our data to characterize the differences in earnings, education, and family backgrounds between future members of Congress and the general public.

3.1 Simple Comparisons of Future MCs and the General Public

We first compare the self-reported earnings of future members of Congress to the general public. Table 1 shows key summary statistics for future members of Congress and the whole

[15] These occupations are coded from the raw census occupation strings by IPUMS. For more on this process, see: https://usa.ipums.org/usa-action/variables/OCC1950

Table 1 – Descriptive Comparisons of Future Members of Congress vs the Population. The first panel compares the earnings and education of future members of Congress who were adults in 1940 to the whole population of individuals in that age range in 1940, in order to understand personal selection into political careers. The second panel compares the earnings, education, and immigration status of the parents of future members of Congress who were children in 1940 to the population of parents of similarly aged children in 1940, in order to understand socioeconomic inequalities in access to political careers.

	Adults in 1940	
Statistic	Future MCs	Population
Wage Earnings, Mean	1464.63	493.05
Wage Earnings, 25th Percentile	0.00	0.00
Wage Earnings, Median	800.00	160.00
Wage Earnings, 75th Percentile	2500.00	800.00
Wage Earnings > $5,000	0.12	0.00
Non-Wage Earnings > $50	0.53	0.15
No Earnings and Not in School	0.05	0.32
High School Completion	0.84	0.35
College Attendance	0.73	0.12
College Completion	0.53	0.05
Share White	0.98	0.90
Share Man	0.98	0.51
N	494	44,486,892
	Youth in 1940	
Statistic	Future MCs	Population
Family Wage Earnings, Mean	1915.62	971.04
Family Wage Earnings, 25th Percentile	0.00	0.00
Family Wage Earnings, Median	1140.00	624.00
Family Wage Earnings, 75th Percentile	2508.00	1500.00
Family Member with Earnings > $5000	0.10	0.01
Family Member with Non-Wage Earnings > $50	0.59	0.45
No Family Earnings	0.03	0.04
Family Member with High School Completion	0.70	0.33
Family Member with College Attendance	0.49	0.12
Family Member with College Completion	0.32	0.05
Immigrant Family Member	0.14	0.17
N	359	18,406,686

The first column reports summary statistics for the distribution of annual earnings in 1940 for people who go on to be members of Congress. The second column reports the same statistics for people who never become members of Congress but who were born in the same window. Adults are defined as individuals born between 1900 and 1922. Youth are those born between 1923 and 1938.

population of similarly aged individuals. The upper panel studies adults who are 18-40 years old in the 1940 Census. We chose the lower threshold of 18 to focus on future MCs who could conceivably be earning an informative amount of money in the 1940 Census; we chose the upper threshold of 40 because we wanted to observe future MCs *before* they become politicians, to the extent possible. In addition to this age cutoff, we also remove any individuals who are already members of Congress in 1940.

As the top panel shows, future members of Congress earn far more than the general population, on average. Future members in 1940 earned on average $1,464, which is roughly $26,000 today, while the average for similarly aged individuals in the population was about $493, which is roughly $8,800 today. Future members exhibit a much more right-skewed distribution of earnings, with large differences at the median and the 75th percentile, and with a much higher probability of being top-coded for having earnings of $5,000 or more (roughly $90,000 in 2018). Future members are also much more likely to have non-wage earnings of $50 or more and less likely to be out of school without any earnings.

Separate from their earnings, future members of Congress were overwhelmingly more likely to complete high school, attend college, and complete college than the population as a whole. Almost three quarters of future members had attended college by 1940, while only 12% of the population of similarly aged individuals had. Finally, and here we are merely stating the obvious for any observer of the history of American politics, future members of Congress in 1940 were almost entirely white men.

The bottom panel studies the families of children in 1940, defined as anyone under the age of 18. Consistent with existing research, we see that future members of Congress come from much higher-earning households. Average parental labor income is about $1915 for future members, roughly $34,000 today, but only $971 for the parents of similarly aged children, or roughly $17,000 today. As such, parents of future members of Congress earned about twice as much labor income as the average parent of similarly aged children.

Parents of future members also exhibit a large education gap, with 49% of future members living in households where at least one family member had attended college—more than four times the rate at which other similarly aged children lived in such households. Interestingly, we find much less difference in the probability that future members of Congress live in households with at least one family member who immigrated to the U.S. Future members who are children in 1940 are only 3 percentage points less likely to live in an immigrant household than the general population.

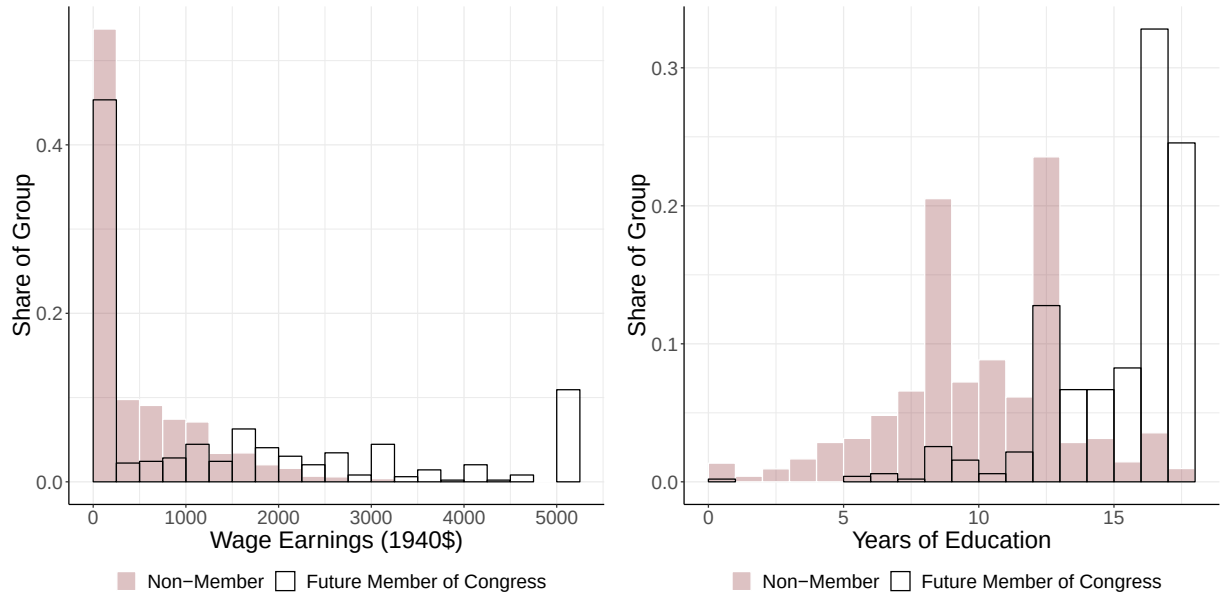
3.2 Distribution of Earnings for Future MCs and General Public

Figure 1 presents histograms of the distribution of earnings and education for each of the two groups. As the left panel of the figure shows, future members of Congress tended to out-earn other individuals, with their whole distribution shifted to the right. The large spike at the far right of the plot indicates top-coding, and the histogram shows that members of Congress were far more likely to be top-coded than other individuals. The spike at the far left of the plot indicates individuals who reported no labor income—which includes both those who simply made no money in 1940 but also those who had considerable non-wage income in 1940, such as business owners or farmers, as well as those who were over 18 but in school in 1940.

The right panel of the figure likewise shows a pronounced rightward shift in the educational attainment of future members of Congress, as measured by total years of education. A remarkably large proportion of future members of Congress attend or complete college, a level of attainment unusual in the general population in this time period.

These descriptive patterns show that members of Congress substantially out-earn the general public and have much higher average levels of education. What these patterns do not reveal, however, is how much of the overall earnings gap reflects built-in advantages of future members of Congress, vs. how much reflects positive political selection for those with

Figure 1 – Differences in the Earnings and Education of Future Members of Congress and the General Population, 1940.



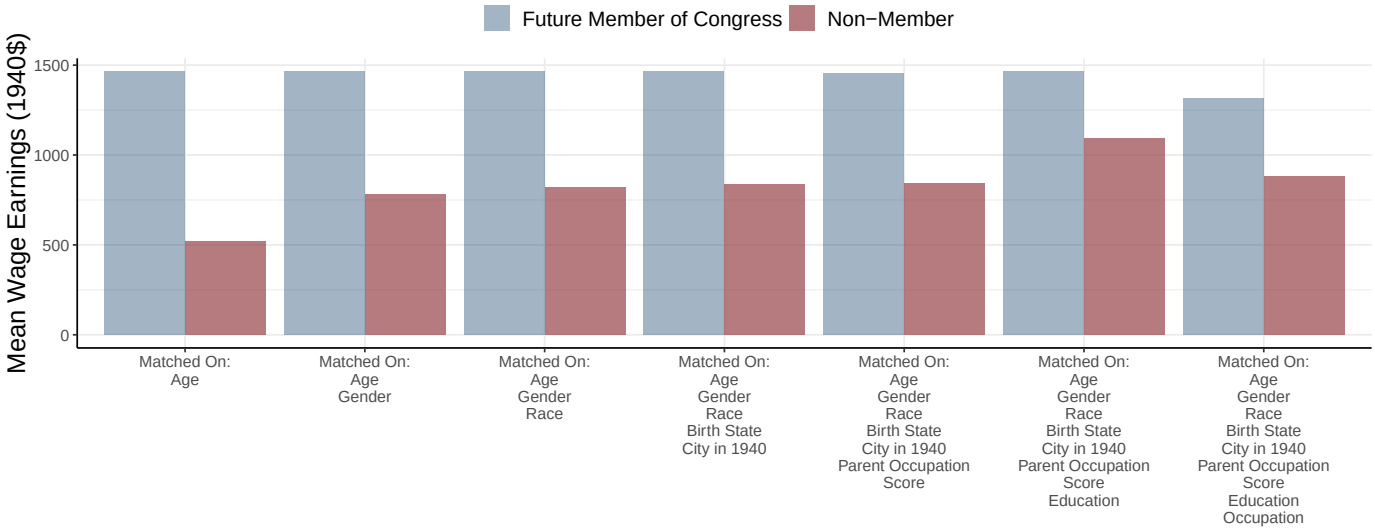
higher earning ability. Among the relatively privileged elite who have a much greater chance of becoming politicians, do those with the ability to earn more money pursue politics?

3.3 Decomposing the Overall Earnings Advantage of Future MCs

We now compare future members of Congress to increasingly strictly matched sets of individuals who do not become politicians, finding consistent evidence that future members out-earn relevant comparison sets. The purpose here is not to obtain a causal effect, but rather to explore which of the characteristics that might differ between future members of Congress and non-politicians might explain the overall gap in their earnings. This is why we sometimes match on variables that could be the downstream consequences of individual competence and ability, particularly occupation and education.

Figure 2 visualizes these comparisons. Each pair of bars in the plot indicates the mean 1940 earnings of members of Congress and a comparison set of non-members, with labels under each pair of bars indicating the covariates used to construct exact-matched comparisons sets.

Figure 2 – Differences in Earnings Between Future Members of Congress and Various Matched Comparison Groups.



In the first pair of bars to the leftmost side of the plot, we match exactly on age. For each member of Congress, we find all individuals in the Census who share the same birth year, compute their average earnings, and pair it with the earnings of the member of Congress before computing the difference in means. As we see, future members of Congress earn substantially more than their matched set of same-aged non-members.

Next, we match exactly on gender in addition to age. Here, we see that the comparison group's mean earnings rises by roughly $300, reducing the overall gap between future members of Congress and non-politicians by roughly 30%. This is largely because members of Congress in this time period are overwhelmingly likely to be men, and men are likely to earn more than women in this time period.

Next, we further match exactly on race. This leads to only a very modest increase in earnings for the comparison group. Although members of Congress are nearly all white in this time period (see Table 1), the general population is also 90% white in this time period.

Next, we further match exactly on birth state and city of residence in 1940. This barely changes the overall differences, suggesting that differences in place are not an explanation of the overall labor income gap.

Next, we also match on socioeconomic background, which we measure using parental occupation scores. As a reminder, occupation scores are based on the median earnings in each occupation in the 1950 Census and are constructed by IPUMS. Since this measure takes 80 unique values, we find matches whose family occupation scores are in the same 5-unit bin as future members of Congress. Adding socioeconomic background only modestly shrinks the difference between future members of Congress and their peers—perhaps because occupation scores are a coarse measure of socioeconomic background. As a robustness check, we match on the exact value of parent's occupation score and find very similar results with slightly smaller differences between future members and their peers while losing an additional 20% of members for whom we cannot find exact matches.

The next set of bars suggests that *a major part of the earnings advantage of future members is related to their high levels of education.* When we add years of education to the match set, we see a substantial increase in the comparison group's average earnings. In fact, roughly half of the overall difference between future members of Congress and the general public falls away once we match on education. We explore this pattern in greater detail below. Despite this advantage, though, future members still earn substantially more than other individuals of the same age, race, gender, geographical location, and education level.

Finally, we additionally match exactly on occupation. This actually decreases the comparison group's average earnings—evidently, conditional on education, future members of Congress tend to opt into somewhat lower-paying occupations. However, since the set of future members for whom we can find exact matches also has somewhat lower earnings, the estimated gap in earnings does not change much when we add occupation.

Overall Results: Future Members are Higher Earners

In sum, it appears that roughly half the overall earnings advantage of future members of Congress reflects some combination of their higher education and their gender, while the remaining half reflects differences in earning ability distinct from these factors. Comparing

Table 2 – Differences in 1940 Earnings for Future Members of Congress and Comparison Groups.

	Difference in Annual Wage Earnings, 1940, MCs vs. Non-MCs						
Statistic	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Mean	946.98	683.21	646.69	628.10	616.01	373.07	432.63
	(74.08)	(73.13)	(73.04)	(41.55)	(45.32)	(69.81)	(107.71)
25th Percentile	0.00	-45.00	-50.00	-48.00	-15.00	0.00	0.00
Median	625.00	200.00	175.00	187.50	24.00	-234.00	-35.00
75th Percentile	1636.00	1300.00	1235.00	1200.00	1200.00	700.00	500.00
Wage Earnings > $5,000	0.11	0.11	0.11	0.11	0.11	0.11	0.10
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)
Non-Wage Earnings > $50	0.37	0.31	0.31	0.31	0.29	0.24	0.17
	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.02)	(0.03)
No Earnings and Not in School	-0.27	-0.04	-0.04	-0.03	-0.03	-0.01	-0.02
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)
Members	494	494	494	476	420	253	118
Matches	44,486,892	31,640,716	28,934,288	536,625	97,494	8,037	2,350
Covariates Matched On							
Age	X	X	X	X	X	X	X
Gender		X	X	X	X	X	X
Race			X	X	X	X	X
Birth State				X	X	X	X
City in 1940				X	X	X	X
Parent Occupation Score					X	X	X
Education						X	X
Occupation							X

Each cell reports the difference between members and their matches on a particular summary statistic of the distribution of income in 1940. Each column corresponds to a particular set of covariates. Robust standard errors in parentheses. The standard errors come from regressions that include a fixed effect for ever unique set of values of the covariates. The reported statistics for matched groups are calculated by matching future members of Congress with people who never become members of Congress but who have the exact same values on the matching covariates indicated at the bottom of the table. The non-members are reweighted so that the joint distribution of all covariates is identical for the member and non-member groups.

the rightmost pair of bars with the full set of matched covariates to the leftmost pair, we see that the included covariates together appear to reduce the overall earnings gap between future members of Congress and non-politicians by roughly half, leaving a substantial earnings gap even between future members and individuals with the same education level, occupation, and socioeconomic background.

Formal Estimates

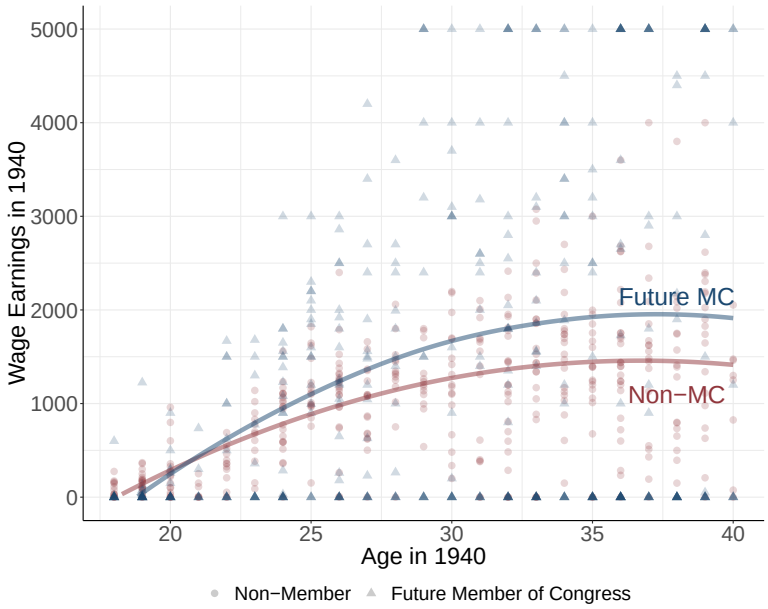
Table 2 presents these graphical comparisons in a more formal manner, allowing the reader to see the precise estimated differences and their standard errors, as well as information about sample sizes and more formal details about the matching approach. Again, the main takeaway is that future members of Congress earn considerably more in 1940 than other people, and that roughly half of this gap persists even after we compare future members only to individuals in the same occupation, with the same level of education, and with the same socioeconomic background. Rather than systematically deterring individuals with high-earning capacity from seeking political office, the American political system appears to encourage their entry and their election to office.

3.3.1 Matching on Neighborhood

Parent's occupation is a coarse measure of the socioeconomic environment in which someone grew up. Unfortunately, the 1910, 1920, and 1930 Censuses where we collect family background information did not measure earnings or education, preventing us from matching on more detailed information about parental background. To further consider socioeconomic background, we construct a measure of average occupation score for adult men with occupations in the childhood neighborhood of each future member of Congress and their potential matches.

Matching on mean neighborhood occupation score for men does not meaningfully change any of the results. Compared to column 7 in Table 2, the earnings differential *increases* by \$24, and the difference in the percent earning over \$5,000 in wages declined by less than one percentage point. We take this as evidence that matching on a larger and larger set of childhood characteristics would not meaningfully change our main findings. Since one third of the 118 future members of congress with a match in column 7 do not also have a match on mean neighborhood occupation score, we have chosen to leave the formal estimates out of the table.

Figure 3 – Earnings Advantage for Future Members of Congress and Matched Controls Across Age in 1940.



3.4 The Earnings Gap Across Age in 1940

The results above include all adults from age 18 to 40 in 1940, but we might think that it is harder to observe earnings ability among individuals who are still teenagers. These individuals may still be in school—part of why we find that members of Congress have a larger spike at 0 labor income—or they may not have had an opportunity yet to advance in their careers to a point where their talent manifests itself.

Figure 3 focuses on future members of Congress and their matched comparisons as in column 7 of Table 2, where matches are based on the full set of covariates, and it presents earnings across respondent age in 1940. As the plot shows, we see little difference in earnings for those still in their teens or early 20s. The earnings gap grows substantially for individuals in their 30s, which is also where we see the appearance of top-coded observations. Based on this plot, in the Appendix we pursue a set of analyses where we subset to future members who are 25 years or older in 1940.

3.5 Political Selection Among Lawyers

One problem with the analyses thus far is that occupation is somewhat coarsely measured. Roughly one quarter of future members of Congress in our sample have an occupation that can't be classified in the 1940 Census, so matching on occupation is not as restrictive a match as we might like. To see how this might affect the results, Table 3 replicates the analysis restricting only to lawyers—constituting roughly 25% of all future MCs in our sample—in the 1940 Census. Now all comparisons are between lawyers who go on to become members of Congress, and other lawyers who do not. As the table shows, we continue to find large differences in earnings. Interestingly, these differences grow significantly when we match on location in 1940, education, and parent occupation score. Future members of Congress who are lawyers in 1940 earn roughly $660 more than lawyers of the same age, race, and sex, born in the same state, living in the same city in 1940, with the same level of education and the same parent occupation score. Again, this seems to suggest that individuals with higher earning capacity selected into political careers.

3.6 Political Selection in Terms of Education

We now turn to studying differences in education between politicians and non-politicians in greater detail. Table 4 shows the estimated differences in education levels between future members of Congress and various comparison groups in the general population, similar to previous tables.

In column 1, we present raw comparisons to the whole population of similarly aged individuals. The first row shows the difference in total years of education between future members and the population, finding a large difference. The second, third, and fourth rows re-estimate these differences in terms of the probability of completing high school, attending college, and completing college, respectively, again finding extremely large differences.

Table 3 – Differences in 1940 Wage Earnings for Future Members of Congress and Comparison Groups, Including Lawyers Only.

Statistic	Difference in Annual Wage Earnings, 1940, MCs vs. Non-MCs (1)	(2)	(3)	(4)	(5)	(6)
Mean	394.36	392.82	394.05	668.63	550.07	662.39
	(157.09)	(157.08)	(156.86)	(144.24)	(270.90)	(340.00)
25th Percentile	0.00	0.00	0.00	0.00	0.00	0.00
Median	0.00	0.00	0.00	0.00	0.00	0.00
75th Percentile	795.00	775.00	795.00	1400.00	1400.00	1600.00
Wage Earnings > $5,000	0.10	0.10	0.10	0.15	0.14	0.21
	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)	(0.06)
Members	124	124	124	74	39	23
Matches	64,092	62,221	61,863	2,320	487	104
Covariates Matched On						
Age	X	X	X	X	X	X
Gender		X	X	X	X	X
Race			X	X	X	X
Birth State				X	X	X
City in 1940				X	X	X
Parent Occupation Score					X	X
Education						X

Each cell reports the difference between members and their matches on a particular summary statistic of the distribution of wage earnings in 1940. Each column corresponds to a particular set of covariates. Robust standard errors in parentheses. The standard errors come from regressions that include a fixed effect for ever unique set of values of the covariates. The reported statistics for matched groups are calculated by matching future members of Congress with people who never become members of Congress but who have the exact same values on the matching covariates indicated at the bottom of the table. The non-members are reweighted so that the joint distribution of all covariates is identical for the member and non-member groups.

As we go across the columns, we continue to find large differences. In column 5, we match on age, gender, race, state of birth, city of residence in 1940, and parent occupation score, and we still find massive differences in education levels.

Finally, in column 6, we match on occupation (a "post-treatment" variable from the perspective of education, if we were focused on a question of causal inference.) Here, the differences do shrink, but they are still extremely large. Compared to individuals of the same age, gender, race, birth state, city of residence in 1940, parent occupation score, and occupation, future members of Congress have almost 3 more years of schooling, on average, are 36 percentage points more likely to complete high school, and are 28 percentage points more likely to finish college.

Table 4 – Differences in 1940 Education for Future Members of Congress and Comparison Groups.

	Average Differences, MCs vs. Non-MCs					
Statistic	(1)	(2)	(3)	(4)	(5)	(6)
Years of Education	5.24	5.37	5.08	4.73	4.33	2.86
	(0.11)	(0.10)	(0.10)	(0.07)	(0.09)	(0.14)
High School Completion	0.57	0.60	0.58	0.54	0.49	0.36
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)
College Attendance	0.67	0.67	0.66	0.63	0.60	0.40
	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.02)
College Completion	0.53	0.52	0.52	0.50	0.48	0.28
	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.02)
Members	509	509	509	492	430	251
Matches	48,609,040	35,330,240	32,247,202	557,737	93,469	12,521
Covariates Matched On						
Age	X	X	X	X	X	X
Gender		X	X	X	X	X
Race			X	X	X	X
Birth State				X	X	X
City in 1940				X	X	X
Parent Occupation Score					X	X
Occupation						X

Each cell reports the difference between members and their matches on a particular summary statistic of the distribution of educational attainment in 1940. Each column corresponds to a particular set of covariates. Robust standard errors in parentheses. The standard errors come from regressions that include a fixed effect for ever unique set of values of the covariates. The reported statistics for matched groups are calculated by matching future members of Congress with people who never become members of Congress but who have the exact same values on the matching covariates indicated at the bottom of the table. The non-members are reweighted so that the joint distribution of all covariates is identical for the member and non-member groups.

In sum, future members of Congress possess much higher levels of education, compared both to the general population and to matched individuals with the same demographic characteristics and family backgrounds. This gap may indicate positive political selection on the basis of skills and earnings ability, since education causes an increase in the ability to earn (e.g., Card 1999). As Besley and Reynal-Querol (2011: 552) explains, "education is a particularly interesting aspect of political selection in view of the strong correlation found between educational attainment and earnings, which is consistent with education either enhancing skills or signaling ability. Education is also strongly correlated with civic engagement. Education is thus a compelling indicator of a leader's quality." Although other work questions the link between education and political competence (e.g., Carnes and Lupu

2015), the education gap we find is striking and at the very least indicates that individuals with higher opportunity costs are opting into politics.

3.7 Political Selection Among Male Siblings

Thus far, we have shown evidence that there is positive selection in terms of earning ability and education in the United States Congress, as measured in 1940. In decomposing these differences, we compared individuals from different households whose parents worked in occupations with similar average earnings as a proxy for socioeconomic background.

In this subsection, we pursue a different strategy to hold fixed family advantages. We locate the male siblings of future members of Congress in the 1910, 1920, and 1930 Censuses, and follow them forwards to the 1940 Census so that we can track their earnings alongside those of their politician siblings. Since these siblings grew up in the same households with the same parents, this allows us to hold family background fixed in a much stronger way than matching on parent occupation scores, which are only a crude measure since they do not capture any within-occupation variation in earnings. This could lead our previous estimates to be upward biased, if parents of future members of Congress are more likely to be high earners than parents of non-politicians working in the same occupation. On the other hand, while the sibling comparison solves this problem, because competence and earning ability may be strongly correlated within family, it may also induce a downward bias in these new estimates.

Table 5 presents the results. In the first column, we include no controls; in the latter two columns, we control for age (we cannot match exactly on age since siblings have no overlap in age), either linearly or quadratically. The rows present the estimated differences between future members of Congress and their siblings on a variety of outcome variables.

In the first row, we see that future politicians out-earn their siblings in 1940, whether or not we control for age. These differences are smaller than they were in previous analyses and we cannot reject the null of no difference, suggesting the strong link between family

Table 5 – Difference in Earnings and Education for Male Future Members of Congress and Their Brothers.

Variable	Diff Btwn MCs and Brothers		
Wage Earnings	135.44	133.79	120.99
	(95.14)	(95.35)	(89.69)
Earnings > $5,000	0.05	0.05	0.05
	(0.02)	(0.02)	(0.02)
Non-Wage Earnings > $50	0.12	0.13	0.12
	(0.03)	(0.03)	(0.03)
No Earnings and Not in School	0.02	0.02	0.02
	(0.01)	(0.01)	(0.01)
Years of Education	1.33	1.30	1.29
	(0.22)	(0.20)	(0.20)
High School Completion	0.07	0.07	0.07
	(0.03)	(0.03)	(0.03)
College Attendance	0.11	0.10	0.10
	(0.03)	(0.03)	(0.03)
College Completion	0.12	0.12	0.12
	(0.03)	(0.03)	(0.03)
Covs	None	Age	Age, Age^2
Members	324	324	324
N	1611	1611	1611

Each cell reports an estimate of the difference between members of congress and their siblings. Each row presents the differences for a different variable, including earnings from labor (wage earnings), labor earnings at or above the top of the coded earnings distribution (earnings > $5,000), at least $50 in non-wage earnings (non-wage earnings > $50), no reported capital or labor earnings and not in school (no earnings and not in school), years of education, high school completion, college attendance, and college completion. The first column is an unadjusted difference estimated using regressions that include unreported family fixed effect. The second and third columns report differences from the same fixed effects regressions after also adjusting for age differences in the siblings. Robust standard errors reported in the parentheses.

background, earnings ability, and political ambition. Nevertheless, we still see patterns of positive selection for politics, especially when we investigate other outcomes. Future members of Congress are a little more likely than their brothers to have top-coded labor income, and they are much more likely than their brothers to have non-wage income. Finally, they have more years of schooling and are substantially more likely to complete high school (7 percentage points), attend college (10 percentage points), and complete college (12 percentage points), even after controlling for age, suggesting again an important link between education and selection into politics.

It is possible that remaining error in the process that links future members to their siblings overstates these differences. We know that raw differences between future members of Congress and the general population are large; if some of the "siblings" in this analysis are not true siblings, they may upward bias the resulting comparisons. In the Appendix, we conduct two exercises to investigate this issue. First, we re-estimate the results using a subset for which we are more confident of the links; and second, we conduct a bounding exercise under the null hypothesis of no difference between siblings. Both exercises suggest that true differences between siblings may be relatively modest, but probably not zero. Especially when it comes to non-wage income and years of schooling, future members of Congress still seem importantly different than their non-politician siblings.

4 Socioeconomic Background and Access to Political Office

We already saw in Table 1 that future members of Congress are not drawn evenly from all ranks of society, but rather tend to come from families with higher-earning and more educated parents. In this section, we investigate these patterns in more detail.

Table 6 presents differences in socioeconomic background for future members of Congress who are under the age of 18 in 1940, as compared to similarly aged children who do not

Table 6 – Difference in Summary Statistics for Future Members of Congress and Different Peers Groups Between Ages 2 and 17 in 1940.

Statistic	Diff, Future MCs vs. Non-MCs		
Family Wage Earnings, Mean	946.83	898.81	685.96
	(183.71)	(183.30)	(129.44)
Family Wage Earnings, 25th Percentile	0.00	0.00	-170.00
Family Wage Earnings, Median	516.00	440.00	150.00
Family Wage Earnings, 75th Percentile	1008.00	948.00	700.00
Family Member with Wage Earnings > $5000	0.09	0.09	0.09
	(0.02)	(0.02)	(0.01)
Family Member with Non-Wage Earnings Over $50	0.14	0.14	0.17
	(0.03)	(0.03)	(0.01)
No Family Earnings	-0.01	-0.01	-0.00
	(0.01)	(0.01)	(0.01)
Family Member with High School Completion	0.38	0.36	0.28
	(0.02)	(0.02)	(0.01)
Family Member with College Attendance	0.37	0.36	0.31
	(0.03)	(0.03)	(0.01)
Family Member with College Completion	0.27	0.27	0.24
	(0.02)	(0.02)	(0.01)
Immigrant Family Member	-0.03	-0.04	-0.06
	(0.06)	(0.06)	(0.05)
Members	359	359	349
Matches	18,406,686	17,277,712	534,925
Age	X	X	X
Gender		X	X
Race		X	X
Birth State			X
City in 1940			X

Each column in this table represents a set of covariates on which we exactly match future members of Congress to members of the public who never become members of Congress. Each cell reports the difference between members and their matches on a particular summary statistic of the distribution of annual earnings in 1940. Robust standard errors in parentheses. The reported statistics for matched groups are calculated by matching future members of Congress with people who never become members of Congress but who have the exact same values on the matching covariates indicated at the bottom of the table. The non-members are reweighted so that the joint distribution of all covariates is identical for the member and non-member groups.

become members of Congress in the future. The table goes beyond the raw differences presented in Table 1 by matching on demographic covariates (we do not match on education or occupation since we do not observe these for individuals who are still children in 1940.)

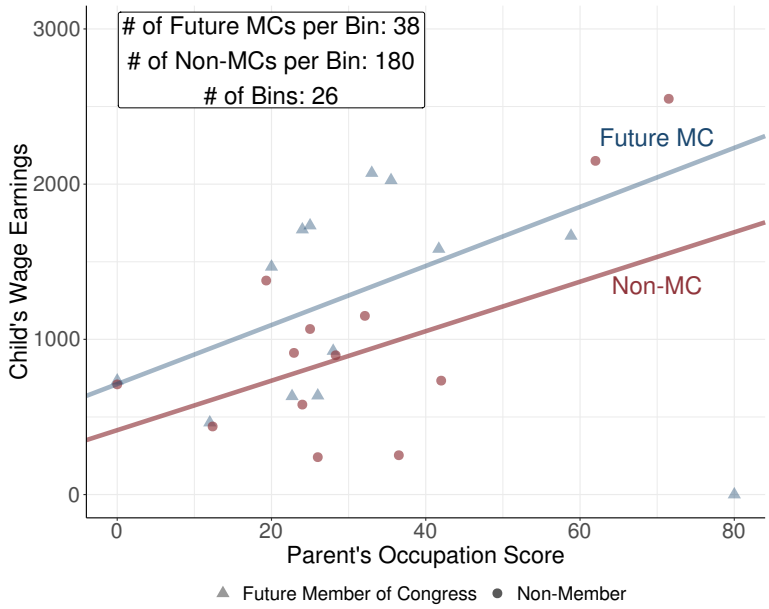
As the table shows, we continue to find significant socioeconomic advantages for future members of Congress. In fact, in the final column when we match on age, gender, race, birth state, and city of residence in 1940, we see that the earnings gap for parents of future members of Congress, at $685.96, is slightly larger than the earnings gap for adult future members of Congress themselves when matched to the population on the same covariates ($628.10; column 4 of Table 2).

We continue to find large gaps in parental education as well. In the fully matched specification, we see that parents of future members of Congress who are children in 1940 are 31 percentage points more likely to have attended college, a very large difference. We also find that future members of Congress come from households that are less likely to have an immigrant family member, but these differences are relatively small and we cannot reject the null of no difference.

Figure 4 examines the relationship between parental occupation scores and earnings for adults ages 18 to 40 in the 1940 Census. Lines are estimated separately for future members of Congress and for the comparable set of non-members exactly matched on age, race, gender, birth state, city of residence in 1940, education, and occupation, as in Table 2. As the figure shows, intergenerational labor income is highly correlated in the United States. Evidence for political selection can be seen in the plot, as the line for future members of Congress is shifted up from the line for non-members; across levels of parents' earnings, future members of Congress earn more than comparable non-politicians.

This figure is important because it helps to square our evidence from the United States with the Swedish evidence from Dal Bó et al. (2017). In Sweden, intergenerational mobility is high, so that parental income does not correlate strongly with children's income. Positive

Figure 4 – Parental Earnings, Child's Earnings, and Selection into Political Office.



electoral selection in terms of earning ability then means that Swedish politicians earn more than comparable non-politicians, but are drawn from across the parental income spectrum.

In the United States, in contrast, positive electoral selection in terms of earning ability—even if it were purely meritocratic—guarantees inequality in political access, because parental earnings and children's earnings are highly correlated.

5 Possible Mechanisms for Positive Selection into Politics

We now investigate some possible factors that could help explain how the American system is able to persuade high-earning and highly educated individuals with lucrative outside options to run for political office.

A common theme in many models of electoral accountability and running for office is the idea that larger benefits of office encourage the entry and selection of high-ability individ-

Table 7 – Comparing Political Selection Effects Across Relevant Dimensions of Heterogeneity.

	Wage Earnings, 1940		College Attendance	
	(1)	(2)	(3)	(4)
Future Member of Congress	628.10	451.02	0.56	0.52
	(41.55)	(64.97)	(0.01)	(0.02)
Future Member of Congress × Senator		696.16		0.07
		(153.38)		(0.03)
Future Member of Congress × South		44.93		0.15
		(96.48)		(0.03)
Future Member of Congress × Democrat		114.94		-0.02
		(92.92)		(0.02)
Members	476	476	539	539
Observations	537,101	537,101	632,330	632,330
Covariates Matched On				
Age	X	X	X	X
Gender	X	X	X	X
Race	X	X	X	X
Birth State	X	X	X	X
City in 1940	X	X	X	X

Regressions reported in all four columns include unreported fixed effects for each unique set of covariate values. In columns two and four, these fixed effects absorb the main effects of Senator, South, and Democrat. Robust standard errors are reported in parentheses below each estimate.

uals (Besley and Coate 1997; Besley 2004; Osborne and Slivinski 1996).[16] Empirical work also suggests that larger politician salaries leads to stronger electoral selection for quality (Gagliarducci and Nannicini 2013) and ideological moderation (Hall 2019). Although we do not have any variation in salaries to exploit—members of the House and Senate generally earn the same salary, except if they hold leadership positions—we can take advantage of the fact that the Senate is a more prestigious and more sought-after office than the House. Senators serve longer terms and hold more power by virtue of their lesser numbers, making the office more appealing. Competition is greater for senate seats, senatorial candidates raise more money, on average, and they serve on corporate boards after retiring at higher rates than members of the House (Palmer and Schneer 2016, 2018).

[16] In citizen-candidate models, higher office benefits leads to equilibria with more moderate candidates, a particular kind of "ability."

Table 7 compares the earnings gap between future members of Congress and non-members across contexts. In the first two columns the outcome variable is labor income in 1940, while in the second two columns the outcome is an indicator for attending college.

Column 1 simply replicates the overall earnings gap for future members from column 4 of Table 2, for comparison purposes. This is the specification matching on age, gender, race, and geographical location. We do not match on education in this table because it is an outcome variable in the second two columns.

In column 2, we break out the earnings gap by category. In the second row, we see that future senators have a much larger earnings gap than future House members—in fact, the earnings gap is more than twice as large for future senators.

In the third row, we see that the earnings doesn't appear to be meaningfully different in the south vs. in the rest of the country. In the third row, we see that the earnings gap is estimated to be somewhat larger for Democrats than for Republicans, but we cannot reject the null of no difference.

Columns 3 and 4 replicate these analyses for college attendance. As the second row shows, we again find a premium for future senators. The gap in rates of college attendance for future Senators is roughly 7 percentage points higher than the gap for future members of the House. The other comparisons are more mixed. We see that future members from the south are more likely to have attended college, in contrast to the earnings analysis in column 2. Also in contrast to column 2, we find no evidence that future Democratic members have more education than future Republican members.

In sum, the main pattern we find is that positive earnings and education selection is much stronger for future senators. This is consistent with the idea that higher benefits of office attract more competent individuals, and it is also consistent with evidence in Dal Bó et al. (2017), where higher salaries are shown to correlate with greater positive political selection.

6 Conclusion

This paper has studied both the earnings ability and education of future members of Congress, as well as their socioeconomic backgrounds, taking advantage of a unique historical dataset allowing us to make fine-grained comparisons between future politicians and over 40 million comparably aged citizens.

We find that future members of Congress earned substantially more money in 1940 than did individuals in the same occupation, with the same level of education, living in the same area and sharing the same race and gender, and coming from comparable socioeconomic backgrounds. Moreover, future members of Congress are substantially more educated than comparable individuals who do not go into politics. Despite reasons to think that democratic systems might struggle to attract highly skilled and highly educated individuals who have lucrative outside options in the private market, the U.S. system is evidently able to make such individuals a compelling offer. The greater selection patterns we find for the Senate compared to the House suggest that the benefits of political office are an important factor in this ability to attract high-earning and highly educated individuals to political office.

On the other hand, our work also builds on existing research showing the high levels of inequality in access to political office in America. Our data allows us to observe the parents of future members of Congress in much greater detail than existing work. We find that parents of future members of Congress earn much more than the general public, and are much, much more likely to have attended college. This is a key difference from previous findings in Sweden (Dal Bó et al. 2017). Because intergenerational mobility is much lower in the United States than in Sweden, political meritocracy on the basis of earnings ability and education in America translates into high inequality in political opportunities. While the American system evidently attracts highly educated individuals able to command high salaries in the private sector, it restricts the ability of large swaths of society to enter politics.

Our study has three main limitations. First, the ability to earn more than comparable people measures only one component of personal ability, and could be uncorrelated or even

negatively correlated with the attributes that make individuals good representatives. While the positive results of our study are important for understanding how the American political system, and democracies in general, operate, and how in particular they face problems of adverse political selection, on their own they have no clear normative implications for American government or democracy. Attracting high earners might lead to better political outcomes, under some assumptions, while under others, it could lead to worse outcomes. The same could be said of education.

Second, because the U.S. Census is only made available to researchers after a 72-year gap, we can only study people who were alive—and, in many of our analyses, already adults—in 1940. This means that our analyses primarily speak to political selection in the United States from roughly 1950 to 1970. While an important period in its own right, many factors have changed about American politics from then to today.

In particular, it appears that running for office today, in an era of severe campaign-finance demands, 24-hour tabloid media coverage, and social-media pitchfork mobs is more difficult than it used to be. Moreover, salaries for members of Congress have not kept up with inflation, meaning that holding a seat in Congress may have higher opportunity cost than it used to (Hall 2019). This may deter high-earners more than it did in the past.

Third, and finally, we are not able to identify individuals who ran for the U.S. House or Senate but who did not win office, because it is impossible to merge individuals from election data to the U.S. Census reliably without additional information, due to the high rate of full-name duplicates. Future work should separate the two stages of political selection—that is, the candidate's choice to run for office, and the electorate's choice to elect a candidate to office. Doing so will require collecting novel data on the birthdates and other characteristics of individuals who ran for but did not win political office, a significant undertaking since biographical information for such people is scant.

These limitations aside, our paper adds to the relatively new literature on political selection, exploring the conditions under which governments can solve adverse selection problems

and attract highly skilled types with appealing non-political career options. Some of the most exciting work in this field uses micro-level data to understand the selection process, but laws in many countries barring access to personal data mean that this work has been limited to Scandinavian countries. Our use of historical Census data allows us to gain some of the same advantages of the Scandinavian data while studying a very different case, the United States. That we continue to find a similar pattern of political selection suggests a somewhat durable ability for wealthy, stable democracies to attract individuals with high earning ability and high levels of education to political office.

References

Abramitzky, Ran, Roy Mill, and Santiago Pérez. 2018. "Linking Individuals Across Historical Sources: a Fully Automated Approach." NBER Working Paper 24324. http://www.nber.org/papers/w24324.

Anzia, Sarah F., and Christopher R. Berry. 2011. "The Jackie (and Jill) Robinson Effect: Why Do Congresswomen Outperform Congressmen?" *American Journal of Political Science* 55(3): 478–493.

Banks, Jeffrey S., and Rangarajan K. Sundaram. 1998. "Optimal Retention in Agency Problems." *Journal of Economic Theory* 82(2): 293–323.

Besley, Timothy. 2004. "Paying Politicians: Theory and Evidence." *Journal of the European Economic Association* 2: 193–215.

Besley, Timothy, and Marta Reynal-Querol. 2011. "Do Democracies Select More Educated Leaders?" *American Political Science Review* 105(3): 552–566.

Besley, Timothy, and Stephen Coate. 1997. "An Economic Model of Representative Democracy." *The Quarterly Journal of Economics* 112(1): 85–114.

Besley, Timothy, Jose G Montalvo, and Marta Reynal-Querol. 2011. "Do Educated Leaders Matter?" *The Economic Journal* 121(554): F205–227.

Bolotnyy, Valentin, and Cristina Bratu. 2018. "The Intergenerational Mobility of Immigrants and the Native-Born: Evidence from Sweden." Working Paper. https://scholar.harvard.edu/bolotnyy/publications/intergenerational-mobility-immigrants-and-native-born-evidence-sweden.

Card, David. 1999. "The Causal Effect of Education on Earnings." In *Handbook of Labor Economics*, ed. Orley C. Ashenfelter, and David Card. Vol. 3 Elsevier pp. 1801–1863.

Carnes, Nicholas. 2013. *White-collar Government: The Hidden Role of Class in Economic Policy Making*. University of Chicago Press.

Carnes, Nicholas. 2018. *The Cash Ceiling: Why Only the Rich Run for Office–and What We Can Do about It*. Princeton University Press.

Carnes, Nicholas, and Noam Lupu. 2015. "What Good is a College Degree? Education and Leader Quality Reconsidered." 78(1): 35–49.

Carreri, Maria, and Edoardo Teso. 2017. "Economic Recessions and Congressional Preferences for Redistribution." Working Paper.

Caselli, Francesco, and Massimo Morelli. 2004. "Bad Politicians." *Journal of Public Economics* 88(3-4): 759–782.

Chattopadhyay, Raghabendra, and Esther Duflo. 2004. "Women as Policy Makers: Evidence from a Randomized Policy Experiment in India." *Econometrica* 72(5): 1409–1443.

Chetty, Raj, Nathaniel Hendren, Patrick Kline, and Emmanuel Saez. 2014. "Where is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States." *The Quarterly Journal of Economics* 129(4): 1553–1623.

Chetty, Raj, Nathaniel Hendren, Patrick Kline, Emmanuel Saez, and Nicholas Turner. 2014. "Is the United States Still a Land of Opportunity? Recent Trends in Intergenerational Mobility." *American Economic Review* 104(5): 141–47.

Dal Bó, Ernesto, and Frederico Finan. n.d. "Progress and Perspectives in the Study of Political Selection." *Annual Review of Economics*. Forthcoming.

Dal Bó, Ernesto, Frederico Finan, Olle Folke, Torsten Persson, and Johanna Rickne. 2017. "Who Becomes A Politician?" *The Quarterly Journal of Economics* 132(4): 1877–1914.

Eggers, Andrew C., and Marko Klašnja. 2018. "Wealth, Fundraising, and Voting in the U.S. Congress." Working Paper.

Enamorado, Ted, Benjamin Fifield, and Kosuke Imai. n.d. "Using a Probabilistic Model to Assist Merging of Large-Scale Administrative Records." *American Political Science Review*. Forthcoming.

Fearon, James D. 1999. "Electoral Accountability and the Control of Politicians: Selecting Good Types versus Sanctioning Poor Performance." In *Democracy, Accountability, and Representation*, ed. Bernard Mann, Adam Przeworski, and Susan C. Stokes. Cambridge University Press pp. 55–97.

Feigenbaum, James J. 2018. "Multiple Measures of Historical Intergenerational Mobility: Iowa 1915 to 1940." *The Economic Journal* 128(612): F446–F481.

Gagliarducci, Stefano, and Tommaso Nannicini. 2013. "Do Better Paid Politicians Perform Better? Disentangling Incentives from Selection." *Journal of the European Economic Association* 11(2): 369–398.

Gilens, Martin. 2012. *Affluence and Influence: Economic Inequality and Political Power in America*. Princeton University Press.

Gilens, Martin, and Benjamin I. Page. 2014. "Testing Theories of American Politics: Elites, Interest Groups, and Average Citizens." *Perspectives on Politics* 12(3): 564–581.

Hall, Andrew B. 2019. *Who Wants to Run? How the Devaluing of Political Office Drives Polarization*. University of Chicago Press.

Jones, Benjamin F., and Benjamin A. Olken. 2005. "Do Leaders Matter? National Leadership and Growth Since World War II." *The Quarterly Journal of Economics* 120(3): 835–864.

Lawless, Jennifer L., and Richard L. Fox. 2005. *It Takes a Candidate: Why Women Don't Run for Office*. Cambridge University Press.

Lawless, Jennifer L., and Richard L. Fox. 2015. *Running From Office: Why Young Americans are Turned Off to Politics*. Oxford University Press.

Meriläinen, Jaakko. 2018. "Politician Quality, Ideology, and Fiscal Policy." Working Paper.

Osborne, Martin J., and Al Slivinski. 1996. "A Model of Political Competition with Citizen-Candidates." *The Quarterly Journal of Economics* 111(1): 65–96.

Palmer, Maxwell, and Benjamin Schneer. 2016. "Capitol Gains: The Returns to Elected Office from Corporate Board Directorships." *Journal of Politics* 78(1): 181–196.

Palmer, Maxwell, and Benjamin Schneer. 2018. "Post-Political Careers: How Politicians Capitalize on Public Office." Working Paper.

Piketty, Thomas, and Emmanuel Saez. 2003. "Income Inequality in the United States, 1913–1998." *The Quarterly Journal of Economics* 118(1): 1–41.

Online Appendix

Intended for online publication only.

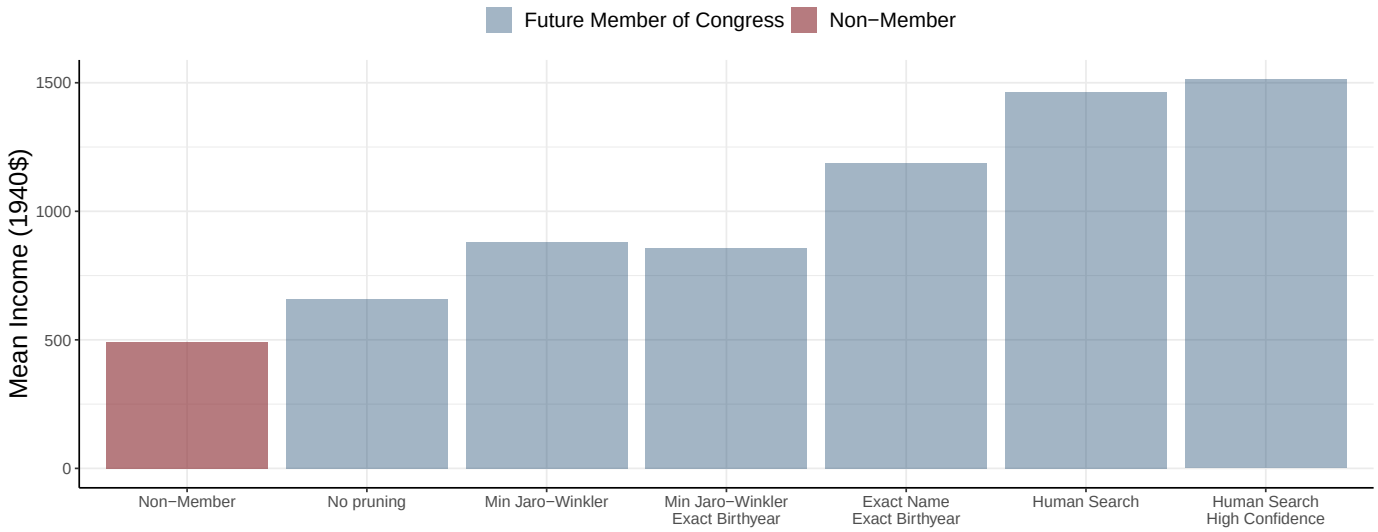
Contents

A.1 Comparing Techniques for Record Linkage 38
A.2 Years Served in Congress in Our Sample 39
A.3 Most Common Occupations in 1940 40
A.4 Comparing Members of the House vs. the Senate 40
A.5 Replicating the Results For 25-40 Year Olds 40
A.6 Socioeconomic Advantages Across Generations 41
A.7 Link Error in the Siblings Analysis 45
A.8 Bias From Future Members We Can't Link 45

A.1 Comparing Techniques for Record Linkage

Here, we compare a variety of linking strategies to validate our hand matching procedure. The red bar in Figure A.1 shows the mean earnings in 1940 dollars among the general population, while each of the blue bars shows the mean earnings in 1940 dollars among members of Congress using increasingly constrained linking criteria. We find that the difference in earnings between the members of Congress and the general population grows as we become more confident in the match accuracy. For example, the "No pruning" bar computes the mean earnings of all Census records we identify in our first pass as potential matches for members of Congress. Specifically, this includes all Census records with the same birth state, birth year within two years of the legislator's, and both last and firstname Jaro-Winkler string distance less than 0.3.[17] Using that linking procedure, most legislators have multiple potential matches in the 1940 Census, so we know there are false positives when computing the mean earnings for the "No pruning" bar. As expected, this attenuates the estimated difference between mean earnings for MCs compared to the general population.

Figure A.1 – Difference from Population Grows with Match Accuracy.



Next, we try constraining the linking criteria in other ways. In the "Min Jaro-Winkler" bar, for each legislator we select the potential Census match that minimizes our name distance metric. We take the Jaro-Winkler distance between the legislator's last name and Census record's last name along with the Jaro-Winkler distance between the legislator's first name and the Census record's first name, and we sum the absolute value of those two quantities to construct our distance metric. The difference in mean earnings between the members of Congress and general population grows when we use this more constrained linking procedure. We try a few other linking strategies, matching on exact name and exact birth year,

[17] We use the **stringdist** package in R to compute the Jaro-Winkler distance, setting the penalty parameter $p = 0.1$.

for example, but for all results in the body of the paper we hand match members to the list of potential matches in the Census. In the paper, we describe the criteria we use to select hand matches. As we can see in Figure A.1, the earnings difference grows even more using the hand matching procedure, which suggests that we are minimizing the presence of false positives in the record linkage. In the final bar, we show the mean earnings for the hand matches where we have high confidence in the match, but to keep as many observations as possible we use all of the hand matched legislators in the body of the paper.

A.2 Years Served in Congress in Our Sample

Our analyses rely on observing future members of Congress in the 1940 Census. In our main analyses, we focus on the earnings of future members of Congress who are adults in 1940, restricting the sample to future members who are between the ages of 18 and 40 as of 1940. Later, we also study those who are under the age of 18 in 1940, in order to understand socioeconomic background. Given the range of ages in 1940, what years do these individuals go on to serve in Congress? Figure A.2 shows the distribution of years that our sample serves in Congress, plotted separately for youths and adults in 1940.

Figure A.2 – Years Served in Congress in Our Sample.

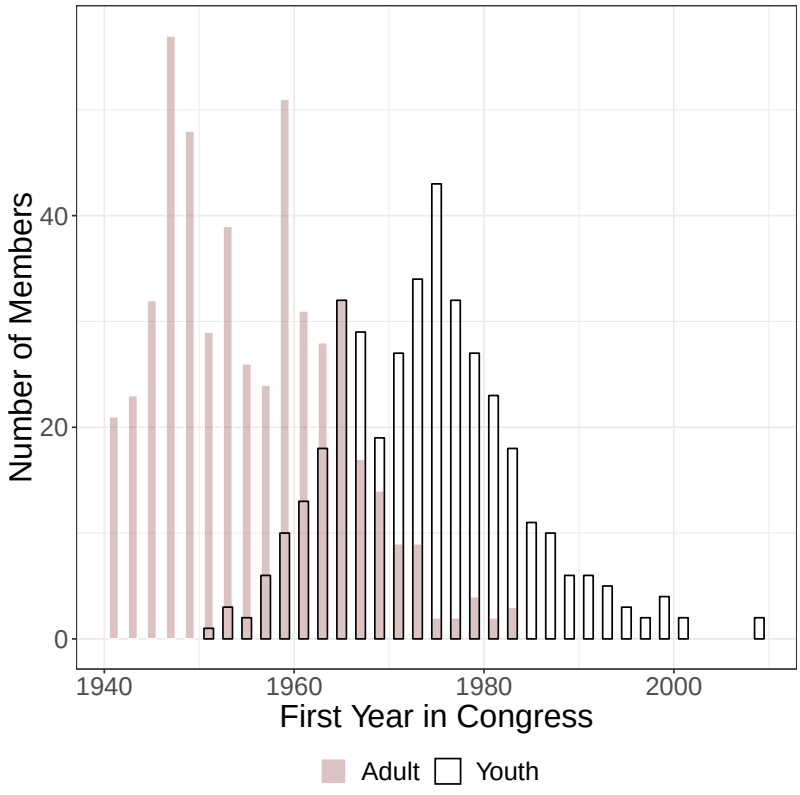


Table A.1 – Top Occupations for Future Members of Congress and Non-Members.

Rank	Future MCs		Non-MCs	
	Occupation	Share of Pop	Occupation	Share of Pop
1	Nonclassifiable	0.26	Nonclassifiable	0.20
2	Lawyer	0.24	Unemployed	0.10
3	Unemployed	0.17	Laborer	0.10
4	Proprietor, Manager	0.06	Farmer	0.06
5	Clerical	0.03	Machine Operator	0.06

A.3 Most Common Occupations in 1940

Table A.1 reports the top five occupations for future members of Congress and the population. 26% of future members of Congress and 20% of the population are working in a role that did not fit into the classifications offered by the Census or were looking for their first job. The second most common group of occupations for future members of Congress was in the law, with 24% of future members working as lawyers or judges. 17% of members of Congress were unemployed. The remaining top occupations were both professional roles. Among the population, 10% of people were unemployed. The remaining top occupations for non-members were all blue collar: laborer, farmer, and machine operator.

A.4 Comparing Members of the House vs. the Senate

As we saw in Table 7, the earnings differential is much larger among future senators than representatives. In table A.2, we track the size of the gap as we match on more and more covariates. The earnings differential between future senators and members of the public their same age in 1940 is roughly $1,550, while that same difference is $925 for future representatives. This leaves a gap between senators and representatives of about $625. By column five, when we are matching on age, gender, race, birth place, city, and education level, the difference between senators and representatives shrinks to $180, but is still present. The remaining columns are difficult to interpret given the relatively small set of future senators who have a match on all covariates in the Census.

A.5 Replicating the Results For 25-40 Year Olds

Table A.3 reports the earnings differentials for future members of Congress who are between 25 and 40 in 1940. Looking at Figure 3, it is clear that this should increase the differences, since the future members earn a similar amount as their peers at 18. The differences between future members and their peers start to appear in their mid 20s. We confirm that the differences are larger when future members are older in Table A.3. Across all seven columns, the differences are larger for future members over 25 than for future members over 18. The

earnings differential for future members is approximately two-thirds larger when we exclude people under 25, based on the estimates using the full set of covariates in column seven.

A.6 Socioeconomic Advantages Across Generations

In Table 6 we presented evidence that future members of Congress who were between the ages of two and seventeen in 1940 came from families with higher earnings and more education. Using the 1910, 1920, and 1930 Censuses, we can compare the families of future members who were between ages two and seventeen in 1940 to those who were between eighteen and forty in 1940. Table A.4 presents these results. We find that youth in 1940 come from families working in higher-earning occupations than the future members who are adults in 1940. This result must be taken with a grain of salt: getting the occupation scores for the parents of adults requires automated record linkage that may introduce some error, whereas the occupation scores the parents of a young person living at home are easily available.

Table A.2 – Differences in 1940 Wage Earnings for Future Members of Congress and Comparison Groups.

	Senators						
Statistic	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Mean	1550.42	1265.48	1221.12	1224.86	1233.17	552.24	1084.77
	(223.70)	(221.74)	(221.48)	(147.48)	(181.96)	(335.93)	(689.67)
Wage Earnings > $5,000	0.24	0.24	0.23	0.24	0.26	0.20	0.23
	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.06)	(0.11)
Non-Wage Earnings > $50	0.42	0.36	0.36	0.34	0.29	0.25	0.33
	(0.05)	(0.05)	(0.05)	(0.03)	(0.04)	(0.08)	(0.12)
No Earnings and Not in School	-0.29	-0.05	-0.05	-0.03	-0.00	0.02	-0.04
	(0.02)	(0.02)	(0.02)	(0.01)	(0.02)	(0.03)	(0.05)
Members	71	71	71	66	53	30	13
Matches	40,448,584	20,525,910	18,512,108	28,876	2,375	122	17
	Representatives						
Statistic	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Mean	924.31	662.20	626.33	611.95	596.75	367.88	457.71
	(76.66)	(75.67)	(75.56)	(42.87)	(46.06)	(71.08)	(114.79)
Wage Earnings > $5,000	0.11	0.10	0.10	0.10	0.10	0.10	0.10
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)
Non-Wage Earnings > $50	0.37	0.31	0.31	0.31	0.29	0.24	0.15
	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.02)	(0.03)
No Earnings and Not in School	-0.27	-0.04	-0.04	-0.03	-0.03	-0.01	-0.02
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)
Members	452	452	452	438	390	238	110
Matches	44,486,892	31,640,716	28,934,288	516,712	95,743	7,953	2,338
Covariates Matched On							
Age	X	X	X	X	X	X	X
Gender		X	X	X	X	X	X
Race			X	X	X	X	X
Birth State				X	X	X	X
City in 1940				X	X	X	X
Parent Occupation Score					X	X	X
Education						X	X
Occupation							X

Each cell reports the difference between members and their matches on a particular summary statistic of the distribution of earnings in 1940. Each column corresponds to a particular set of covariates. Robust standard errors in parentheses. The standard errors come from regressions that include a fixed effect for ever unique set of values of the covariates. The reported statistics for matched groups are calculated by matching future members of Congress with people who never become members of Congress but who have the exact same values on the matching covariates indicated at the bottom of the table. The non-members are reweighted so that the joint distribution of all covariates is identical for the member and non-member groups.

Table A.3 – Differences in 1940 Wage Earnings for Future Members of Congress and Comparison Groups, Aged 25 to 40 in 1940.

	Difference in Annual Wage Earnings, 1940, MCs vs. Non-MCs						
Statistic	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Mean	1280.13	951.71	907.98	881.46	853.36	542.20	726.70
	(102.87)	(102.00)	(101.92)	(56.37)	(61.55)	(100.40)	(172.11)
25th Percentile	0.00	-161.00	-195.00	-182.00	-150.00	-175.00	0.00
Median	1260.00	720.00	640.00	645.00	540.00	140.00	610.00
75th Percentile	2190.00	1750.00	1690.00	1680.00	1680.00	1300.00	900.00
Wage Earnings > $5,000	0.16	0.16	0.16	0.15	0.16	0.16	0.17
	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.02)	(0.03)
Non-Wage Earnings > $50	0.47	0.41	0.41	0.40	0.38	0.31	0.23
	(0.03)	(0.02)	(0.02)	(0.01)	(0.02)	(0.03)	(0.05)
No Earnings and Not in School	-0.29	-0.03	-0.03	-0.03	-0.03	-0.01	-0.03
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)
Members	344	344	344	334	294	170	71
Matches	27,690,636	22,209,552	20,243,256	396,515	77,882	6,363	1,825
Covariates Matched On							
Age	X	X	X	X	X	X	X
Gender		X	X	X	X	X	X
Race			X	X	X	X	X
Birth State				X	X	X	X
City in 1940				X	X	X	X
Parent Occupation Score					X	X	X
Education						X	X
Occupation							X

Each cell reports the difference between members and their matches on a particular summary statistic of the distribution of earnings in 1940. Each column corresponds to a particular set of covariates. Robust standard errors in parentheses. The standard errors come from regressions that include a fixed effect for ever unique set of values of the covariates. The reported statistics for matched groups are calculated by matching future members of Congress with people who never become members of Congress but who have the exact same values on the matching covariates indicated at the bottom of the table. The non-members are reweighted so that the joint distribution of all covariates is identical for the member and non-member groups.

Table A.4 – Difference in Parent Occupation Score Between Future Members of Congress and Their Peers, Adults v Youth.

	Diff in Parent Occ Score [0-80]	
Statistic	Adults	Youth
Mean	3.93	5.78
	(0.69)	(0.41)
25th Percentile	0.00	2.00
Median	1.00	3.00
75th Percentile	7.00	10.00
Members	446	349
Matches	46,407	534,925
Covariates Matched On		
Age	X	X
Gender	X	X
Race	X	X
Birth State	X	X
City in 1940	X	X

Each cell reports the difference between the childhood households of future members and their matches on a particular summary statistic of the distribution of occupation score. Occupation score is an estimate of the typical annual wage for an occupation in 1950 divided by 100 and is top coded at 80. Each column corresponds to a particular set of covariates. Robust standard errors in parentheses. The standard errors come from regressions that include a fixed effect for ever unique set of values of the covariates. The reported statistics for matched groups are calculated by matching future members of Congress with people who never become members of Congress but who have the exact same values on the matching covariates indicated at the bottom of the table. The non-members are reweighted so that the joint distribution of all covariates is identical for the member and non-member groups.

A.7 Link Error in the Siblings Analysis

As we discussed above, we find siblings through an automated record linkage approach while we find future members of Congress using a more manual strategy. Previous work finds that the approximately 15% of the links made by this automated record linkage strategy would be rejected by humans (Feigenbaum 2018). Given the large difference we report between future members of Congress and people with very similar background, labeling an average member of the public as a future member of Congress's brother would increase the gap between future members of Congress and the people we think are their brothers. To address this we change the target of the algorithm–rather than maximizing the average of the positive predictive value (PPV) and the true positive rate (TPR), we minimize the $.02 * TPR + .98 * PPV$. This focus on PPV while essentially ignoring TPR brings our false positive rate below 9%.

Despite using a much higher threshold of certainty for counting a member of the public as the future member of Congress's brother, we find largely similar results. As expected, most of the estimates shrink toward zero to some degree, but all of the estimates continue to show a small but noticeable positive selection.

A.8 Bias From Future Members We Can't Link

One threat to our interpretation of our results throughout the paper is error in our manual process for finding future members of Congress. If it is easier to find higher earners, this would bias our estimate of the raw earnings differential between future members of Congress and the general population. To provide a bound on this, imagine that all of the future members of Congress who we could not find in the 1940 Census were average earners. Since we were able to find 63% of all members, this would reduce our estimate of the raw earnings difference by about 40% to $595. This is a large potential bias in the overall results, but it relies on a set of assumptions about the future members who we failed to match that is incredibly strong and may even suggest the wrong bias. Perhaps instead it is easier to find future members of Congress who earn less on average; this error would produce a bias in the opposite direction. More importantly, these potential biases are not a major concern in the main analysis of the paper.

The main goal of this paper is to measure the degree of positive selection into politics and decompose its sources. The bulk of this exercise involves comparing future members of Congress to people who look like them on many observables and measure the remaining differences in private market earnings and education levels. In these analyses in which we condition on a large set of observables, we only care about biases in our linkage that still exist conditional on that large set of observables. We cannot directly test for this form of bias, but we have no reason to believe this bias exists or is driving our results.

Table A.5 – Difference in Earnings and Education for Male Future Members of Congress and Their Brothers, Certain Matches.

Variable	Diff Btwn MCs and Brothers		
Wage Earnings	77.69	75.85	78.41
	(86.41)	(86.20)	(82.79)
Earnings > $5,000	0.04	0.04	0.04
	(0.02)	(0.02)	(0.02)
Non-Wage Earnings > $50	0.08	0.09	0.09
	(0.03)	(0.03)	(0.03)
No Earnings and Not in School	0.01	0.01	0.01
	(0.01)	(0.01)	(0.01)
Years of Education	0.89	0.85	0.86
	(0.23)	(0.20)	(0.20)
High School Completion	0.02	0.01	0.01
	(0.03)	(0.03)	(0.03)
College Attendance	0.05	0.04	0.05
	(0.03)	(0.03)	(0.03)
College Completion	0.06	0.05	0.05
	(0.03)	(0.03)	(0.03)
Covs	None	Age	Age, Age^2
Members	297	297	297
N	799	799	799

Each cell reports an estimate of the difference between members of congress and their siblings. Each row presents the differences for a different variable, including earnings from labor (wage earnings), labor earnings at or above the top of the coded earnings distribution (earnings > $5,000), at least $50 in non-wage earnings (non-wage earnings > $50), no reported capital or labor earnings and not in school (no earning and not in school), years of education, high school completion, college attendance, and college completion. The first column is an unadjusted difference estimated using regressions that include unreported family fixed effect. The second and third columns report differences from the same fixed effects regressions after also adjusting for age differences in the siblings. Robust standard errors reported in the parentheses.