The Intergenerational Incidence of Government Old-Age Support: Evidence from the Early Social Security Era*

Daniel Fetter
Stanford University and NBER

Lee Lockwood
University of Virginia and NBER

Paul Mohnen
University of Michigan

July 9, 2021

Preliminary draft: please do not cite without permission

Abstract

The efficiency and distributional consequences of government old-age support depend crucially on the nature and strength of the connections between parents and their adult children. To shed light on these connections, we introduce a novel empirical approach for studying the impact of Social Security, based on differing Social Security coverage of similar occupations in the early expansion of the program. We apply this approach in a dataset linking information on the earliest beneficiaries of Social Security to long-run outcomes for their children – in particular, characteristics of the location where the children were living when they died, which we interpret as a proxy for their late-life socioeconomic status. Those whose parents were more likely to have received Social Security, or received it earlier, lived in higher-income and higher-wealth ZIP codes near the end of their lives. These impacts were associated, in part, with a greater likelihood of migration away from their location in 1930. These effects appear stronger for sons than for daughters. We also find that effects are stronger for children from smaller families, consistent with Social Security displacing family-based support of the elderly.

*For helpful comments and discussions, we thank Martha Bailey, Amy Finkelstein, Matt Notowidigdo, Heidi Williams, as well as seminar participants at the University of Michigan and Stanford University. We thank Tamri Matiashvili, Elisa Heinrich Mora, Maxwell Rong, and Qiyi Zhao for excellent research assistance. Mohnen gratefully acknowledges financial support from the Alfred P. Sloan Foundation Pre-Doctoral Fellowship on the Economics of an Aging Workforce, awarded through the NBER. Fetter and Lockwood gratefully acknowledge financial support from National Science Foundation (NSF) Grant SES-1628860.

1 Introduction

The efficiency and distributional consequences of government old-age support depend crucially on the nature and strength of the links between parents and their adult children. Absent such links, expansions of government old-age support redistribute from younger to older generations and crowd out life cycle saving for retirement and the capital stock (see, for example, Feldstein and Liebman, 2002). Strong links between parents and their adult children fundamentally transform the effects of government old-age support programs, since, with strong links, expansions of government old-age support tend to trigger offsetting changes in intergenerational transfers within families (Barro, 1974; Becker, 1974; Bernheim and Bagwell, 1988). This reduces the extent to which such expansions redistribute from younger to older generations and crowd out the capital stock, and it causes such expansions to redistribute from larger families to smaller families (since with family insurance the per-child cost of providing a given level of old-age support is decreasing in family size, whereas with government old-age support it is independent of family size). Since the provision of old-age support often involves physical proximity, this also raises the possibility that government old-age support has a variety of effects that are not often considered, including on the labor supply and migration of recipients' adult children and their families. Understanding the effect of government old-age support on intergenerational transfers within families is therefore central for evaluating these policies.

In this paper, we provide new evidence on the effect of government old-age support given intergenerational connections within families. We do so by exploring the impact of Social Security on children of recipients, focusing on the early years of the program. The introduction and subsequent expansions of the Social Security program coincided with striking changes in the connections between the elderly and their adult children. Figure 1 shows that as Social Security payments increased over the second half of the 20th century, the rate of intergenerational co-residence, a common measure of intergenerational connections in the literature (e.g., Costa, 1997, 1998, 1999; McGarry and Schoeni, 2000; Ruggles, 2007), declined substantially.[1]

Naturally, this time-series pattern does not establish a causal relationship between government-provided old-age support and within-family transfers. To that end, we introduce a novel empirical approach for estimating the causal impact of Social Security. This approach is based on the fact that some types of employment were covered when the program was introduced in 1935, while other types of employment were only covered in 1950 or later as the program was expanded. For example, work as a janitor in the private sector was covered in 1935, while work as a janitor in the non-profit sector was not covered until 1950. As a result, due to their industry of employment,

[1]One can think of co-residence as encompassing transfers of various kinds between generations (monetary and non-monetary), in both directions.

individuals with similar jobs sometimes differed in the likelihood of ever becoming eligible for Social Security, or became eligible for Social Security benefits at different ages – and so received different net transfers from Social Security. We leverage this variation to compare the long-term outcomes of otherwise-similar children whose parents received different levels of net transfers from Social Security.

Implementing this approach requires having sufficient information on parents to predict their likely eligibility for Social Security (cohort, occupation, industry, self-employment status), as well as observing their children's (future) outcomes in adulthood. In practice, datasets typically contain information on one generation but not the other. To overcome this challenge, we build a new dataset linking Social Security death records to the 1930 US Census. The 1930 Census allows us to predict the age at which parents would likely become eligible for Social Security, based on their year of birth and their employment information prior to the passage of the Social Security Act. Social Security death records contain information on children's ZIP code at death, which serves as a proxy measure for their end-of-life socio-economic status (SES) and also allows us to study their long-term migration patterns.

Our main finding is that children whose parents would likely reach eligibility for Social Security at a younger age – and so likely received larger net transfers from Social Security – died in ZIP codes characterized by a higher rank in the national income and house price distribution, in part related to greater rates of long-term migration. These effects are concentrated among children from smaller families, consistent with the idea that displacement of family support should have smaller effects on the average child from a larger family. We also find that these effects are largely driven by sons rather than daughters. Together, these findings suggest that old-age support programs like Social Security have important indirect effects on the next generation, possibly by relaxing location constraints associated with family-provided old-age support.

2 Data

Our linked dataset relies on two data sources: the 1920 and 1930 full count US Censuses, and the public version of the Social Security Numerical Identification (NUMIDENT) File released by the National Archives and Records Administration (NARA). Contained in the NUMIDENT are Social Security death records for 49 million individuals. Our primary goal is to link individuals in death records, where we can observe their ZIP code at death, to their fathers in the 1930 Census, which allows us to predict their fathers' future Social Security eligibility based on their age and 1930 employment information (five years prior to the passage of the Social Security Act). However, compared to other records in the NUMIDENT files, death records only contain basic information on individuals (full name, date of birth, and place of birth). The basis of our linked sample is a second set of records contained in the NUMIDENT: Social Security application (form SS-5) records for 40 million individuals. SS-5 records are immediately linkable to death records via Social Security Numbers, and have the advantage that they additionally list individuals' parents' full names. This allows us to link individuals and their parents as a family unit to households in the 1930 Census, which increases both the number and quality of links we are able to make. To address the fact that some children may not be living with their parents in 1930, we augment our sample by combining analogous linkages to the 1920 Census and linkages between the households in the 1920 and 1930 Censuses.

More specifically, we construct our linked data as follows. We first link siblings in the SS-5 records together using parent names. This allows us to link multiple children and their parents as one single family unit. Second, we link these reconstructed families to households in the 1920 and 1930 Censuses using parent names and information on children (first and middle names, age, and place of birth). Third, we link households in the 1930 Census to households in the 1920 Census using the same information as the previous links, except that we can additionally exploit parents' age and place of birth. Lastly, we restrict our sample to families in the SS-5 data which can be linked to both the 1920 and 1930 Censuses based on any combination of the linkages just described. Although we are mainly interested in observing fathers in 1930, observing individuals' household structure in 1920 and 1930 is important because it allows us to more accurately determine their true family size, which is a key variable in our analysis.[2]

To make each of these linkages, we adopt a supervised machine learning (ML) approach.[3] For each record we want to link to another dataset, we first construct a set of potential candidates. This set is constructed by blocking on key characteristics, computing a string distance score to rank all potential candidates within the block, and selecting the top 20 candidates. For example, when linking households in the 1930 Census to households in the 1920 Census, we block on father state/continent of birth, last name initial, and a +/- 5-year age window around father age in 1920, and compute a composite Jaro-Winkler score which captures the similarity between pairs of households in terms of father first names, father last names, and mother first names. The exact blocking scheme and string distance score varies across linkages depending on the available information. We then draw a random sample of sets and make manual linking decisions. We do this by examining the entire set of potential candidates, and either selecting a single candidate when we are confident enough, or not selecting any candidate when there is no plausible candidate or when there are multiple plausible candidates which we cannot confidently tell apart. We then use the

[2]The Censuses only provide information on family size at a specific point in time, conditional on co-residence. Moreover, reconstructed families in SS-5 records are often incomplete as parent names are not always spelled consistently across siblings, and not all children appear in the SS-5 data to begin with.

[3]The general approach described in this section comes from the Longitudinal, Intergenerational Family Electronic Micro-Database (LIFE-M) Project (https://life-m.org/).

hand-linked data to train a machine learning model to mimic human linking decisions. We use the two-stage model developed by Murray et al. (2020) specifically for one-to-one matching based on names.[4] Lastly, using the trained ML model, linking is scaled to the entire set of records we want to link. For more details on this procedure, see Mohammed and Mohnen (2021).

The key advantage of our supervised ML approach is that it allows us to achieve high match rates while controlling for the accuracy of the links we make. Essentially, disciplining the ML model using hand-linked training data allows us to explicitly control the tradeoff between making more matches and making fewer false matches. This is important because false matches and false non-matches can affect the downstream analysis in unknown ways (Bailey et al., forthcoming). In practice, for all our linkages we select the match rate associated with a false match rate of 3 percent. Note that this false match rate is relative to trainers' decisions, not the truth. In other words, our model-generated links are only as good as the links chosen by human trainers. We attempt to minimize the true error rate in the training data in two ways. First, we instruct trainers to only make links when they are confident in their decision. Importantly, because we display sets of potential candidates which may not only contain potential links but also close candidates, we are able to incorporate the degree of ambiguity in the set into the linking decision. Second, each case is independently reviewed by multiple trainers, and only cases for which there was a consensus among trainers are called links. Note that what allows our supervised ML approach to achieve high match rates in general is that human trainers are able to identify links that are non-obvious due to name misspellings, nicknames, matches on initials, and other subtleties common in name-based linking using historical data (in addition to obvious links which can be made using any approach).

Our population of interest comprises the children of men born between 1875 and 1888, who were alive in 1930, and were covered by Social Security in 1935 or 1950 based their employment information at the time. In the end, our linked sample contains around 1 million such children born between 1895 and 1930, belonging to around 600,000 unique families. The coverage rate of fathers is around 20 percent (see panel (a) in Figure 2), while the coverage rate of children is likely around 12 percent (see panels (b) and (c) in Figure 2).[5] Several factors contribute to incomplete coverage. First, only a subset of children appear in the SS-5 data, either because they never applied for a Social Security Number or due to restrictions placed on the public version of these files. In particular, the public version of the NUMIDENT released by NARA only includes individuals who died prior to 2007 and whose deaths were not state-reported. In practice, the NUMIDENT has near-universal coverage of deaths that occurred between 1988 and 2007, and relatively low

[4] For the sibling linkages which are one-to-many linking decisions, we use a random forest model.

[5] It is not possible to calculate the exact coverage rate of children as some of them are no longer living with their fathers in 1930. Instead, 12 percent is the coverage rate of children whose fathers were born in 1875-1888, regardless of their 1930-based Social Security eligibility status (which is broader than our population of interest).

coverage of deaths that occurred prior to 1988. Second, only a subset of children that appear in SS-5 records also appear in the death records, and only a subset of those have non-missing ZIP at death information. Third, human trainers are only able to link a subset of cases they examine with high confidence due to incomplete/imperfect information. Lastly, the model is only able to link a subset of the links made by human trainers at an error rate of 3 percent. Nevertheless, match rates of this magnitude are common in the historical linking literature, though a key difference between our approach and traditional automated linking methods is the high quality of the links we make.

Table 1 assesses the representativeness of our linked sample relative to the population in terms of 1930 characteristics. A few things are worth noting. First, our linked sample is naturally skewed towards later cohorts of fathers (and later cohorts of children) due to the coverage of SS-5 records. Second, because our linkages rely on children's information (i.e., observing several children improves the chances of making a link), our linked sample is by design skewed towards larger families. Lastly, as is common in name-based linking using historical data, black or lower-SES families are underrepresented in our linked sample (e.g., fathers in our linked sample are more likely to be literate, less likely to be unskilled workers, more likely to be homeowners or own a radio). This is because these individuals tend to report their names less consistently across records, which makes them harder to link. Our linked sample is fairly representative of the population along other dimensions such as occupations and geography (place of birth, place of residence).[6] To account for these differences, we re-weight our linked sample using inverse probability weights (IPW) following Bailey et al. (2020).[7] By design, the weighted means in column 3 are very close to the population means in column 1.

The NUMIDENT also includes a third set of files: Social Security claims records for 25 million individuals. We build a separate dataset linking these claims records to the 1930 Census. Here, we are linking men born in 1875–1888 in the claims records to themselves in the 1930 Census using name, age and place of birth information. These linkages are made using the same supervised ML approach described above. We are able to link around 360,000 men at a 3 percent error rate, which roughly corresponds to a coverage rate of 10 percent. We use this sample in the analysis to assess whether our prediction of fathers' eligibility for Social Security based on their 1930 employment information is predictive of their actual claiming behavior, as measured by their probability of ever appearing in the claims data and the age at which they claimed Social Security benefits conditional on appearing in the claims data.

[6] Despite uneven coverage across cohorts, SS-5 records have excellent geographical coverage. See Mohammed and Mohnen (2021) for more details.

[7] The weighting variables are dummies for father year of birth, father race, father state/continent of birth, father literacy status, homeownership status, radio ownership status, farm status, and urban status.

3 Empirical Strategy

3.1 Eligibility for Social Security retirement benefits

The first monthly retirement benefits under Social Security were paid in 1940, under the 1939 Amendments to the Social Security Act. Under the 1939 and subsequent amendments, a worker became "fully insured" under Social Security – and hence, eligible to receive monthly retirement benefits – by working for a sufficient number of quarters in employment covered by Social Security, provided he or she earned above a threshold earnings level in that quarter. For example, under the 1939 amendments, a worker was fully insured if he or she earned a sufficient amount in covered employment in half of the quarters after 1936 (or the quarter he or she turned 21) and prior to death or age 65, with a minimum of six quarters.

Our empirical approach makes use of variation in eligibility both *across* cohorts due to the required number of quarters of coverage and *within* cohorts due to the types of employment covered by Social Security at different times. The original act and the 1939 amendments covered wage and salary workers only, excluding the self-employed. Wage and salary workers in certain sectors were also excluded from coverage: in particular, domestic workers in private homes, agricultural wage workers, employees of nonprofit organizations, and local, state, and federal government employees.[8] For types of employment covered in the original Act, coverage began on January 1, 1937.[9] The 1950 amendments expanded the definition of covered employment. Coverage was compulsory for the self-employed, except for certain professional groups, farmers, regularly employed agricultural wage workers, and domestic workers. Coverage was elective for most employees of nonprofit organizations and for employees of state and local governments who were not under an existing retirement system. Importantly, the new additions to covered employment applied only to work performed in 1951 and later: coverage was not retroactive. Hence, a worker working solely in 1950-covered employment would not begin to accumulate quarters of coverage until 1951. Additional expansions to eligibility in later years covered self-employed farmers and farm managers (largely in 1954, with a further expansion in 1956), self-employed professionals who had not been covered in 1950, and various other groups.

A fully insured worker was eligible to receive monthly retirement benefits after reaching age 65, subject to an earnings test that withheld benefits in months in which earnings exceeded $15 (in 1939, this amount corresponded to about $230 in 2010 dollars). In addition, the spouse of a retired male worker received spouse's benefits (50 percent of the primary benefit) and the widow of

[8] Other exclusions included railroad workers, who were covered under a separate system under the Railroad Retirement Act, and some other groups of smaller size, such as crew members of ships and fishermen.

[9] Work done at age 65 or older was not covered under the original act, but was covered under the 1939 amendments, with an effective date of January 1, 1939. Hence, work during at least some part of 1937 and 1938 did not contribute to eligibility for workers in the 1873 and earlier birth cohorts.

a male worker received widow's benefits (75 percent of the primary benefit). The primary monthly amount was a function of the beneficiary's earnings in covered employment. Replacement rates for the median earner were around 20 percent through the 1940s, and 30 percent through the 1950s and most of the 1960s.

As has been frequently noted before (e.g., Moffitt, 1984), the initial generation of beneficiaries paid little in taxes, received large benefits, and hence received large net transfers under Social Security. A simple calculation illustrates this point. Consider a married couple with a husband born in 1880 and a wife born in 1883. Suppose that the wife did not work, but the husband worked continuously in employment covered under the 1935 act, until he was 65, and in each year earned the average for all workers in covered employment. Also suppose that both husband and wife lived exactly their remaining life expectancy at age 65. The net present value at retirement of future benefits the couple received, minus taxes paid, is on the order of $84,000 in 2010 dollars. By comparison, the median home value in 1940 was about $45,774 in 2010 dollars.[10]

3.2 Parameterizing variation in Social Security eligibility

In our empirical analysis, we map an individual's birth cohort and type of employment in 1930 into that person's likely eligibility for Social Security. Doing so addresses two issues. First, we do not directly observe taxes to or transfers from Social Security. Second, we want a measure of likely Social Security eligibility that is not a function of behavioral responses to Social Security. To parameterize variation in Social Security, we use the 1935 Act and the 1939 and 1950 amendments to calculate the earliest possible age at which an individual from a given birth cohort and with a given employment type could stop working and be permanently fully insured, assuming that workers do not change the type of their employment to one that was covered earlier.[11] Although some workers surely changed from 1950-covered employment to 1935-covered employment, we will be able to assess the degree to which employment in 1930 is predictive of later claiming of Social Security.

Figure 3, panel (a), illustrates the minimum age at which a worker from a given birth cohort would become permanently fully insured under the 1950 amendments, under the assumption that those in 1950-covered employment do not engage in any 1935-covered employment. Following this assumption, we will sometimes refer to workers as "1935-covered workers" or "1950-covered workers" even though workers could, of course, switch between 1935-covered and 1950-covered

[10] See https://www2.census.gov/programs-surveys/decennial/tables/time-series/coh-values/values-unadj.txt, accessed 2021-03-18.)

[11] "Permanently" fully insured means that a worker could stop working in covered employment and still be fully insured upon reaching age 65. The distinction arises from the fact that a worker can be "fully insured" in case of death but lose fully insured status if he or she ceased to work in covered employment. In the discussion that follows, we use the terms "permanently fully insured" or "permanently insured."

work. For hypothetical workers from a given birth cohort in different types of employment, the required number of quarters of coverage is the same; the difference between the two in the earliest age of reaching permanently insured status lies in the 14-year delay between 1935-covered work contributing to eligibility (beginning in 1937, except for those 65 or older in 1937 and 1938) and 1950-covered work contributing to eligibility (beginning in 1951). Compared to a worker in 1950-covered employment, a worker in 1935-covered employment is more likely to have worked long enough to ever be eligible, due to exit from the labor force at older ages. A worker in 1935-covered employment will also be permanently insured at an earlier age, conditional on ever being eligible. Although the difference in age at reaching permanently insured status is largely constant across birth cohorts, the effect of this difference on the likelihood of ever being eligible is heterogeneous given latent retirement behavior. Since exit from the labor force at older ages (say, 65 to 80) is greater than exit at younger ages (say, 50 to 65), the difference between 1935-covered workers and 1950-covered workers in ever being eligible tends to be greater for older cohorts, at least over some range. In addition, panel (b) of Figure 3 shows the earliest age at which a beneficiary could begin collecting monthly benefits, which is the primary parameterization of "exposure" to Social Security that we use in our analysis. As is also evident from panel (a), a worker who remains in 1950-covered employment reaches eligibility at an older age. But since no workers received monthly benefits until age 65, the difference in the earliest possible age of receiving monthly benefits is smaller for later birth cohorts.[12]

3.3 Comparing similar workers with differing Social Security eligibility

We get empirical traction on this parameterization of likely Social Security eligibility by observing an individual's age (and hence birth cohort) and employment information in the 1930 Census. For all workers, the Census recorded an occupation, industry, and class of worker (e.g. self-employed, working for wages or salary, or unpaid). We classify all combinations of these three variables into a year when this type of employment was covered by Social Security. A priori, one would expect that the types of workers who were covered in 1935 would systematically differ from those covered in 1950; the self-employed, domestic workers, and agricultural wage workers would likely have different underlying characteristics than wage and salary workers in manufacturing, for example. To the extent that these differences are level differences, constant across birth cohorts, one could net these differences out in a specification of the form

$$y_i = \alpha_{b(i)} + \beta_{c(i)} + \gamma(\text{min age of SS eligibility})_i + \Lambda'\mathbf{x}_i + \varepsilon_i \tag{1}$$

[12]Men were eligible to claim at ages 62 to 64 starting in 1961, which is relevant only to cohorts born in 1897 and later.

regressing the outcome y for household i on fixed effects for father's birth cohort b and father's coverage year c, with γ the coefficient of interest and $\mathbf{x}_i$ a vector of controls.

In practice, when making all possible comparisons between families of workers covered in the different years, the average characteristics of the groups in 1930 appear to be different. Hence, our primary specifications are based on a finer comparison. In particular, the exclusion of certain industries in the original act generates variation that allows us to make comparisons of wage and salary workers (that is, excluding the self-employed) of the same occupation, but working in different industries. As noted earlier, employees of nonprofit organizations, state and local governments, and private households were not covered under the 1935 Act or 1939 Amendments, even though work in the same, or similar, occupations in other sectors were covered by Social Security. For example, work as a janitor for a manufacturing firm would have been covered in 1935, but work as a janitor in a school, a nonprofit hospital, or a church would not have been covered until at least 1950. Differing treatment of similar workers did not escape the notice of the Social Security Board: Arthur J. Altmeyer, its Chairman, noted in 1944 that many employees of nonprofit institutions were not covered, despite the fact that their "skills, tasks, and earnings…do not usually distinguish them from comparable employees in commerce or industry" (Altmeyer, 1944).

To compare workers of the same occupation but with coverage status differing due to industry of employment, we exclude self-employed workers from our sample and estimate a specification of the form

$$y_i = \alpha_{b(i)} + \beta_{c(i)} + \gamma(\text{min age of SS eligibility})_i + \delta_{b(i)o(i)} + \Lambda'\mathbf{x}_i + \varepsilon_i \tag{2}$$

where, relative to equation (1), we include fixed effects for the interaction of father's birth cohort and father's 1930 3-digit occupation (subscripted by o) in order to isolate variation in minimum age of eligibility arising from work in different industries. Equation (2) serves as our preferred empirical specification. We use the industry classification of the Census to identify employees who are less likely than others to have been covered by the 1935 Social Security Act. Work for private households was identified as a separate industry category. Nonprofit organizations were concentrated most heavily in hospitals, educational institutions, and welfare and religious organizations, and there were also some other nonprofit membership organizations. Hospitals and educational services were also important areas of state and local government employment, in addition to more general state and local public administration. Hence, we treat wage and salary workers in these industries as not covered until 1950. It is worth noting that many of the workers we classify as not covered by Social Security until 1950 – and state and local government workers in particular – may have been covered by existing retirement systems. To the extent that workers who were not covered by Social Security had separate retirement plans, it would likely bias our reduced form estimates towards zero.

3.4 First stage

The discussion above suggests two predictions that we test in our data. First, workers whose pre-Social Security employment would be covered in 1935 will be more likely to ever claim Social Security, and conditional on claiming will claim at earlier ages, compared to workers whose employment would be covered in 1950. Second, these differences should be smaller for later birth cohorts than earlier birth cohorts, at least over some range. We begin with a descriptive comparison and subsequently show "first stage" specifications that focus on wage and salary workers and isolate variation in coverage status due to industry of employment. Figure 4 shows the distribution of age at claim, conditional on appearing in the Social Security claims data and being linked to the 1930 Census, separately for men whose 1930 employment was covered in 1935 vs. 1950, and for the first and last birth cohorts in our sample (1875 and 1888). Note that these distributions are conditional on ever claiming, and the 1950-covered men are also less likely to appear in the claims dataset at all. For the 1875 birth cohort, it is evident that 1935-covered workers claimed significantly earlier, but for the 1888 birth cohort, although differences are still present, the distribution of age at claim is significantly more similar between workers of differing coverage status.

Our parameterization of Social Security eligibility is also predictive of ever claiming Social Security, and age at claim, conditional on comparisons within fine occupation categories. Table 2 reports estimates of equation (2), which exploits variation arising from workers with the same occupation but working in different industries. Column (1) offers evidence that workers who would not reach eligibility until a later age were indeed less likely to ever claim Social Security. Eligibility one year earlier is associated with a 0.33 percentage point increase in the likelihood of being linked to a claim for monthly benefits. As a benchmark, we were able to link roughly 10 percent of men born in 1875–1888 in the 1930 Census to the claims data. Moreover, conditional on being linked to the claims data, a one-year earlier minimum age at Social Security eligibility is associated with claiming monthly benefits 0.12 years earlier. We do not formally treat these estimates as a first stage – and in particular do not scale our reduced form estimates by them – because neither fully captures the way in which a father's earlier eligibility for Social Security may influence outcomes for his children. But they do indicate that there is a robust relationship between our parameterization of Social Security eligibility and subsequent claiming behavior, even conditional on comparisons within fine occupational categories.

3.5 Selection into the linked sample

As with nearly all applications of probabilistic linking using historical data, the households we are able to link to Social Security records are a selected subset of the population. Nonrandom selection into the linked sample will matter for our primary estimates to the extent that it is correlated with the variation we use. To assess the degree to which this type of selection is a concern, we examine all households in our population of interest and test whether being in our linked sample is correlated with our variation. The results are reported in Table 3. Reassuringly, we find no evidence that a father having a lower minimum age of Social Security eligibility is correlated with the likelihood of his household appearing in our linked sample.

3.6 Placebo results in the 1930 Census

A useful check on whether our empirical approach allows for valid causal comparisons is to test whether our variation is correlated with characteristics of households in the 1930 Census, prior to the introduction of Social Security. Table 4 reports estimates based on two approaches, first testing for differences in the full population of households in the Census, and second testing for differences within the subset of households contained in our linked sample. The results confirm that conditional on making comparisons between households with fathers in the same occupation, there were few or no significant pre-existing differences between households whose fathers would reach Social Security eligibility earlier or later. With one exception out of 18 coefficients, all coefficients are statistically insignificant, and small in magnitude. To further confirm that underlying differences across households are not driving our results, in the following section we show that our results are robust to controlling for 1930 household characteristics.

4 Results

Having established that the variation we use is not correlated with pre-existing characteristics or selection into the sample, and is correlated with observed Social Security claiming patterns, we next investigate how it affected outcomes for the children of the direct beneficiaries. Throughout, we allow for heterogeneity across two key dimensions. First, we allow for heterogeneity by the number of children in the family. It is likely, for example, that the burden of caring for aging parents is greater for the average child from a small family than a large family, simply because there are fewer children who can contribute to the support of their parents. Second, motivated by the literature on intergenerational co-residence that has tended to emphasize potential differences by sex of children (e.g., Elman and Uhlenberg, 1995; Choi, 2003), we examine impacts on sons and daughters separately.

Children whose parents were ever eligible for Social Security or eligible at an earlier age lived in higher-income and higher-wealth locations late in their lives, as measured by the average Adjusted Gross Income (AGI) in 2001 and the median house price in 2000 in the ZIP code in which the children were living when they died.[13] Figure 5 plots the results of estimating equation (2), where the dependent variable is the percentile rank of the ZIP code in the national distribution of average AGI or median house prices, where outcomes are averaged across all the children observed in our linked sample for each family. In each panel, we show the coefficient on the father's minimum age of Social Security eligibility imposing a common effect across all family sizes, and also the coefficients from a separate regression where this minimum age is interacted with a set of family size fixed effects to allow for heterogeneous effects across family sizes. Note that a lower minimum age of Social Security eligibility is associated with greater likelihood of eligibility (and earlier eligibility conditional on ever being eligible), so that negative coefficients imply that fathers' greater or earlier eligibility for Social Security led their children to live in higher-income or higher-house price locations at the end of their lives.

Impacts of father's Social Security eligibility on both AGI (Panel (a) of Figure 5) and median house prices (Panel (b)) show similar patterns. Consistent with the expectation that a shift from family-based to government-based old-age support has greater impacts on children from small families, effects are greater in magnitude, and more often statistically distinguishable from zero, for children from smaller families. Interestingly, although there is some suggestive evidence that daughters' late-life locations are affected by Social Security, we find that effects tend to be larger, and more often statistically distinguishable from zero, for sons. In terms of magnitude, a ten-year difference in the earliest age of father's eligibility – roughly the difference between 1935- and 1950-covered fathers from the 1877 birth cohort – is associated with a 2 to 4 percentile difference in the ZIP at death income or house price rank, for sons from families with up to 5 children. These effects roughly correspond to 3 to 6 percent of the baseline means in our linked sample (63 and 68 respectively). Interpreting children's location late in life as a proxy for their socio-economic status, these results suggest that larger net transfers to the early beneficiaries of Social Security affected not just the beneficiaries themselves, but also had positive effects on their children's long-term outcomes.

These results are robust to a wide range of alternative specifications. We show results from these alternatives in Figures 6 and 7, reporting baseline coefficients alongside those from specifications that include controls for 1930 household characteristics, 1930 county fixed effects, and specifications that do not use inverse probability weights to adjust for non-representativeness of

[13] Data on ZIP-level mean AGI comes from the Statistics of Income (SOI) Division of the Internal Revenue Service while data on ZIP-level median house prices comes from the National Historical Geographic Information System (Manson et al., 2020).

the linked sample. All specifications are qualitatively similar.

One possible channel through which children's outcomes may be affected by their parents' receipt of Social Security is through a higher likelihood of migration – if, for example, government old-age support substitutes for family-based in-kind support that requires physical proximity. To assess the degree to which parents' receipt of Social Security affected children's likelihood of migrating, Figure 8 estimates similar specifications examining the effect of Social Security on children's likelihood of migrating out of their 1930 county or 1930 state of residence, where final location is measured by their ZIP code of residence at death. Parents' eligibility for Social Security led to greater rates of out-migration from children's 1930 county and 1930 state of residence, once again with larger impacts on children from smaller families than for larger families, and larger impacts for sons than for daughters. A ten-year difference in the earliest age of father's eligibility is associated with a 5 to 8 percentage point greater likelihood of migrating to a different county, and a 2 to 5 percentage point greater likelihood of migrating to a different state, for sons from families with up to 5 children. These effects roughly correspond to 6 to 13 percent of the baseline means in our linked sample (0.63 and 0.37 respectively).

In our full analysis, we will explore various further dimensions of heterogeneity in the effects of Social Security. Here we report estimates of one dimension, heterogeneity by father's race, and for one outcome, average AGI of children's ZIP code at death (estimates for median house prices reveal similar patterns). It is likely that Social Security would have had differential impacts depending on the wealth or income of the parents and their families. Consistent with this idea, Figure 9 shows that father's Social Security eligibility had a substantially larger impact on average incomes in sons' ZIP code at death for sons of black fathers than for sons of white fathers. A specification not interacted with family size suggests that a father reaching eligibility for Social Security ten years earlier was associated with an increase of black sons' ZIP at death income rank of 8 to 9 percentiles. Notably, and perhaps consistent with the results above, we see little evidence of differencial effects by race for daughters, although the estimates for daughters of black fathers are too imprecise to reach any definitive conclusion.

5 Conclusion

The incidence and welfare impacts of government old-age support programs like Social Security depend crucially on how they affect intergenerational relationships within families. Developing our empirical understanding of the impact of Social Security on recipients' children has been difficult, however, due to the scarcity of data with information on both parents and non-co-resident children, in addition to the challenge of finding empirical variation in Social Security for otherwise similar individuals. To address these challenges, we introduce a novel empirical approach for studying the early Social Security program and implement it in a new dataset linking Census records to Social Security administrative data. Taken together, our results indicate that parents' eligibility for Social Security had substantial impacts on children's long-run outcomes, particularly for sons and children from smaller families. These results suggest the importance of accounting for behavioral responses within the family for evaluating the incidence or welfare impacts of pay-as-you-go old age programs such as Social Security.

References

Altmeyer, Arthur J., "The Desirability of Extending Social Security to Employees of Nonprofit Institutions," *Social Security Bulletin*, August 1944, *7* (8).

Bailey, Martha J., Connor Cole, and Catherine Massey, "Simple Strategies for Improving Inference with Linked Data: A Case Study of the 1850–1930 IPUMS Linked Representative Historical Samples," *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 2020, *53* (2), 80–93.

_ , _ , Morgan Henderson, and Catherine Massey, "How Well Do Automated Linking Methods Perform? Lessons From U.S. Historical Data," *Journal of Economic Literature*, forthcoming.

Barro, Robert J., "Are Government Bonds Net Wealth?," *Journal of Political Economy*, 1974, *82* (6), 1095–1117.

Becker, Gary S., "A Theory of Social Interactions," *Journal of Political Economy*, 1974, *82* (6), 1063–1093.

Bernheim, B. Douglas and Kyle Bagwell, "Is Everything Neutral?," *Journal of Political Economy*, April 1988, *96* (2), 308–338.

Carter, Susan B., Scott Sigmund Gartner, Michael R. Haines, Alan L. Olmstead, Richard Sutch, Gavin Wright, Price V. Fishback, and Melissa A. Thomasson, *Historical Statistics of the United States Millennial Edition Online*, New York: Cambridge University Press, 01 2006.

Choi, Namkee, "Coresidence between Unmarried Aging Parents and Their Adult Children: Who Moved in With Whom and Why?," *Research on Aging*, 07 2003, *25*, 384–404.

Costa, Dora L., "Displacing the Family: Union Army Pensions and Elderly Living Arrangements," *Journal of Political Economy*, 1997, *105* (6), 1269–1292.

_ , *The Evolution of Retirement* National Bureau of Economic Research Books, University of Chicago Press, 1998.

_ , "A House of Her Own: Old Age Assistance and the Living Arrangements of Older Nonmarried Women," *Journal of Public Economics*, 1999, *72* (1), 39–59.

Elman, Cheryl and Peter Uhlenberg, "Co-residence in the Early Twentieth Century: Elderly Women in the United States and Their Children," *Population Studies*, 1995, *49* (3), 501–517.

Feldstein, Martin and Jeffrey B. Liebman, "Social Security," in A. J. Auerbach and M. Feldstein, eds., *Handbook of Public Economics*, Vol. 4 of *Handbook of Public Economics*, Elsevier, 2002, chapter 32, pp. 2245–2324.

Manson, Steven, Jonathan Schroeder, David Van Riper, Tracy Kugler, and Steven Ruggles, *IPUMS National Historical Geographic Information System: Version 15.0 [dataset]*, Minneapolis, MN: IPUMS, 2020.

McGarry, Kathleen and Robert Schoeni, "Social Security, Economic Growth, and the Rise in Elderly Widows' Independence in the Twentieth Century," *Demography*, May 2000, *37* (2), 221–236.

Moffitt, Robert A., "Trends in Social Security Wealth by Cohort," in "Economic Transfers in the United States" NBER Chapters, National Bureau of Economic Research, Inc, December 1984, pp. 327–358.

Mohammed, Shariq and Paul Mohnen, "Black Economic Progress in the Jim Crow South: Evidence from Rosenwald Schools," Working paper 2021.

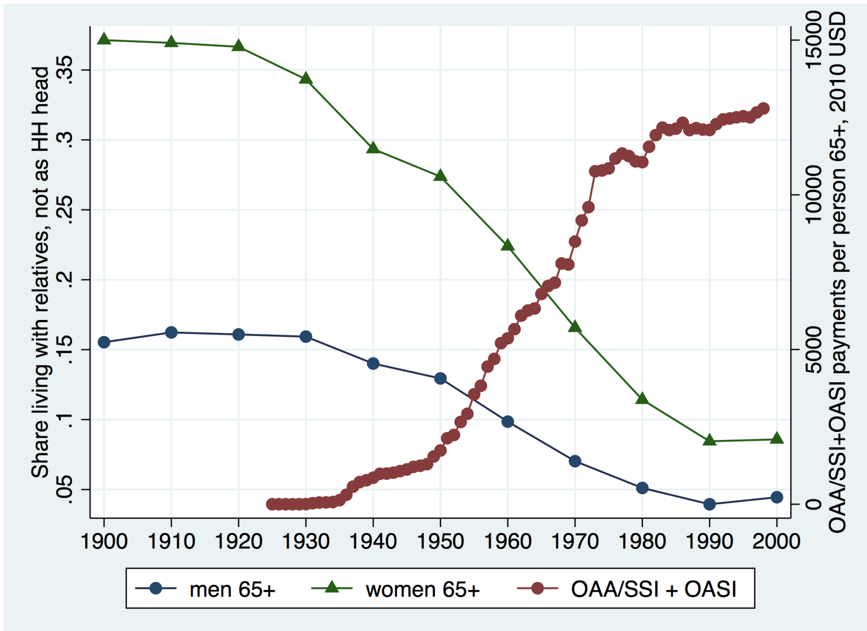
Murray, Jared, Martha J. Bailey, and Connor Cole, "The Highlander Probability Model: Power and Precision from Imposing Constraints in One-to-One Matching," Working paper 2020.

Parker, Florence E., "Experience Under State Old-Age Pension Acts in 1935," *Monthly Labor Review*, October 1936, *34* (4), 811–837.

Ruggles, Steven, "The Decline of Intergenerational Coresidence in the United States, 1850 to 2000," *American Sociological Review*, 2007, *72* (6), 964–989.

Figures and Tables

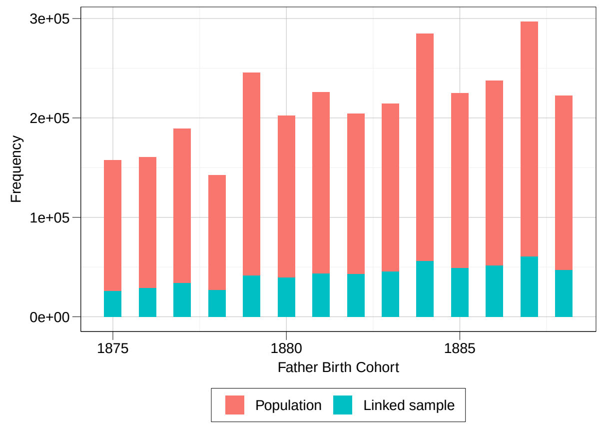
Figure 1: Intergenerational co-residence over the 20th century



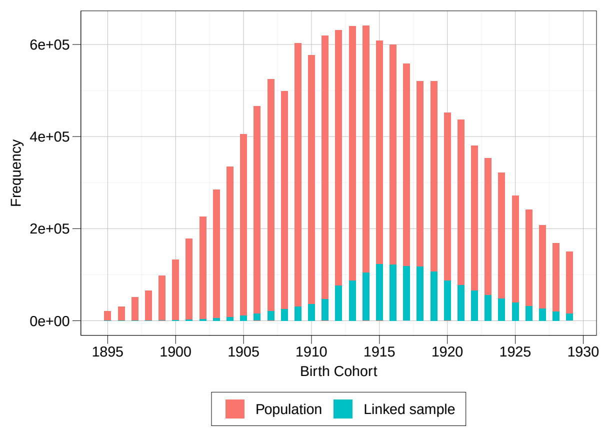
Notes: This graph plots the share of men and women aged 65 or older living with a relative, not as a household head or spouse of the household head. Also displayed are total payments under Old Age Assistance (OAA) and Old Age and Survivors Insurance (OASI) divided by the 65+ population (in 2010 US dollars). OAA payments data come from Parker (1936) for 1925 to 1935 and Series Bf621 of Carter et al. (2006) for 1936 onwards. OASI payments data come from Series BF396 of Carter et al. (2006).

Figure 2: Coverage of men born in 1875-1888 and their children born in 1895-1929

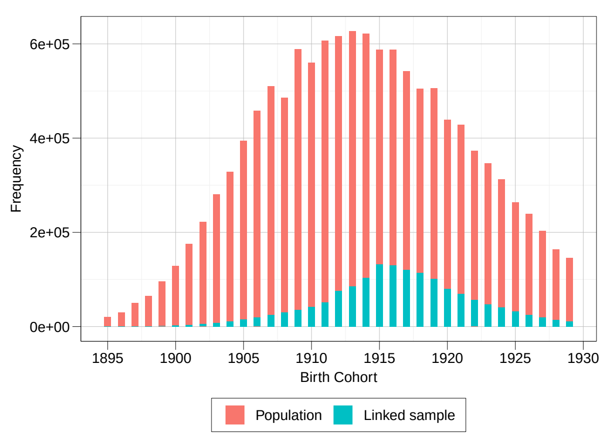
(a) Men born in 1875-1888, covered by Social Security in 1935 or 1950 based on their employment in 1930



(b) Sons of men born in 1875-1888



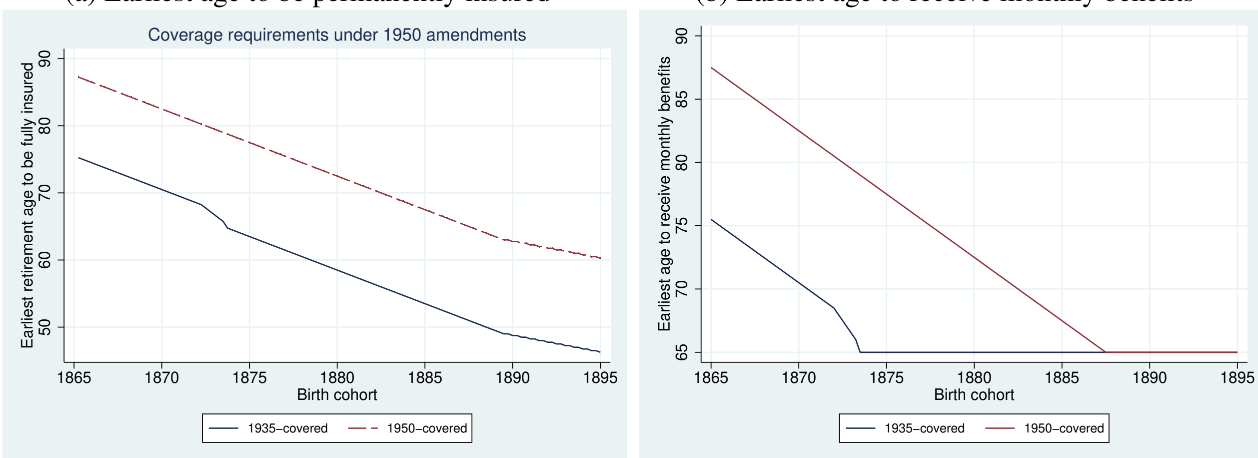
(c) Daughters of men born in 1875-1888



Notes: Population in panel (a): men born in 1875-1888 in the 1930 Census, whose employment was covered by Social Security in 1935 or 1950 based on their employment in 1930. Linked sample in panel (a): fathers in linked sample with analogous restrictions. Population in panels (b) and (c): boys/girls born in 1895-1909/1910-1919/1920-1929 in the 1910/1920/1930 Censuses, whose fathers were born in 1875-1888. Linked sample in panels (b) and (c): boys/girls born in 1895-1929 in linked sample, whose fathers were born in 1875-1888.

Figure 3: Earliest age when permanently insured and earliest age for benefits

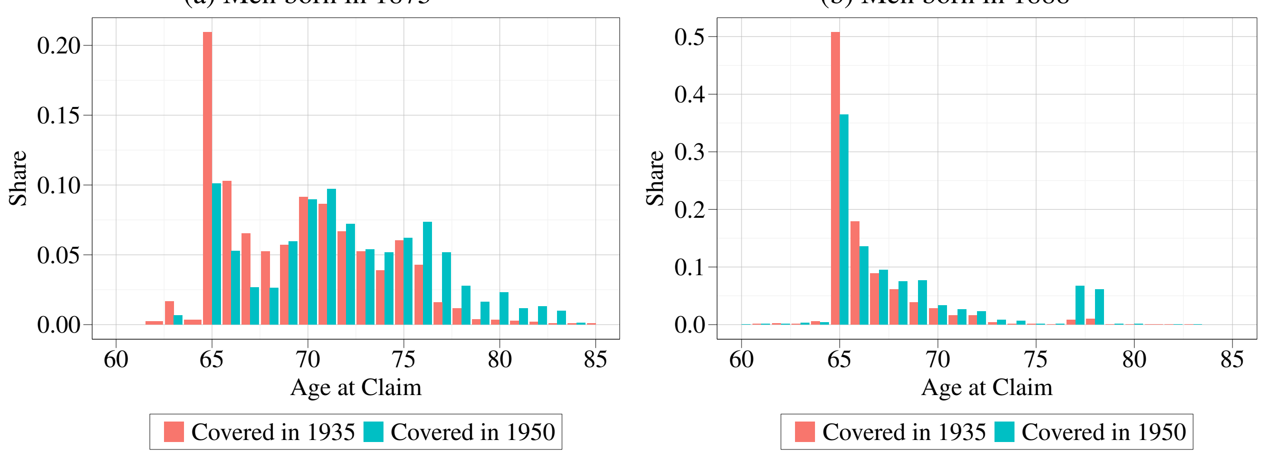
(a) Earliest age to be permanently insured (b) Earliest age to receive monthly benefits



Notes: minimum ages based on Social Security Act of 1935, 1939 amendments, and 1950 amendments, assuming that a worker in 1950-covered employment has no employment that was covered in 1935. The minimum ages shown are those under the 1950 amendments.

Figure 4: Age at claim by coverage of 1930 employment

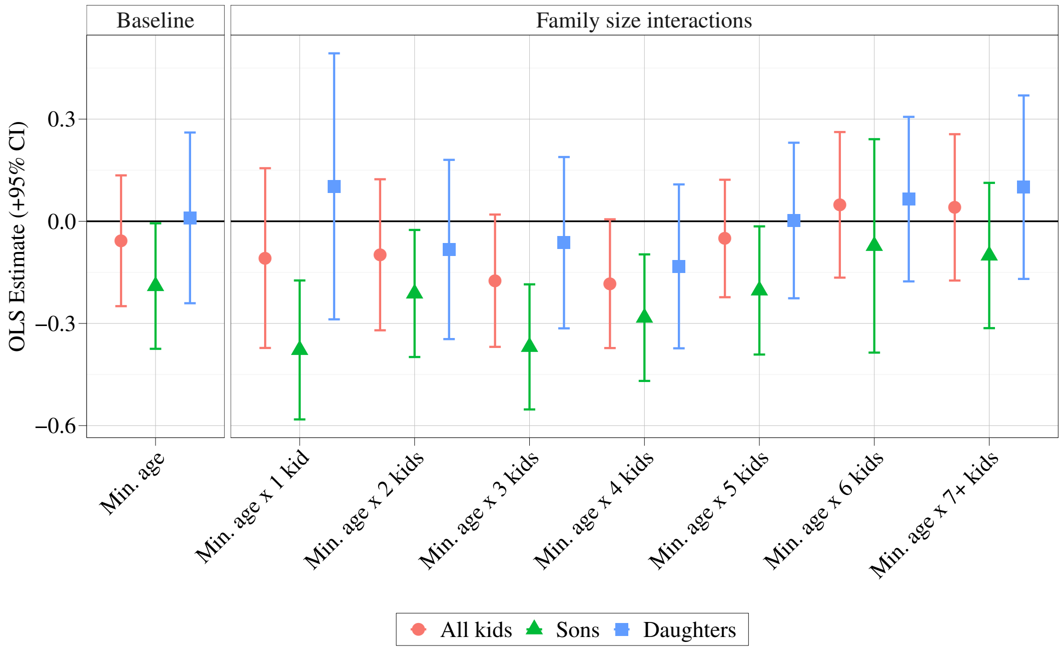
(a) Men born in 1875 (b) Men born in 1888



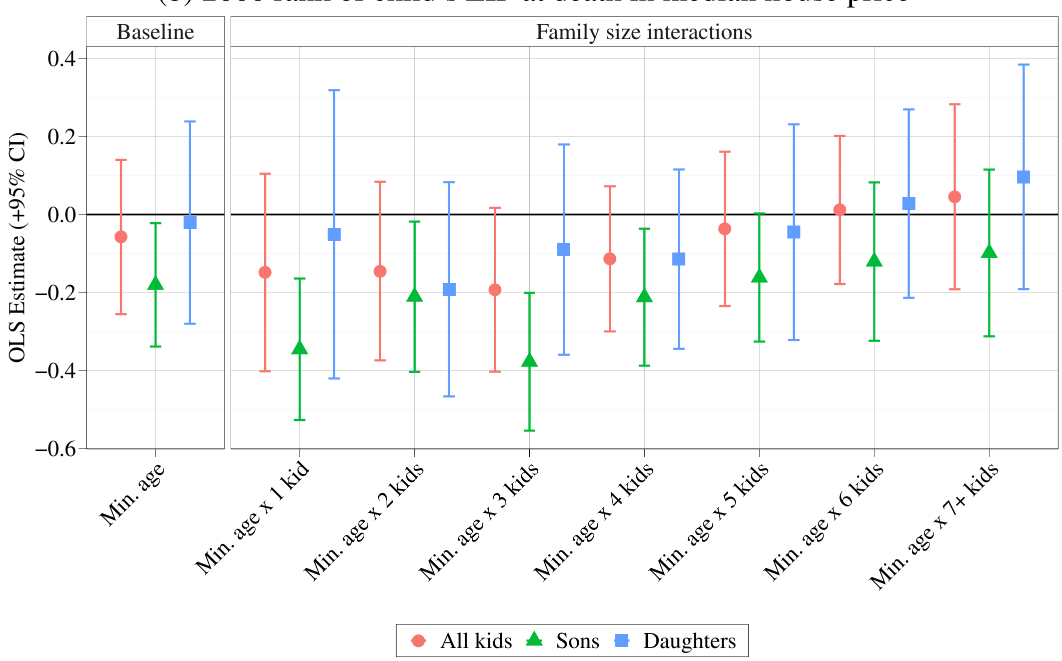
Notes: Graphs based on claims-1930 Census linked sample. Observations are weighted using IPWs.

Figure 5: Effect of later Social Security coverage of father's employment on children's outcomes

(a) 2001 rank of child's ZIP at death in average Adjusted Gross Income (AGI)



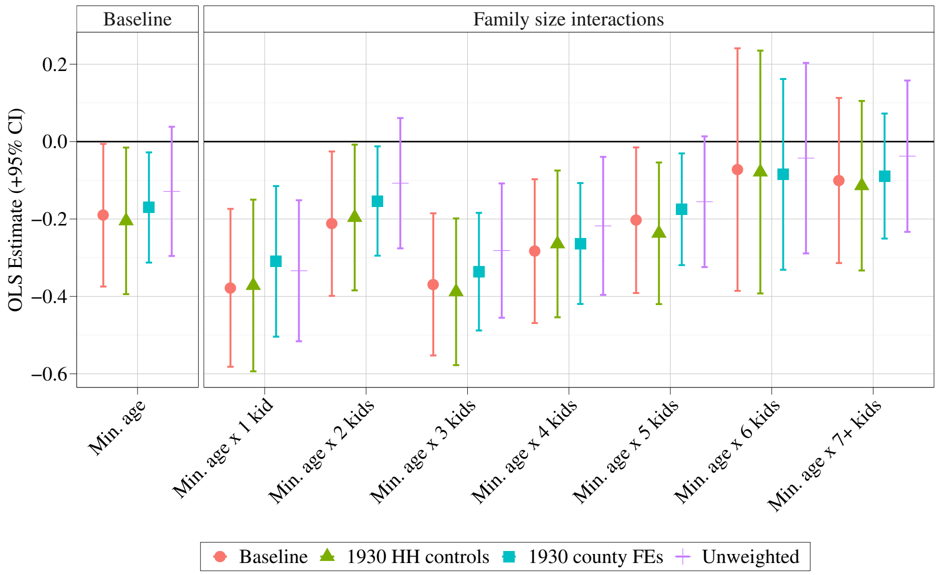
(b) 2000 rank of child's ZIP at death in median house price



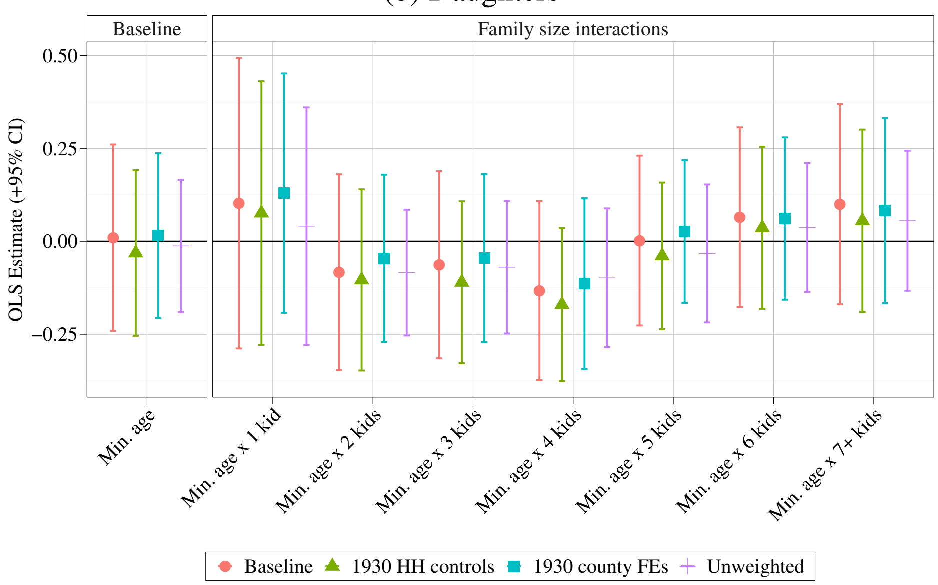
Notes: Figures show coefficients on father's minimum age at Social Security eligibility in estimates of equation (2), where the outcome is the average percentile rank of his sons' or daughters' ZIP code at the time of their death in the distribution of AGI (the unit of observation is the household). Baseline estimates (left) do not interact minimum age with family size; family size interactions (right) interact minimum age with fixed effects for the number of children in the family observed across the 1920 and 1930 Censuses. All regressions include fixed effects for father's occupation $\times$ year of birth, Social Security coverage year, race, and state/country of birth. Observations are weighted using IPWs. 95% confidence intervals are based on standard errors clustered at the level of father's industry of employment.

Figure 6: Robustness of result on AGI of children's ZIP code

(a) Sons



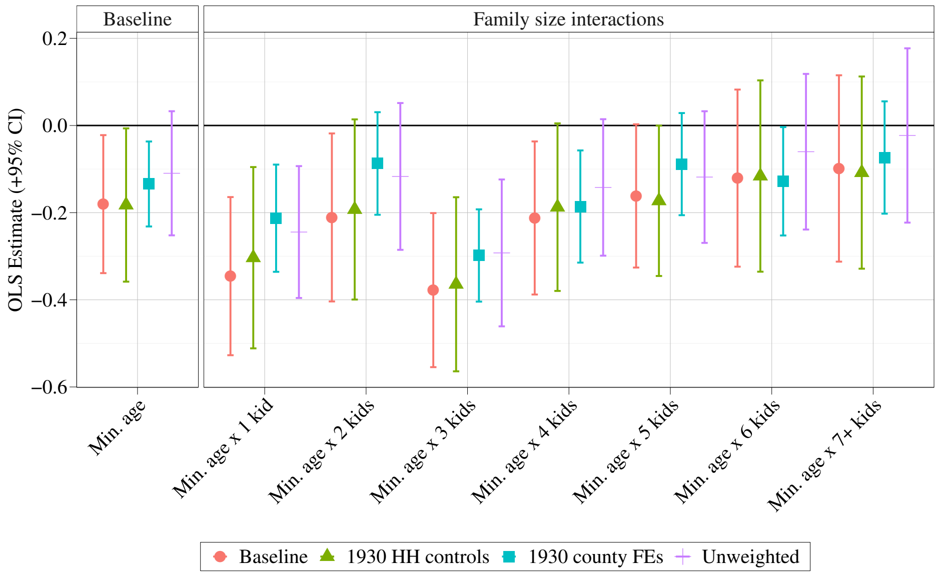
(b) Daughters



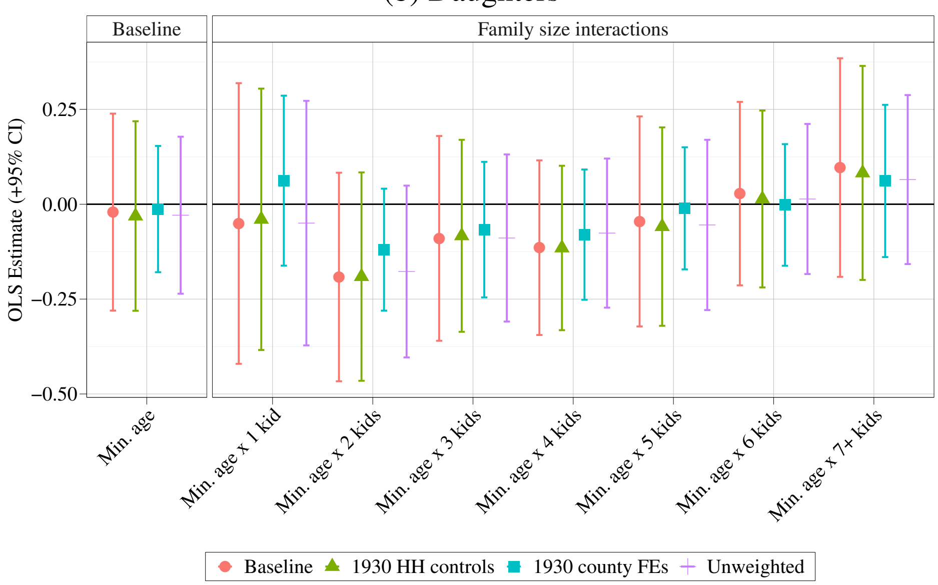
Notes: Figures show coefficients on father's minimum age at Social Security eligibility in estimates of equation (2), where the outcome is the average percentile rank of his sons' or daughters' ZIP code at the time of their death in the distribution of AGI (the unit of observation is the household). Baseline estimates (left) do not interact minimum age with family size; family size interactions (right) interact minimum age with fixed effects for the number of children in the family observed across the 1920 and 1930 Censuses. All regressions include fixed effects for father's occupation $\times$ year of birth, Social Security coverage year, race, and state/country of birth. Observations are weighted using IPWs. 95% confidence intervals are based on standard errors clustered at the level of father's industry of employment.

Figure 7: Robustness of result on median house value of children's ZIP code

(a) Sons



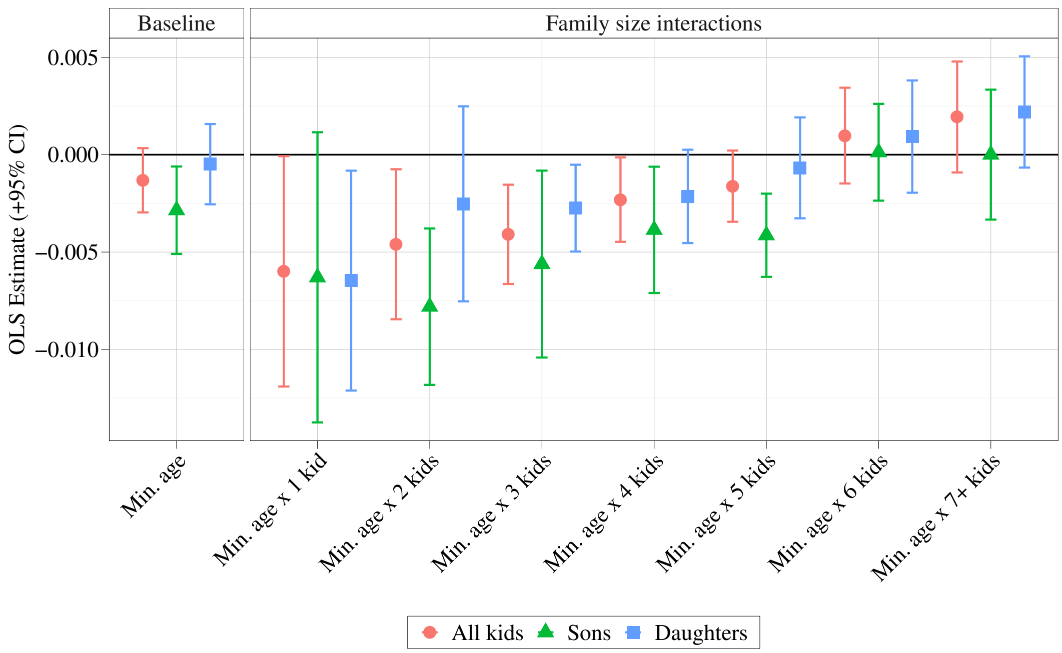
(b) Daughters



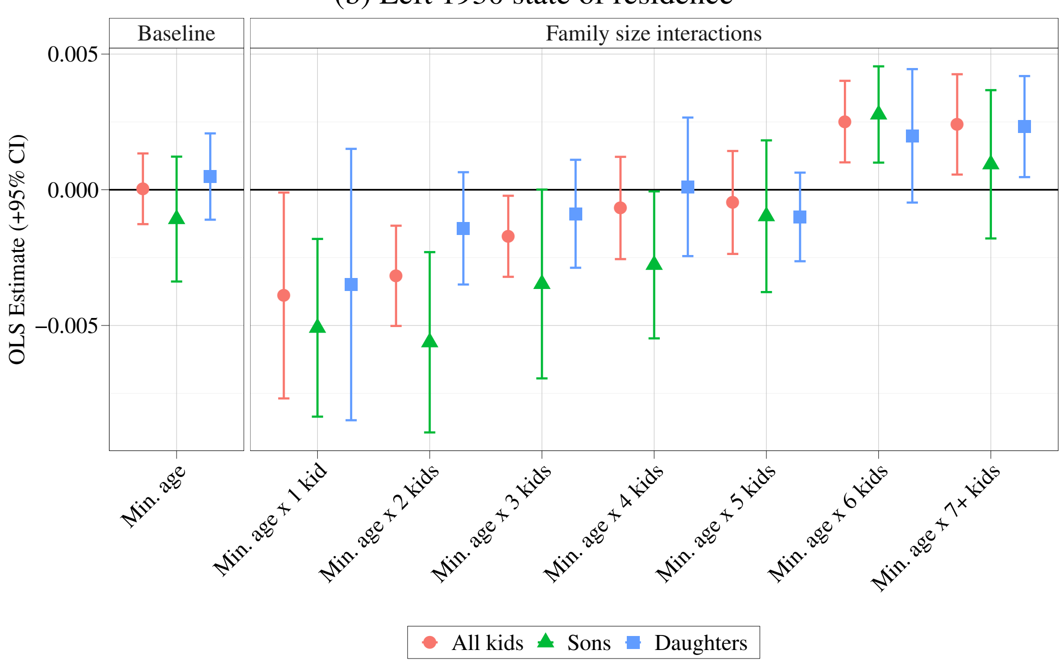
Notes: Figures show coefficients on father's minimum age at Social Security eligibility in estimates of equation (2), where the outcome is the average percentile rank of his sons' or daughters' ZIP code at the time of their death in the distribution of AGI (the unit of observation is the household). Baseline estimates (left) do not interact minimum age with family size; family size interactions (right) interact minimum age with fixed effects for the number of children in the family observed across the 1920 and 1930 Censuses. All regressions include fixed effects for father's occupation $\times$ year of birth, Social Security coverage year, race, and state/country of birth. Observations are weighted using IPWs. 95% confidence intervals are based on standard errors clustered at the level of father's industry of employment.

Figure 8: Effect of later Social Security coverage of father's employment on children's migration

(a) Left 1930 county of residence



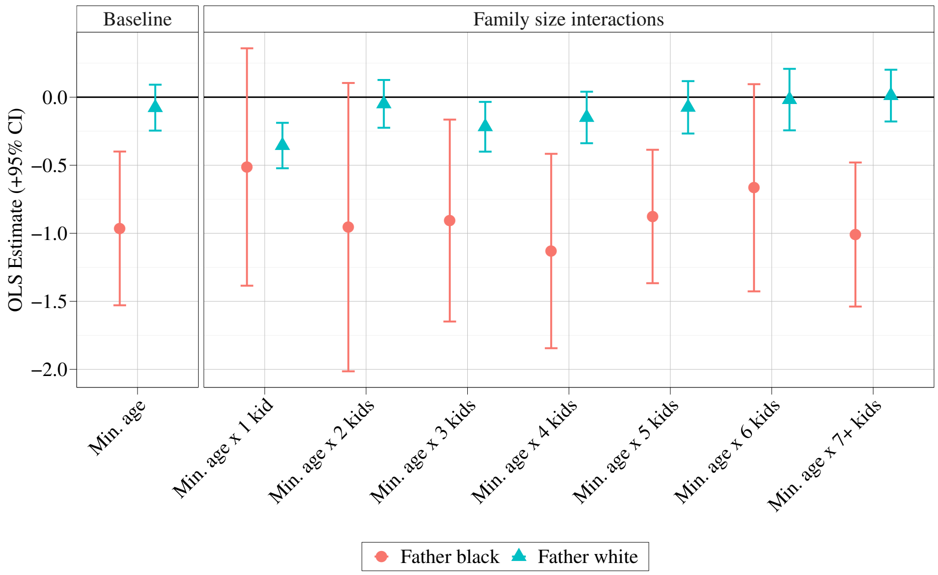
(b) Left 1930 state of residence



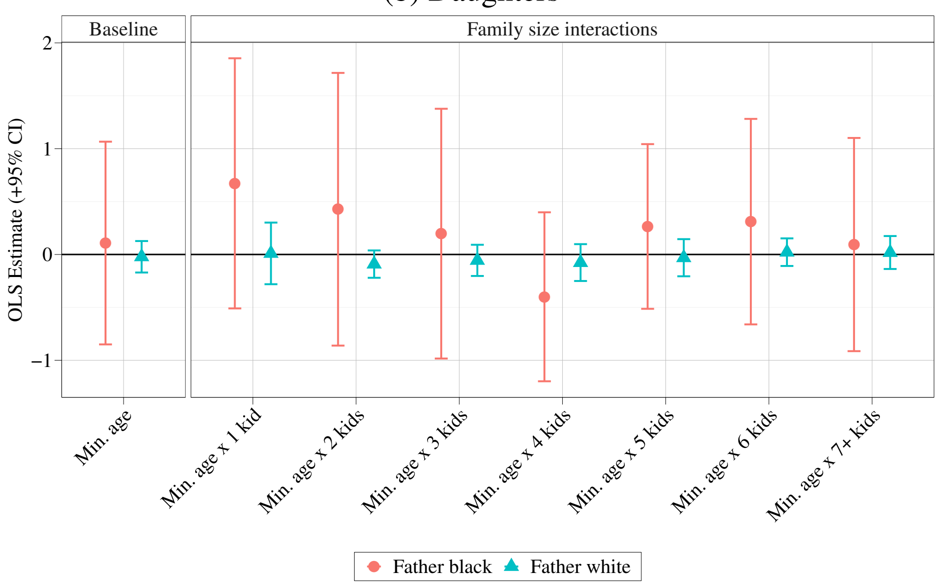
Notes: Figures show coefficients on father's minimum age at Social Security eligibility in estimates of equation (2), where the outcome is the average percentile rank of his sons' or daughters' ZIP code at the time of their death in the distribution of AGI (the unit of observation is the household). Baseline estimates (left) do not interact minimum age with family size; family size interactions (right) interact minimum age with fixed effects for the number of children in the family observed across the 1920 and 1930 Censuses. All regressions include fixed effects for father's occupation $\times$ year of birth, Social Security coverage year, race, and state/country of birth. Observations are weighted using IPWs. 95% confidence intervals are based on standard errors clustered at the level of father's industry of employment.

Figure 9: Heterogeneity of impact on AGI of child's ZIP code by race

(a) Sons



(b) Daughters



Notes: Figures show coefficients on father's minimum age at Social Security eligibility in estimates of equation (2), where the outcome is the average percentile rank of his sons' or daughters' ZIP code at the time of their death in the distribution of AGI (the unit of observation is the household). Baseline estimates (left) do not interact minimum age with family size; family size interactions (right) interact minimum age with fixed effects for the number of children in the family observed across the 1920 and 1930 Censuses. All regressions include fixed effects for father's occupation $\times$ year of birth, Social Security coverage year, race, and state/country of birth. Observations are weighted using IPWs. 95% confidence intervals are based on standard errors clustered at the level of father's industry of employment.

Table 1: Representativeness of linked sample

	Mean			Difference (two-sided t-test)	
	Population (20%)	Linked sample (unweighted)	Linked sample (weighted)	Linked sample (unweighted)	Linked sample (weighted)
	(1)	(2)	(3)	(4)	(5)
Panel A: Father characteristics in 1930					
Black	0.082	0.036	0.081	-0.046	-0.001
				[0]	[0.091]
Born in 1875-1879	0.297	0.264	0.298	-0.033	0.001
				[0]	[0.475]
Born in 1880-1884	0.377	0.385	0.377	0.008	0
				[0]	[0.766]
Born in 1885-1888	0.326	0.35	0.325	0.025	0
				[0]	[0.697]
Born abroad	0.311	0.306	0.312	-0.005	0.001
				[0]	[0.443]
Born in Northeast	0.193	0.204	0.193	0.011	0
				[0]	[0.943]
Born in Midwest	0.244	0.264	0.244	0.02	0
				[0]	[0.939]
Born in South	0.228	0.203	0.227	-0.025	-0.001
				[0]	[0.403]
Born in West	0.024	0.022	0.023	-0.001	0
				[0]	[0.662]
Literate	0.934	0.948	0.935	0.014	0.001
				[0]	[0.011]
Covered by SS in 1935	0.875	0.88	0.874	0.005	-0.001
				[0]	[0.035]
Covered by SS in 1950	0.125	0.12	0.126	-0.005	0.001
				[0]	[0.035]
White-collar worker	0.225	0.221	0.21	-0.004	-0.015
				[0]	[0]
Farmer	0.004	0.004	0.004	0	0
				[0]	[0.251]
Skilled/semi-skilled worker	0.472	0.507	0.492	0.035	0.02
				[0]	[0]
Unskilled worker	0.299	0.268	0.294	-0.031	-0.006
				[0]	[0]
(table continues on next page)					
N	602,163	650,848			

Notes: Columns (3) and (5) weighted using IPWs. p-values for two-sided t-test of equality of means in brackets.

Table 1 (cont.): Representativeness of linked sample

	Mean			Difference (two-sided t-test)	
	Population (20%) (1)	Linked sample (unweighted) (2)	Linked sample (weighted) (3)	Linked sample (unweighted) (4)	Linked sample (weighted) (5)
Panel B: Household characteristics in 1930					
Live in Northeast	0.36	0.374	0.371	0.013	0.011
				[0]	[0]
Live in Midwest	0.312	0.336	0.32	0.024	0.008
				[0]	[0]
Live in South	0.214	0.197	0.214	-0.017	0.001
				[0]	[0.376]
Live in West	0.114	0.094	0.094	-0.02	-0.019
				[0]	[0]
Live in urban area	0.706	0.675	0.705	-0.031	-0.001
				[0]	[0.226]
Live on farm	0.065	0.076	0.065	0.012	0
				[0]	[0.752]
Homeowner	0.478	0.538	0.481	0.06	0.003
				[0]	[0.001]
Own a radio	0.438	0.469	0.44	0.03	0.002
				[0]	[0.088]
Number of sons	1.154	2.029	2.043	0.875	0.889
				[0]	[0]
Number of daughters	1.07	1.887	1.912	0.817	0.841
				[0]	[0]
N	602,163	650,848			

Notes: Columns (3) and (5) weighted using IPWs. p-values for two-sided t-test of equality of means in brackets.

Table 2: "First stage": variation is predictive of Social Security claiming

	Dependent variable:	
	1[in linked sample]	Age at claim
	(1)	(2)
Minimum age of SS eligibility	-0.0033***	0.1163***
	(0.0004)	(0.0237)
Sample	Population	Linked sample
Weighting	Unweighted	Weighted
Mean of dep. var.	0.1	67.79
R^2	0.0234	0.1682
N	3,796,856	364,278

Notes: Unit of observation is a father. All regressions include fixed effects for occupation × year of birth, Social Security coverage year, race, and state/country of birth. Observations weighted using IPWs in column (2). Robust standard errors in parentheses, clustered at the industry level. *** 1%, ** 5%, * 10% significance.

Table 3: Test for sample selection

	Dependent variable: 1[in linked sample]
Minimum age of SS eligibility	-0.0006
	(0.0006)
R^2	0.0188
N	2,971,208

Notes: Unit of observation is a household. All regressions include fixed effects for occupation × year of birth, Social Security coverage year, race, and state/country of birth. Robust standard errors in parentheses, clustered at the industry level. *** 1%, ** 5%, * 10% significance.

Table 4: Placebo tests in 1930 Census

	Dependent variable: Household characteristic in 1930								
	Homeowner	House value (asinh)	Own a radio	Live in urban area	Live on farm	Number of children	Dad literate	Mom literate	Mom in labor force
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: Population (unweighted)									
Minimum age of SS eligibility	-0.0005	-0.0050	0.0003	-0.0003	-0.0002	-0.0038	-0.0001	-0.0000	0.0003
	(0.0012)	(0.0112)	(0.0015)	(0.0009)	(0.0002)	(0.0042)	(0.0006)	(0.0008)	(0.0009)
Mean of dep. var.	0.48	3.96	0.44	0.71	0.06	2.22	0.93	0.93	0.11
R^2	0.0640	0.0884	0.2154	0.2562	0.2332	0.1078	0.1495	0.1908	0.0625
N	2,964,018	2,899,011	2,971,208	2,971,208	2,971,208	2,971,208	2,971,208	2,708,885	2,709,477
Panel B: All families in linked sample (weighted)									
Minimum age of SS eligibility	-0.0009	-0.0011	-0.0003	0.0003	-0.0005***	-0.0007	0.0001	0.0003	0.0003
	(0.0011)	(0.0110)	(0.0012)	(0.0006)	(0.0002)	(0.0031)	(0.0005)	(0.0006)	(0.0013)
Mean of dep. var.	0.48	4	0.44	0.7	0.06	3.95	0.93	0.93	0.09
R^2	0.0752	0.1045	0.2146	0.2602	0.2194	0.1292	0.1426	0.1822	0.0620
N	650,247	638,367	650,848	650,848	650,848	650,848	650,848	637,236	637,328

Notes: Unit of observation is a household. All regressions include fixed effects for occupation $\times$ year of birth, Social Security coverage year, race, and state/country of birth. Observations weighted using IPWs in Panel B. Robust standard errors in parentheses, clustered at the industry level. *** 1%, ** 5%, * 10% significance.