Long-Run Intergenerational Effects of Social Security*

Daniel K. Fetter
Dartmouth College
and NBER

Lee M. Lockwood
University of Virginia
and NBER

Paul Mohnen
Federal Reserve Bank of Atlanta

November 18, 2024

Abstract

Both historically and today, much of the support of the elderly by their adult children has taken the form of in-kind transfers that require shared location, such as shared housing. To the extent that Social Security substitutes for these forms of support, it might thereby relax a "location constraint," with potentially wide-ranging intergenerational effects. Motivated by this fact, we investigate intergenerational effects of Social Security by combining a novel empirical approach—exploiting within-occupation, cross-industry differences in Social Security coverage in the early years of the program—with a dataset linking information on parents to the long-run outcomes of their children. We find that individuals whose parents were more likely to have received Social Security, or received it earlier, lived farther from their childhood location in adulthood, earned more, and lived in ZIP codes of higher socioeconomic status near the end of their lives. A variety of evidence suggests that migration to better-matched labor markets was a key driver of these gains. The magnitude of the estimates suggests that the impact of government old-age support programs on the labor market outcomes of recipients' children may be central to their overall welfare effects.

*For helpful comments and discussions, we thank Ran Abramitzky, Amy Finkelstein, Sasha Indarte, Hilary Hoynes, Olivia Mitchell, Matt Notowidigdo, Jim Poterba, Kent Smetters, Heidi Williams, as well as seminar/conference participants at the ASSA, Atlanta Fed, Berkeley, Boston College Center for Retirement Research, Boston Fed, Chicago Fed, Clemson, Dartmouth, Emory, Harvard, Hitotsubashi, IPUMS, KU Leuven, Lehigh, Michigan Retirement and Disability Research Center, NBER Trans-Atlantic Public Economics Seminar, NBER Economics of Mobility, NBER Summer Institute (DAE, ESS), Philadelphia Fed, Princeton, UC Davis, UCLA, UIUC, University of Nebraska-Lincoln, Rutgers, Stanford, SEA, SOLE, and Wharton BEPP. We thank David Lee, Tamri Matiashvili, Elisa Heinrich Mora, Maxwell Rong, Amanda Wahlers, and Qiyi Zhao for excellent research assistance. Fetter and Lockwood gratefully acknowledge financial support from National Science Foundation (NSF) Grant SES-1628860. Fetter gratefully acknowledges financial support from the Stanford Institute for Economic Policy Research. Lockwood gratefully acknowledges financial support from the Bankard Fund for Political Economy. 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. The views expressed herein are those of the authors and not necessarily those of the Federal Reserve Bank of Atlanta or the Federal Reserve System.

1 Introduction

There is considerable evidence that government old-age support programs, like Social Security, affect intergenerational transfers between adult children and their elderly parents (e.g., Costa, 1997, 1999; McGarry and Schoeni, 2000; Engelhardt et al., 2005). That such changes in family transfers can fundamentally transform the effects of such programs is well known (e.g., Barro, 1974; Becker, 1974; Bernheim and Bagwell, 1988). Yet leading models, which have guided subsequent empirical work, focus on a special case in which family support is, at least implicitly, in the form of frictionless cash transfers. Both historically and today, however, a large share of family support has been in kind, such as shared housing, assistance with household chores, and personal care. While such support has important advantages, a potential cost is that it often requires parents and their adult children to live close to one another. To the extent that Social Security reduces the need for such transfers, it might help children of recipients by effectively relaxing a constraint that they live near their parents, which could be especially helpful if their parents' location is a poor match for their skills.

Motivated by these possible effects of Social Security, we provide new evidence on the intergenerational effects of government old-age support by studying the long-run effects of Social Security on recipients' children. 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 government old-age payments—primarily through Social Security—increased over the second half of the 20th century, the rate of intergenerational coresidence, a common measure of intergenerational connections that encompasses transfers of various kinds (e.g., Costa, 1997, 1998, 1999; McGarry and Schoeni, 2000; Ruggles, 2007), declined substantially.

Naturally, this time-series pattern does not establish a causal relationship between government-provided old-age support and family connections. 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 employment in some industries was covered when the program was introduced in 1935, while similar employment in other industries was covered only in 1950 or later as the program was expanded. For example, work as a janitor in a private, for-profit firm was covered in 1935, while work as a janitor in a nonprofit hospital or educational institution was not covered until 1950. As a result, due to their industry of employment, 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. Moreover, differences in coverage year had differential "bite" across cohorts since eligibility was conditional on being 65 or older; a worker staying in 1950-covered employment and who turned 65 in the early 1950s, for example, was much less affected than one who turned 65 ten years earlier. We leverage this variation to compare the long-run outcomes of otherwise-similar individuals whose parents had different levels of coverage from Social Security.

Implementing this approach requires observing children's outcomes in adulthood as well as having sufficient retrospective information on their non-co-resident or deceased parents to predict their likely eligibility for Social Security. To overcome this challenge, we build a new dataset linking Social Security death records to the 1930 and 1950 US Censuses. The 1930 Census allows us to predict the age at which fathers 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 the ZIP code in which their children last lived, which we show is correlated with lifetime earnings using restricted data from the Health and Retirement Study and therefore serves as a proxy measure for their end-of-life socioeconomic status (SES). The 1950 Census allows us to study children's migration patterns by observing their location 10 years after Social Security first started paying monthly benefits, when the children in our sample were in their early- to mid-adulthood. We supplement this dataset with the 1973 Survey of Occupational Changes in a Generation (OCG), which contains sufficient retrospective information on respondents' fathers to determine their likely Social Security eligibility age and also contains direct measures of respondents' income and occupational standing in adulthood.

Conceptually, we emphasize that an important omission from standard models of the intergenerational consequences of government old-age support is that family support is often non-monetary. Figure 1 shows that co-residence was historically quite common, and both historical and modern data show that various types of in-kind care, such as performing errands, preparing food, or providing personal care, are also common (e.g., Sussman, 1965; Horowitz, 1985; Folbre, ed, 2012). An important characteristic of these kinds of connections is that they require spatial proximity. Hence, to the extent that Social Security displaces these connections, it may affect where people live, and hence their labor market opportunities. Also important is that both contemporary and historical evidence reveals differences between sons and daughters in the amount and type of care provided (e.g., Horowitz, 1985; Folbre, ed, 2012; Grigoryeva, 2017). Daughters more often provide types of care that, especially historically, have been more difficult to purchase on the market, or that families may prefer not to purchase on the market, such as caregiving. Hence, it is likely that sons and daughters would be affected differently by their parents' receipt of government old-age support.

Our main empirical finding is that children whose parents were more likely to ever be eligible for Social Security benefits, or were likely eligible at a younger age, lived in higher socioeconomic status ZIP codes near the end of their lives (around the late 20th century or beginning of the 21st century), as measured by income, house prices, and a neighborhood quality index that captures a broader set of socioeconomic indicators (e.g., college attainment rates, poverty rates, etc.). These effects are concentrated among sons rather than daughters, consistent with the prediction that daughters might be less affected because of differences in the types of care they typically provided. Direct evidence on income and occupational standing from the OCG supports the interpretation that residence in higher-income ZIP codes reflects higher lifetime income for sons.

The size of the estimates suggests that Social Security conferred large, long-lasting financial gains on recipients' children. Under our baseline set of assumptions, we estimate that a ten-year difference in fathers' predicted earliest age of Social Security eligibility—roughly the difference between 1935- and 1950-covered fathers from the 1877 birth cohort—conferred a financial gain on recipients' sons of $740 per year or $15,010 in expected present value (amounts in 2020 dollars). Given the long-lasting nature of these gains and the fact that the average recipient had 2.3 sons, a comparison of these gains to an estimate of the corresponding difference in parents' net transfers from Social Security suggests that the total financial gain to recipients' children may have been comparable to or even exceeded the direct financial gain to recipients themselves.

A variety of evidence suggests that a key driver of the financial gains to recipients' children was that Social Security led some to migrate to better-matched labor markets, and that this may have been driven by changes in in-kind transfers between children and their parents. First, we find that sons whose parents were more likely to receive Social Security were more likely to have migrated from their parents' 1930 state or census division of residence by 1950. In contrast, there is no such effect for daughters. The size of this effect is large enough that it could plausibly account for a large share of the total gains. Second, the labor market gains and the differential effects on sons versus daughters are also consistent with migration and the displacement of in-kind family transfers that were typically provided by sons being a key driver. In contrast, alternative potential drivers, such as greater net downstream monetary transfers, are unlikely to account for more than a small share of the total gains and typically would not be expected to lead to labor market gains or highly differential effects on sons versus daughters.

Our results suggest that the effects of Social Security on recipients' children may be central to its overall distributional and welfare effects. While much research has documented Social Security's large statutory redistribution from younger to older cohorts (see, e.g., Moffitt, 1984; Leimer, 1994; Feldstein and Liebman, 2002), our results suggest that the economic incidence of Social Security was much less concentrated on older cohorts. Indeed, our results suggest that even Social Security's large net transfers to early recipients may have conferred larger financial gains on recipients' children than on recipients themselves. While much of the literature on Social Security and related programs has focused on the effects on recipients' labor supply and savings (e.g., Coile, 2015; Attanasio and Brugiavini, 2003), our results suggest that the effects on recipients' children are another important determinant of the overall welfare effects.

As noted earlier, the theoretical literature on Social Security has recognized the importance of potential changes in family relationships for the effects of Social Security (Barro, 1974; Becker, 1974; Bernheim and Bagwell, 1988), though the focus has been, at least implicitly, on the special case of frictionless monetary transfers. In parallel, an empirical literature has emphasized that Social Security and other old-age transfer programs displaced in-kind transfers within families in the form of shared housing (Costa, 1997, 1999; McGarry and Schoeni, 2000; Engelhardt et al., 2005), and, in more recent periods, that Social Security increased downstream monetary transfers and reduced upstream non-monetary transfers (Mukherjee, 2022). However, due to data limitations, little is known about how these changes have affected recipients' children in the long run, which is a key input into the overall welfare effects of Social Security. Thanks to newly available data and our ability to follow the same individuals over long periods of time, our work builds on these findings to examine long-term effects on recipients' children. Our finding of large net financial gains to recipients' children, including through greater migration, is also consistent with evidence on the impact of the introduction of pensions in modern-day developing countries.[1]

Taken together, our results suggest that to an important degree, the introduction and expansion of Social Security was not so much a large transfer from one generation to another as a replacement of one type of old-age support by another. This view is in accordance with that of work by historians such as Gratton (1996), who alludes to the significant costs of a family-based system and notes that in public opinion surveys, Social Security was as popular with the young as it was with the old.

2 Background

2.1 Eligibility and benefits in the early Social Security Program

The Social Security Act was passed in 1935, but it was under the 1939 Amendments to the Social Security Act that the first monthly retirement and survivors' benefits under Old Age and Survivors Insurance (henceforth, "Social Security") were paid, in 1940. Under the 1939 and subsequent amendments, a worker became 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.[2] For example, under the 1939 Amendments, a worker would be eligible for retirement benefits 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.

[1] Similar themes and findings have been documented in the context of South Africa (Jensen, 2003; Edmonds et al., 2005; Ardington et al., 2009; Hamoudi and Thomas, 2014), China (Chen et al., 2018; Eggleston et al., 2018; Huang and Zhang, 2021), and Ghana and Indonesia (Bau, 2021).

[2] The original act in 1935 specified a different criterion for coverage, which required both sufficient earnings in covered employment and at least one day of covered employment in each of five years after 1936 and prior to reaching age 65.

The requirement for a sufficient length of employment meant that relatively few of the birth cohorts who were already advanced in age in 1940 were eligible for monthly Social Security payments. The major source of expansion of the program in subsequent decades was the aging-in of cohorts who had worked a sufficient length of time to be eligible for benefits. The growth of the program as it matured is shown in Figure 2, which shows the share of men receiving retirement benefits over time, separately by 5-year birth cohorts. Only about 5 percent of the 1871–75 birth cohort received monthly payments in 1940 at ages 65–69, rising to about 42 percent of those surviving to 1955 (ages 80–84). The 1896–1900 birth cohorts, born 25 years later, received Social Security at far higher rates: nearly 70 percent of men collected benefits in 1965, at ages 65–69, and 96 percent of those surviving to 1980 collected them then, at ages 80–84.

It will be central to our empirical analysis that there was also growth in eligibility due to expansion of 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.[3] For types of employment covered in the original Act, coverage began on January 1, 1937.[4]

Amendments to the Social Security Act, particularly between 1950 and 1956, significantly expanded the definition of covered employment. In our analysis, our focus will be on wage and salary workers rather than the self-employed, so we first discuss those extensions. The first extension of coverage to additional categories of wage workers was under the 1950 Amendments. Under these amendments, coverage was made compulsory for regularly employed agricultural wage workers and domestic workers. The amendments also extended coverage on an elective basis to most employees of nonprofit organizations and to 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. Amendments in 1954 extended coverage to state and local employees already under a retirement system, still on an elective basis, and amendments in 1956 further liberalized the pro-

[3]See Appendix Section A.2.1 for a more detailed description of the initial exclusions and subsequent expansion. Nonprofit organizations were excluded due to concerns that it would jeopardize their tax-exempt status (McCamman, 1948), and state and local government employees were excluded because of perceived constitutional limitations on federal taxation of state and local governments (Clark et al., 2011). 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.

[4]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.

visions under which state and local government employees could be covered. Although coverage was on an elective basis for nonprofits, take-up was fairly rapid: as of 1951, 56% of employees of nonprofit organizations were covered under Social Security, rising to 76% by 1960 (Subcommittee on Social Security, 1976, p. 27). For state and local governments, take-up was slightly slower than for nonprofits, but still rose to high levels: about 11 percent of all state and local government employees were covered by Social Security at the beginning of 1951 (the first year in which they could be), and by 1958 the majority of state and local government employees (54 percent) were covered by Social Security (Kestenbaum, 1982).

Coverage of self-employment was also important to the growth of Social Security, although it does not play a role in our empirical analysis. The 1950 Amendments extended coverage on a compulsory basis to the self-employed, except for certain professional groups and farmers. Self-employed farmers and farm managers were largely brought into the system in 1954, with a further expansion in 1956. Self-employed professionals who had not been covered in 1950 were mostly covered in either 1954 or 1956.

Under the 1939 Amendments, a worker who had worked a sufficient number of quarters in covered employment 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 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 25 to 30 percent through the 1950s and most of the 1960s; for a low earner replacement rates were somewhat higher, around 25 to 30 percent in the 1940s and 40 to 50 percent in the 1950s and 1960s (SSA Office of the Chief Actuary, 2013, Supplementary Table V.C7). Tax rates for both the employer and employee were 1 percent beginning in 1937, the first year any employment was covered; these rose to 1.5 percent in 1950 and to 2 percent in 1954, and gradually rose in later years.

As has been frequently noted (e.g., Moffitt, 1984; Schieber and Shoven, 1999; Gruber, 2013), the initial generation of beneficiaries paid little in taxes, received large benefits, and hence received large net transfers from Social Security. Famously, the first recipient of monthly benefits, Ida May Fuller, paid $24.75 in taxes and, by her death at age 100, had received $22,888.92 in benefits (all amounts nominal). As a more typical case, 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 until he was 65 in employment covered under the 1935 act, 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.[5]

2.2 Transfers and shared resources between the elderly and their adult children

For interpreting our empirical results, it is useful to discuss the typical nature of support between the elderly and their adult children prior to and in the early years of Social Security. We focus on support that children provided to their parents ("upstream" transfers), but discuss other types of relationships at the end of this section.

One of the major forms of support between adult children and their parents in the first half of the 20th century was co-residence. Figure 1 shows that in 1930, roughly 27 percent of women 65 and older, and about 12 percent of men 65 and older, lived with their children and were not the household head; co-residence rates of the elderly fell significantly in the subsequent decades.[6] Because we are interested in effects of Social Security on adult children, it is also useful to consider the share of adult children co-residing with parents, and hence what an adult's expectations might have been regarding their future living arrangements. Panel (a) of Figure 3 shows the share of native-born adults who were either household head or the spouse of the household head and who had one or more parents or parents-in-law in the household, tracing individual 10-year birth cohorts over their life cycles. For each birth cohort co-residence rises between ages 25–34 and 35–44 and then declines, presumably due to parents' mortality. For all cohorts shown here, this form of co-residence peaked at ages 35 to 44. For cohorts that reached their mid-20s by 1940, about 7 percent had a parent or parent-in-law in their household at ages 35–44, serving as a lower bound on the share of children who ever lived as household heads with their parents in the household. Note that since all children within a family are included in the base, with an average of between 4 and 5 children per family, the share of families in which children ever lived with their elderly parents was significantly higher. Consistent with evidence that Social Security led to reductions in co-residence of the elderly (e.g., McGarry and Schoeni, 2000), co-residence patterns changed significantly for cohorts reaching age 25 in 1950 or later, who lived with their parents at much lower rates.

Systematic data on other types of non-monetary support from adult children to their elderly parents, especially prior to Social Security, is more scarce. But evidence from the 1950s, when many elderly cohorts were still not eligible for Social Security, indicates that such support was

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

[6]This series shows a decline prior to 1940, which Costa (1998) attributes in large part to rising income. The decline from 1930 to 1940 is most likely not due to Social Security retirement benefits, for which regular payments did not begin until 1940, but Old Age Assistance played an important role in the 1930–40 decline (Fetter, Lockwood and Mohnen, 2018). We focus on this measure of co-residence rather than a broader one in which parents are named the household head as a way to distinguish it from changes in children's propensity to leave their childhood home in young adulthood.

widespread (Schorr, 1960). Indeed, such support remains widespread even in more recent periods (as described, e.g., by Horowitz, 1985). Several sociological studies of family connections of the elderly were carried out in the 1950s and early 1960s, some studies of a sample of households within a specific city, and some national. Notably, many of the major forms of non-monetary help that children provided their parents other than shelter required both time and spatial proximity. These types of help included physical care, shopping, and performing household tasks (Sussman, 1965). Help during illness, including long-term illness, was particularly common (Streib, 1958; Sharp and Axelrod, 1956).

Monetary transfers from children to their parents were present, but as emphasized by observers at the time (e.g. Schorr, 1960), connections within extended families did not primarily take this form. In a supplement to the 1952 Current Population Survey, Steiner and Dorfman (1957) found that for couples and unmarried men over 65, less than 1 percent received any regular monetary contributions from people outside their household; for unmarried women, less than 2 percent received regular monetary contributions as a major source of income. Irregular monetary contributions—such as paying a doctor or hospital bill in cases of serious illness—were probably more prevalent than regular monetary contributions. In a national survey of the elderly in 1957, Shanas (1961) found that there were contributions from children in about 20 percent of the cases in which the elderly had a recent serious illness. Nevertheless, Steiner and Dorfman (1957) regarded these irregular monetary contributions as being less important than co-residence as a form of support that adult children provided their parents. The reasons that monetary transfers were uncommon in comparison to other forms of help may have included a lack of markets for assistance with household tasks or shopping, or the ability to economize on both living costs and the time costs of physical or other care through co-residence. But it is also worth noting that in addition to their material needs, the elderly had strong desires for love and affection from their children. Respondents in Streib (1958), for example, reported that affection from their children was as or more important than financial help. Shanas (1961) found that in opinion surveys, many older people were concerned that asking for financial help from their children would threaten the affectionate character of their relationship with their children.

As they do now (Horowitz, 1985; Grigoryeva, 2017), daughters and sons differed in their connections with aging parents. For co-residence, asymmetry between daughters and sons widened in the early years of the Social Security era, then began to converge again. Panel (b) of Figure 3 shows time trends in the share of US-born men and women aged 25–54 living with their own parents. At the beginning of the 20th century, co-residence rates of older parents with daughters and sons were not substantially different, and in fact in 1900, co-residence with sons was slightly more common than with daughters. Over the middle decades of the 20th century, however, the share of sons co-residing with parents fell significantly faster than the corresponding rates for daughters; by 1960, daughters were much more likely to live with their parents than sons were. Interestingly, these rates converged in 1990 before diverging again. There is, unfortunately, little systematic evidence from the early 20th century on children's provision of non-housing services, which are the type of support that survey data have found to be disproportionately provided by daughters in more recent periods (Horowitz, 1985). But in the 1957 National Survey of the Aged, Shanas (1961) found that approximately 60 percent of people 65 and older named a son or daughter as the person they would ask for help during an extended illness, and that they were about twice as likely to name a daughter as to name a son. Given the similarity to evidence from recent periods, it seems plausible that the greater reliance on daughters for these types of care was also the case in the 1940s and earlier.

Downstream transfers, and resource-sharing more generally, were also prevalent. There was, for example, a significant degree of co-residence of adult children with older parents as the household heads, and regardless of headship status, co-residence may have been a matter not of transfers in one direction or another but rather of mutual economic benefit (Schorr, 1960). Older parents also provided a substantial amount of non-housing services, such as help caring for young children or help during illness (Schorr, 1960; Sussman, 1965). Finally, there is more evidence for significant downstream financial transfers than for upstream financial transfers; several sources comment that the balance of financial transfers tended to be downstream (Sussman, 1953; Streib, 1958; Schorr, 1960; Sussman, 1965). It should be noted, however, that much of the evidence for significant downstream financial transfers is at relatively young ages, when parents are younger than 65 or not much older, and in the early years of children's marriage.

3 Data

3.1 Linked Numident-census data

Our primary linked sample relies on two data sources: the full count US Censuses from 1920, 1930 and 1950 (Ruggles et al., 2024a), 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, in which we can observe the ZIP code where the individuals last lived, to their fathers in the 1930 Census, which allows us to predict their fathers' future Social Security eligibility based on their age and employment information in 1930, five years prior to the passage of the Social Security Act. Because 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 Numident records: Social Security application (form SS-5) records for 40 million individuals. The majority of SS-5 records are immediately linkable to death records via Social Security Numbers (SSNs), and have the advantage that they additionally list individuals' parents' full names. This allows us to link individuals and their parents as one family unit to households in the 1930 Census. To address the fact that some children may not be living with their parents anymore in 1930, we augment our sample by combining analogous linkages to the 1920 Census and linkages between households in the 1920 and 1930 Censuses. We also link children in SS-5 records to themselves in the 1950 Census in order to observe their location in 1950, ten years after Social Security started paying out monthly benefits, and test our hypothesis that Social Security relaxed location constraints for children of recipients.

We highlight two key advantages of our linked dataset relative to existing longitudinal datasets based on historical records. Typically, these datasets link individuals across records solely based their own information (e.g., name, year of birth, place of birth), and tend to focus on men, since women have historically changed their surname at marriage. Our household linking strategy allows for significantly more disambiguation, increasing the number and quality of links we are able to make. In addition, our dataset covers sons and daughters at similar rates. This is because men and women are present in SS-5 records at similar rates, and with information on parents' names, linking individuals to their childhood household is gender-neutral.[7] Moreover, because we can observe both married and birth names for women, we are able to find them in 1950, regardless of their marital status at the time.[8]

Appendix Figure A.1 provides an overview of how we construct our linked sample. We start by linking siblings in SS-5 records together based on parent names. We then link these reconstituted families to households in the 1920 and 1930 Censuses based on parent names and information on children (first and middle name, age, and place of birth). Next, 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 link children in SS-5 records to themselves in the 1950 Census based on age, place of birth, and name information (leveraging both married and birth names for women).

Each of the linkages described above involves linking records primarily based on names. In this paper, we adopt the supervised machine learning (ML) approach to historical record linking from the Longitudinal, Intergenerational Family Electronic Micro-Database (LIFE-M) Project (Bailey et al., 2023).[9] Like other supervised ML approaches, this method involves making manual linking

[7] Although women participated in the labor market at lower rates than men historically, SS-5 records contain the spouses of primary earners as spouses were eligible for benefits after the 1939 Amendment to the Social Security Act.

[8] We always observe women's birth names via their father's last name. For women who were married when they first filed form SS-5, their last name is typically their married name. Women who subsequently got married typically filed a second form SS-5 to update their last name information for administrative purposes.

[9] Other supervised ML approaches to historical record linking include the IPUMS Linked Representative Samples (Ruggles, 2006), Feigenbaum (2016), and the IPUMS Multigenerational Longitudinal Panel (Helgertz et al., 2024). A common alternative approach, though not the one we take here, is to use automated linking algorithms (Ferrie, 1996; Abramitzky et al., 2020, 2021).

decisions for a random sample of records, which are then used to train an algorithm to make "similar" linking decisions for millions of records. A key feature of the LIFE-M training process is that human trainers examine an entire set of potential candidates ("potentials") for each record we want to link ("primary"), and decide whether or not to make a link and which link to make. Human input is useful as there are often multiple plausible candidates due to name commonality, and deciding which candidate to choose or concluding there is too much ambiguity to do so is not something that can easily be automated using deterministic rules. One valid concern is that the quality of the model links will only be as good as the quality of the human links (Abramitzky et al., 2021). We mitigate this in two ways: (1) by instructing trainers to only make links when they are very confident, and (2) by aggregating decisions from multiple trainers, as links that are independently chosen by different trainers are more likely to be objectively good links.[10]

The resulting hand-linked data is used to discipline a machine learning algorithm to mimic human decisions. We use the two-stage model from Bailey et al. (2023), which is specifically designed for one-to-one linking.[11] The output of the model is an estimated match probability for every primary-potential pair. The model-based links are then determined using a threshold rule: potentials that have a match probability exceeding the threshold are called links. A key decision is how to set this threshold. In general, the lower the threshold the higher match rate but the lower the agreement rate with trainers ("precision rate"). Conversely, the higher the threshold the lower the match rate but the higher the agreement rate, creating a natural trade-off between making more links and fewer false links. Training data allows us to control this trade-off, and we select the threshold associated with a 97 percent precision rate for all our linkages (except when linking children to 1950, where we use a 95 percent precision rate). Using the trained model and the chosen threshold, linking can be scaled to the full sample of records. Details for each specific link, including how we generate sets of potential candidates, what information is shown to trainers, and how well our models approximate trainer decisions, are provided in Appendix A.1.

After combining links between SS-5 records, death records and the 1920 and 1930 Censuses, we restrict the sample to our population of interest: children whose fathers were born between 1875

[10]Following the LIFE-M training process, trainers first complete a training module and are given feedback so that they can familiarize themselves with each specific linkage. Two trainers then make linking decisions for the full random sample of primaries, and their decisions are compared. Disagreement cases are shown to three additional trainers (mixed in with a random sample of other cases). Only links chosen by the two original trainers, or links chosen by four out of five trainers are considered true links for the purpose of training the model.

[11]In the first stage, the probability that there exists a link in the set of potential candidates is modeled using a random forest model (Friedman et al., 2009). The random forest learns to label entire sets as containing a link or not based on set-level features, and is fitted on the training data collapsed at the primary level. In the second stage, the probability that any link in the set is the true link conditional on there being a link in the set is modeled using a logit-style discrete choice model. The explanatory variables in the logit model are pair-level features between each primary and potential, and the associated parameters are estimated via maximum likelihood using the subset of primaries for which a link was found. The stage 1 and stage 2 probabilities are generated using 10-fold cross-validation to minimize overfitting, and multiplied to obtain the unconditional match probability for each primary-potential pair.

and 1888 and covered by Social Security in 1935 or 1950 based their employment information in 1930, excluding self-employed workers.[12] The resulting linked sample constitutes our primary sample and contains around 1.5 million children born between 1895 and 1930, belonging to around 890,000 unique families. Appendix Figure A.2 plots the cohort distribution of fathers and children in the population and in our linked sample. The coverage rate of fathers in our population of interest is 19 percent and the coverage rate of children born by 1930 is 15 percent.[13] We are able to link around 52 percent of sons and 47 percent of daughters in our primary sample to themselves in the 1950 Census, for a final sample size of around 730,000 children belonging to around 550,000 unique families. We will use this secondary sample to study the impact of Social Security on children's location choices in 1950.

Several factors contribute to incomplete coverage. First, only a subset of children appear in the Numident, either because they never applied for a Social Security Number or due to restrictions placed on the public version of these files. The Numident file released by NARA only includes individuals who died prior to 2007 and whose deaths were not state-reported. In practice, this file has near-universal coverage of deaths that occurred between 1988 and 2007, and relatively low coverage of deaths that occurred prior to 1988 (Goldstein and Breen, 2022). Second, ZIP at death information is missing for around 17 percent of death records in our linked sample for our population of interest. Lastly, human trainers are only able to confidently link a subset of records, and in turn the model is only able to replicate a subset of those links at a 3 percent error rate.

Incomplete coverage raises two types of selection concerns: (1) which families appear in our linked sample, and (2) which children appear in our linked sample. Appendix Table A.1 assesses the representativeness of our linked sample in terms of 1930 household characteristics. Overall, our sample is relatively balanced, though a few patterns are worth noting. First, as is common in historical linked samples, Black and low-socioeconomic status families (e.g., as measured by father literacy or home ownership) are underrepresented as they tend to have more common names and report their names less consistently across records. Second, because the population includes men that never had any children and because observing multiple children improves our chances of making a link, our linked sample is naturally skewed towards families with a greater number

[12]We choose the cutoff points of 1875 and 1888 for two main reasons. First, to identify a father-child relationship between two people, we must observe children with their father in the same household in 1920 or 1930. Including men born earlier will increase the number of children who have left the household by 1920, and including men born later will tend to increase the number of children not yet born by 1930. Second, as shall be evident in Section 4, the birth cohorts from 1875 through 1888 cover close to the full range of differences in our key policy variable between 1935- and 1950-covered workers.

[13]We cannot compute the exact coverage rate of children since our population of interest conditions on fathers' employment status in 1930 (and not all children live with their fathers in 1930). Instead, 15 percent is the coverage rate of children whose fathers were born in 1875–1888 regardless of their 1930-based Social Security coverage status, which we compute by pooling the number of children aged under 10 (or 15) across the 1910–1940 Censuses. Our coverage rate is an underestimate due to the fact that some of those children never reached adulthood.

of children in the household in 1930. To account for these differences, we re-weight families in our linked sample using inverse propensity score weights (IPWs) generated using the procedure described in Bailey et al. (2020).[14] By design, the resulting weighted means in column 3 are very close to the population means in column 1.

Intuitively, selection of families into our linked sample is mild because most families in our population of interest had at least one child who ever applied for an SSN and died between 1988 and 2007. However, it is clear that our sample contains the subset of children who satisfy these criteria. In particular, as was evident in Appendix Figure A.2, our linked sample is skewed towards later cohorts of children as they were more likely to survive until 1988. This reflects the selection of cohorts into SS-5/death records, prior to linking (see Appendix Figure A.3). On the other hand, selection in terms of place of birth or place of death is less of a concern given that SS-5/death records have excellent geographic coverage (see Appendix Figure A.4). Selection in terms of which families or which children appear in our linked sample matters only to the extent that it is correlated with the Social Security coverage variation we exploit. In Section 4.4 we show that selection into our sample is not correlated with the Social Security coverage variation we exploit, alleviating the concern that selection would affect the conclusions of our analysis.

The Numident also includes a third set of files: Social Security claims records for 25 million individuals. We build a separate dataset linking men born in 1875–1888 in the claims records to themselves in the 1930 Census. 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 or the age at which they claimed Social Security benefits conditional on appearing in the claims data. These linkages are made on the basis of name, age, place of birth, and place of residence information using the same supervised ML approach described earlier (see Appendix A.1 for more details). Ultimately, we are able to link around 450,000 men covered by Social Security in 1935 or 1950 at a 97 percent precision rate, or around 11 percent of the population. We re-weight this sample using the same procedure used to re-weight our main sample (see Appendix Table A.2 for summary statistics).

3.2 Survey of Occupational Changes in a Generation

While our core analysis is based on the Numident-census linked sample described in the previous section, we also present complementary survey-based evidence. Specifically, we exploit the 1973 wave of the Survey of Occupational Changes in a Generation. The OCG was a supplement to the

[14] The weighting variables are fixed effects for father year of birth, race, place of birth, and literacy status, as well as fixed effects for household home ownership, radio ownership, farm status, and urban status. Other examples of studies using re-weighting techniques to adjust for selection include Pérez (2017) and Zimran (2019).

1973 Current Population Survey (CPS) that focused on civilian males aged 20–65 and collected detailed information on the respondents themselves as well as their parents and spouses. Crucially, it contains retrospective information on the respondent's father's occupation, industry, and self-employment status when they were aged 16, as well as the father's year of birth. Combining this information enables us to predict fathers' likely Social Security eligibility age.[15]

A unique feature of the 1973 CPS is that it was linked to tax records from the Internal Revenue Service (IRS) and earnings histories from the Social Security Administration (SSA) in a joint effort with the U.S. Census Bureau. In Appendix A.1, we describe how we link together the 1973 OCG file (Blau et al., 1994), the 1973 CPS file (United States Bureau of the Census, 1992), the Social Security Records: Exact Match Data file (Social Security Administration, 2005), and the 1937–1975 Social Security Longitudinal Earnings file (Social Security Administration, 2008), despite the absence of unique identifiers.[16] After restricting the sample to sons born in 1907–1930 whose fathers were born in 1875–1894 and covered by Social Security in 1935 or 1950 based on their retrospective employment information (excluding self-employed workers), we are left with 3,020 observations.[17] In addition to containing detailed income information, a key advantage of the OCG sample is that it is nationally representative.

4 Empirical Strategy

4.1 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 (and hence eligible to receive benefits upon reaching age 65), assuming that workers do not change the type of their employment to one that was covered earlier.[18] Although some workers surely changed from 1950-

[15] A first wave of the OCG was carried out in 1962, but this survey did not collect information on fathers' year of birth, which is needed to determine their likely Social Security eligibility age.

[16] In short, we exploit the fact that these files contain common variables that uniquely identify nearly all individuals we are interested in.

[17] 1907 is the earliest birth cohort in the OCG data (aged 65 at the time of the 1973 CPS), while 1930 is chosen as an upper bound for consistency with our Numident-census linked sample and because we are interested in children's outcomes in adulthood. Relative to the Numident-census linked sample, we also expand the set of father cohorts to increase sample sizes.

[18] "Permanently" fully insured means that a worker could stop working in covered employment and still be fully insured—eligible to receive benefits—upon reaching age 65. The distinction arises from the fact that a worker can be

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 4, 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 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. Panel (b) of Figure 4 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.[19]

4.2 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

"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." In our analysis, the distinction is of modest importance because the cohorts we focus on were largely bound by the 6-quarter minimum.

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

a year when this type of employment was covered by Social Security. Full details are provided in Appendix Section A.2.1, and Appendix Table A.3 shows that in the 1940 Census, our coding is highly predictive of workers reporting that they had a Social Security Number. 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

$$y_f = \alpha_{b(f)} + \delta_{g(f)} + \beta \cdot (\text{min. age of Social Security eligibility})_{b(f)g(f)} + \Lambda' \mathbf{x}_f + \varepsilon_f, \qquad (1)$$

regressing the outcome y for family f on fixed effects for father's birth cohort b and father's coverage category g, with β the coefficient of interest and $\mathbf{x}_f$ 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 specification is based on a finer comparison. In particular, the exclusion of certain industries in the original act generates variation that allows us to exclude the self-employed and to make comparisons of wage and salary workers with 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 occupation in other sectors was 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 in 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

$$y_f = \alpha_{b(f)o(f)} + \delta_{g(f)} + \beta \cdot (\text{min. age of Social Security eligibility})_{b(f)g(f)} + \Lambda' \mathbf{x}_f + \varepsilon_f, \qquad (2)$$

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 focus on variation in minimum age of eligibility arising from work in different industries. Equation (2) serves as our preferred empirical specification when analyzing family or parental outcomes.[20] When analyzing the impact of father's Social Security eligibility on children's outcomes, we estimate a variant of equation (2)

$$y_{if} = \theta_{s(i)c(i)} + \alpha_{b(f)o(f)} + \delta_{g(f)} + \beta \cdot (\text{min. age of Social Security eligibility})_{b(f)g(f)} + \Lambda' \mathbf{x}_f + \varepsilon_{if}, \quad (3)$$

where the unit of analysis are individual children i and we additionally include fixed effects for the interaction of children's sex and birth cohort (subscripted by s and c).[21]

As described in detail in Appendix Section A.2.1, we use the industry classification of the Census to identify employees who are less likely than others to have been covered by the original 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, in addition to 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.[22] Note that while most employment in hospitals and educational institutions was in nonprofit organizations or in state and local governments, there were a handful of for-profit organizations in these sectors, as we discuss in Appendix Sections A.2.1 and A.3. Any misclassification of this sort should tend to attenuate our results.

The occupations that form the basis of our comparison are those that were present in both for-profit and either nonprofit or public enterprises, and to a lesser extent private households.[23] Appendix Table A.4 shows the 20 largest occupation categories in the population and our linked samples in which there is non-trivial variation in Social Security coverage status (defined here as the smaller group comprising at least 1 percent of that occupation). Among the largest categories

[20] The three coverage categories are: covered in 1935, covered in 1950 (domestic work), covered in 1950 (nonprofit or state and local government sector). Our baseline controls are fixed effects for father race and state/country of birth. Observations are weighted using the inverse probability weights from Appendix Table A.1.

[21] In these regressions, individuals are weighted by the baseline family-level inverse propensity score weights normalized by the number of children that appear in our linked sample. This avoids mechanically placing more weight on larger families.

[22] As described in Section 2.1, nonprofit employees and state and local workers not covered under an existing retirement system were covered on an elective basis starting in 1950, while state and local workers already covered under a retirement system were not covered by Social Security until 1954. In the Census we do not observe whether or not workers are covered by a retirement system, nor can we generally distinguish between employees of nonprofit organizations and employees of state or local governments. Hence, we treat all such workers as covered in 1950.

[23] Note that many of the wage workers covered in 1950 for whom it is less clear that there would be similar 1935-covered workers do not play a significant role in our comparisons. In particular, most agricultural wage workers and many domestic workers are in their own occupation codes in the census classification. They are included in our sample to help identify covariates, but with no variation in their coverage status, do not contribute significantly to our estimates of interest. We do exploit a handful of occupations that appear both in the private sector and in private households, such as chauffeurs and gardeners.

are carpenters, clerical workers, truck drivers, and janitors.

For interpretation of our estimates, it is useful to note that some workers we classify as not covered by Social Security until at least 1950—state and local government workers in particular—may have been covered by existing retirement systems. We discuss existing retirement systems for government and nonprofit organizations in detail in Appendix Section A.3, and here provide a few notable facts. At the beginning of 1942, when the Census Bureau carried out the first systematic national survey, 46 percent of all state and local government employees were covered by a retirement system (United States Bureau of the Census, 1943, Table 2). However, the types of employees most often covered by these retirement systems, such as police, firefighters, and teachers, are either excluded from our sample or have little effect on the estimates because these occupations do not appear in private industry. Government employees that form the primary basis for our comparisons—such as custodial and maintenance employees—were less likely to be covered under existing retirement systems. As for nonprofit organizations, about a third of private, nonprofit institutions of higher education had retirement plans in the 1940s, but the maintenance and nonprofessional employees who are our focus were generally not eligible (McCamman, 1948). Only about 8 percent of nonprofit hospitals had pension plans at the beginning of the 1940s (Hayhow, 1940). Finally, it is also useful to note that the benefits under retirement systems for both public and nonprofit employment typically differed from Social Security in important ways. Most relevant to our analysis, these plans tended to have either no provision or less generous provisions for widow's benefits (McCamman, 1953). Widows are a particularly relevant case for our study, since co-residence and other types of family old-age support were especially common among widows.

4.3 Predicted Social Security eligibility and subsequent claiming behavior

The discussion above suggests two testable predictions. First, workers whose pre-Social Security employment would be covered in 1935 should be more likely to ever claim Social Security than workers whose employment would only be covered in 1950, and conditional on claiming they should claim at earlier ages. Second, these differences should be smaller for later birth cohorts than earlier birth cohorts, at least over some range. We test these predictions using our sample of Social Security claims records linked to the 1930 Census. For individuals found in the claims data, we can observe the age at which they claimed Social Security benefits. While we cannot distinguish between individuals who never claimed Social Security and individuals who did but were not linked to the claims data (or simply do not appear in it), ever claiming should be positively correlated with the probability of being linked to a claims record.

We begin with a descriptive comparison. First, in the 1930 Census, 1935-covered men born in 1875 are three times more likely than 1950-covered men to be linked to a claims record. This figure drops to one and a half for men born in 1888. Next, focusing on the subset of men linked to a claims record, Figure 5 plots the distribution of age at claim by coverage status, separately for men born in 1875 and 1888. For the 1875 birth cohort, it is evident that 1935-covered workers claimed benefits earlier (3 years on average). In contrast, the age at claim distributions are much more similar for the 1888 birth cohort (average difference of 2 years), which reflects the fact that even workers who never switched out of 1950-covered employment would have satisfied the eligibility requirement by the time they reached age 65 in 1953.

Our parameterization of Social Security eligibility is also predictive of ever claiming Social Security and age at claim. Table 1 reports estimates from equation (2), which exploits variation in eligibility age arising from workers with the same occupation but working in different industries. In column 1, the sample is the 1930 population of men born in 1875–1888 and covered in 1935 or 1950 (excluding the self-employed), and the dependent variable is an indicator for being linked to a claims records. In column 2, we focus on the subset of men linked to a claims record and the dependent variable is the age at which they claimed Social Security benefits. Column 1 shows that eligibility one year earlier is associated with a 0.35 percentage point increase in the likelihood of being linked to a claim for monthly benefits, relative to a baseline mean of 11 percent. Column 2 shows that, conditional on being linked to a claims record, eligibility one year earlier is associated with claiming monthly benefits 0.1 years earlier. Note that we do not formally treat these estimates as a first stage—and in particular do not use them to scale our reduced-form estimates—because neither fully captures the way in which a father's earlier eligibility for Social Security may influence his children's outcomes. Nevertheless, they do establish that there is a strong relationship between our parameterization of Social Security eligibility and subsequent claiming behavior, even when restricting comparisons to fine occupational categories.

4.4 Selection into the linked sample

Due to the coverage of Numident records and the fact that we can only link a selected subset of them to the censuses, the families and children that appear in our linked sample are a selected subset of the population. Even though we re-weight families in our sample to resemble the population in terms of observable characteristics, non-random selection could potentially affect the internal validity of our estimates if it is correlated with the variation we exploit. To assess the extent to which selection is a concern, we examine all households in our population of interest in 1930—those headed by a man born in 1875–1888 and covered by Social Security in 1935 or 1950 based on his employment information (excluding self-employed workers)—and test whether appearing in our linked sample is correlated with our Social Security eligibility variation. The results are reported in Table 2. Reassuringly, we find no evidence that fathers' minimum age of Social Security eligibility is correlated with the likelihood of the household appearing in our linked sample, the likelihood of any son or daughter appearing in our linked sample, or the number of sons or daughters appearing in our linked sample.[24]

4.5 Placebo results in the 1930 Census

A key check on whether our empirical approach allows for valid causal comparisons is to test whether our variation is correlated with household characteristics in 1930, five years prior to the introduction of Social Security. The set of characteristics we examine are home ownership, house value, radio ownership, urban and farm status, number of children present in the household, father and mother literacy status, and mother labor force participation status. Table 3 reports estimates based on two approaches, first testing for differences in the full population of households in panel (a), and then testing for differences within the subset of households contained in our linked sample in panel (b). The results confirm that conditional on making comparisons between households with fathers in the same occupation, there were no significant pre-existing differences between households whose fathers were predicted to be eligible for Social Security earlier versus later. With the exception of homeownership and house value in the population, which are statistically significant at the 10 percent level but small in magnitude, all coefficients are close to zero and statistically insignificant. To further confirm that observable differences across households are not driving our results, we will show that our results are robust to controlling for 1930 household characteristics.

5 Results

5.1 Characteristics of late-life location for sons and daughters

Having established that the variation we exploit is correlated with observed Social Security claiming patterns, but not with pre-existing family characteristics or selection into the linked sample, we now explore how fathers' Social Security eligibility affected the long-run outcomes of their children. We proxy for children's long-run socio-economic status using the "quality" of the ZIP code they last resided in prior to their death. In practice, we measure neighborhood quality using ZIP-level mean adjusted gross income (AGI) in 2001, median house values in 2000, and a neighborhood quality index that captures a broader set of socioeconomic indicators. Partly following Bailey et al. (2024), this index additionally includes the following ZIP characteristics: (1) the homeownership rate in 2000, (2) median rents in 2000, (3) the share of college-educated individuals in 2000, (4) the share of individuals under the poverty line in 1999, (5) the share of households receiving public assistance in 1999, (6) the share of single-parent families in 2000, and (7) a measure of upward

[24] In Appendix Table A.5, we also find no evidence of selection of children from specific cohort ranges, with the exception of daughters born in 1895–1910. As we will show, our main findings are driven by sons so we do not view this as a major concern.

income mobility from Chetty et al. (2018) (the average income percentile rank among children whose parents were at the 25th percentile of the parental income distribution).[25]

Intuitively, children's last known ZIP code of residence should be informative about how well they fared over their lives, as access to good neighborhoods is partly a function of accumulated income. We provide empirical evidence to support this claim in Appendix A.4 where we document a strong positive relationship between lifetime earnings and late-life ZIP of residence quality using data from the Health and Retirement Study (HRS), a nationally representative longitudinal survey of Americans aged 50 and older. Specifically, we leverage restricted-access information on HRS respondents' ZIP code of residence at the time of the surveys as well as information on their lifetime earnings from SSA administrative records. We then show that for respondents aged 65+ at the time of their last interview, a one percentile rank increase in mean AGI in their ZIP code of residence is associated with a 0.35 percentile rank increase in their household lifetime earnings. This result is robust to alternative ways of ranking ZIP codes and holds for both men and women (for more details see Appendix A.4).

Throughout the analysis, we estimate separate models pooling all children or restricting the sample to sons or daughters. As described in Section 2.2, sons and daughters differed in the types of support that they provided their parents. In particular, while sons and daughters co-resided with their parents at broadly similar rates in the early 20th century, co-residence rates with sons declined more rapidly over the course of the 20th century and the evidence suggests that there were substantial differences in their propensity to provide support that was not easily purchased on the market, such as caregiving. This suggests that Social Security might displace transfers from sons to a greater degree than transfers from daughters, and hence that sons may exhibit a larger out-migration response to their parents' receipt of Social Security.

Figure 6 plots the OLS estimates from equation (3), where the dependent variable is the ZIP percentile rank within the national distribution of ZIP codes in terms of mean AGI (top row), median house values (middle row), or the neighborhood quality index (bottom row). As mentioned, we run separate regressions pooling all children, restricting to sons, or restricting to daughters. The negative coefficients in both panels of Figure 6 imply that a *lower* minimum age of Social Security eligibility—in other words, fathers' greater likelihood of eligibility or earlier eligibility—has a *positive* impact on children's end-of-life neighborhood quality. Consistent with the prediction of differential effects for sons and daughters, we find that the effects are entirely concentrated among sons, with little to no effects among daughters. In terms of magnitudes, a ten-year difference in

[25] Data on ZIP-level mean AGI comes from the Statistics of Income Division of the IRS, while data on all other ZIP-level SES indicators (except upward income mobility) comes from the National Historical Geographic Information System (NHGIS, Manson et al. 2024). The neighborhood quality index is constructed in three steps: (1) normalize each sub-index to have a mean of 0 and standard deviation of 1, (2) take the average across all (normalized) sub-indices, (3) normalize the resulting composite index to have a mean of 0 and standard deviation of 1.

fathers' minimum age of Social Security eligibility—roughly the difference between 1935- and 1950-covered fathers from the 1877 birth cohort—is associated with a 1.5 percentile difference in the rank of sons' ZIP at death according to income, house prices, or the neighborhood quality index, which is around 2 to 3 percent of the in-sample means (see Appendix Table A.6). Alternatively, this corresponds to a 2.1 percent increase in ZIP-level income and a 3.5 percent increase in ZIP-level house prices (see Appendix Figure A.6, in which ZIP ranks are replaced with logs of ZIP outcomes).

Appendix Figure A.7 shows results based on ranking ZIP codes according to each of the sub-indices underlying the neighborhood quality index (apart from income and house prices).[26] Across all outcomes we find qualitatively similar patterns, though the coefficients for sons are generally smaller in magnitude and some are only significant at the 10 percent level (poverty rate) or insignificant (home ownership, upward income mobility).

Appendix Figure A.8 shows that the ZIP quality results are robust to controlling for 1930 household characteristics, including 3-digit occupation-by-coverage category fixed effects, including 3-digit industry fixed effects, and including 1930 county fixed effects. In particular, the occupation-by-coverage and industry fixed effects reinforce the notion that our identifying variation comes from differential cohorts trends across workers covered in 1935 vs. 1950 (not industry-based differences across workers covered 1935 vs. 1950 that are common to all cohorts), while the county fixed effects alleviate the concern that our findings might reflect differences in characteristics of 1930 locations correlated with our Social Security eligibility variation (e.g., differential exposure to the Great Depression).

As we have emphasized, a possible channel through which parents' receipt of Social Security may affect their children is the children's migration behavior in adulthood. To the extent that Social Security displaces upstream transfers that require physical proximity, it may induce some children to move to a location more distant from their parents. Figure 7 shows the impact of fathers' Social Security eligibility on children's propensity to live outside their parents' 1930 place of residence (county, state, census division, census region) in 1950. We treat parents' location in 1930, when they are in their 40s and 50s, as a proxy for where they spent their last remaining years (and therefore where children would have had to live in order to take care of them). In particular, leaving parents' state, census division or census region of residence captures the extent to which children migrated to more distant local labor markets, which may have offered them better job opportunities.

The negative coefficients for sons in Figure 7 imply that fathers' earlier eligibility for Social

[26] Note that the signs of the coefficients for the share of individuals under the poverty line, the share of households receiving public assistance, and the share of single-parent families have been inverted so that a negative coefficient is "desirable" for all outcomes.

Security has a positive effect on sons' propensity to live outside their parents' 1930 place of residence by 1950. In terms of magnitude, a ten-year difference in father's earliest age of Social Security eligibility is associated with a 2.6 percentage point greater likelihood of moving to a different state and a 3.1 percentage point greater likelihood of moving to a different census division, which roughly correspond to 20 percent of the baseline means in our linked sample (see Appendix Table A.7).[27] The coefficient for cross-region migration is of similar magnitude but not statistically significant. In contrast, the coefficient for cross-county migration is close to zero and insignificant, which implicitly means that the cross-state and cross-division migration effects are largely offset by a significant decline in within-state cross-county migration. In other words, Social Security appears to have prompted some sons to migrate farther away than they otherwise would have, rather than inducing more sons to migrate (i.e., intensive rather than extensive margin effect). We find no evidence of any significant migration response among daughters, consistent with our hypothesis that caregiving duties prevented them from out-migrating to the same extent as sons.

5.2 Labor market outcomes for sons

The findings in the previous section imply that sons whose fathers were more likely to be eligible for Social Security, or eligible at an earlier age, ended up in higher-SES ZIP codes. To what extent does this reflect real improvements in individual-level outcomes? In this section we provide some suggestive evidence based on the 1973 OCG survey, in which we can observe earnings and occupational standing in adulthood for a small but nationally representative sample of sons.

Figure 8 displays OLS estimates from equation (3), for different measures of occupational standing and income. One minor difference relative to the empirical setup for Numident-based results is that the baseline controls are fixed effects for sons' race and state/country of birth (rather than fathers), and observations are weighted using OCG sampling weights. Note that OCG respondents were asked about their father's employment information when they were 16, which might have endogenously responded to the introduction of Social Security in 1935 for fathers of sons born after 1919. We therefore show separate results for sons born in 1907–1919 to ensure that our findings are not driven by this endogeneity issue.

The top three rows in Figure 8 look at three measures of occupational standing: the occupational income score associated with sons' first job, the job they held in 1962, and their job at the time of the survey in 1973.[28] The bottom two rows look at two measures of income: AGI in 1972, and IRS salary and wage in 1972. All outcomes are in logs. The negative coefficients,

[27] Appendix Figure A.9 shows that our cross-division migration results are also robust to alternative specifications.

[28] Occupational income scores are obtained by mapping OCG occupations into 1950 Census occupations and assigning to each occupation its corresponding median total income in the 1950 Census (Ruggles et al., 2024b). See Appendix Table A.8 for summary statistics.

though imprecisely estimated, broadly suggest that sons whose fathers were more likely to be eligible for Social Security, or were eligible earlier, had better jobs and higher income in adulthood, consistent with the death neighborhood quality gains documented earlier. The effect on the occupational income score of sons' first job is negative but not statistically significant (note that the "first" jobs may predate parents' Social Security receipt for some sons). The coefficients for sons' occupational income scores in 1962 and 1973 are also negative, but slightly larger in magnitude and statistically significant at the 5 or 10 percent level.

The coefficients for AGI and IRS salary are all negative and statistically significant at the 5 percent level except the AGI coefficient for sons born in 1907–1930, which is statistically insignificant. Even though Appendix Figure A.10 shows that the results are not driven by a differential likelihood of having non-missing and positive occupational income scores/income, Appendix Figure A.11 shows results in levels that include zeros. Given that the AGI and IRS salary results are somewhat sensitive to the functional form, whether we include zeros or not, and the cohort range of sons, we put more stock in the more stable occupational income score results when interpreting magnitudes. In particular, a ten-year difference in father's minimum age of Social Security eligibility is associated with a 25 to 30 percent increase in sons' occupational income score during their prime to late adulthood.

6 Implications

6.1 Social Security conferred large financial gains on recipients' children.

Our results suggest that Social Security conferred substantial, long-lasting effects on the children of recipients. The most direct evidence of the financial gains from Social Security to recipients' children is the improvement in the quality of the neighborhoods in which they lived late in life. Neighborhood quality, like other components of consumption, presumably reflects a household's long-term financial position. So the effect of Social Security on the late-life neighborhood quality of recipients' children likely reflects the long-term net financial gains to recipients' children from Social Security: the financial benefit of greater earnings less the financial costs of moving, foregone economies of scale from shared living, foregone childcare from grandparents, and any other financial costs. We estimate that a ten-year difference in the predicted earliest age of father's eligibility—roughly the difference in the earliest age of eligibility between 1935- and 1950-covered fathers from the 1877 birth cohort—is associated with the average son's ZIP at death having 3.5% higher median house prices and 2.1% higher average income. If recipients' children enjoyed the same 2.1% increase in income starting in 1950 (when their fathers were 73 years old), their financial gain would have been equivalent to a $740 per year annuity or roughly $15,010 in expected present value, in 2020 dollars.[29,30]

What about daughters? The vast majority of the children of fathers affected by our instrument are married and so have two sets of parents, one set for each spouse. For a husband-and-wife couple, the effect of both of their fathers having earlier eligibility for Social Security is the sum of our estimates for sons and daughters. Given that the estimates for daughters tend to be small, we focus on the estimates for sons and treat them as applying to married couples as well.[31]

Our findings suggest important modifications to prevailing views on the incidence of Social Security. For example, to ask what the estimated gain of $15,010 in expected present value implies for the intergenerational incidence of Social Security, we can compare this amount to an estimate of the Social Security benefits received by the recipients themselves. Once again, consider father birth cohorts around the 1877 birth cohort. We simulate lifetime Social Security benefits for workers earning the median earnings in covered employment, under the assumptions that workers did not change their employment coverage type, retired according to the profile of labor force participation by age observed in 1930, and died according to the observed mortality-age profile. Under these assumptions, a ten-year difference in earliest eligibility age would lead to a difference in the expected present value of lifetime Social Security benefits to the parents, measured in 1950, of roughly $38,410. Note that the difference in the expected present value to the parents associated with a ten-year difference in the *predicted* earliest age would be smaller, except under the extreme assumption that differences in the predicted earliest age of eligibility translate one-for-one into

[29]These estimates come from discounting a gain of 2.1% of average family income in each calendar year starting in 1950 (when the 1877 cohort reaches age 73) to reflect time discounting (real interest rate of 3% per year) and mortality (based on the cohort life table for males in the 1910 cohort from the SSA).

[30]In general, people living in a location with higher income or higher house prices do not necessarily enjoy higher living standards. In simple spatial equilibrium models without individual heterogeneity, for example, everyone enjoys the same real living standard despite potentially large differences in nominal income and house prices. In reality, such spatial equilibrium forces linking labor and housing markets likely operate mainly at the level of labor markets (e.g., CZs or counties), not within labor markets (e.g., ZIP codes), so within-labor-market differences in income and house prices likely mainly reflect differences in real living standards across different types of households. As discussed in more detail below and consistent with spatial equilibrium forces toward the equalization of real living standards across labor markets, we find little association between Social Security and labor market characteristics, so the strong association between Social Security and ZIP characteristics comes from within-labor-market sorting to better ZIP codes rather than across-labor-market sorting to labor markets with higher income or house prices. The strong association between Social Security and ZIP characteristics thus likely reflects differences in real living standards.

[31]Given that the vast majority of the children of fathers affected by our instrument are married, we interpret our estimates as showing that for husband-and-wife married couples, the Social Security of the husband's father is a more important driver of their migration and living standards than the Social Security of the wife's father. This is in keeping with the idea that sons and daughters-in-law play a different role in family old-age support arrangements than daughters and sons-in-law do, with the role of sons and daughters-in-law being more substitutable with Social Security benefits, perhaps because such couples tend to provide more resource-intensive support whereas daughters and sons-in-law couples provide more time- or care-intensive support. The small effects on daughters of their father's Social Security does not imply that unmarried daughters gain little from their father's Social Security. Unfortunately, it is not straightforward to estimate heterogeneity in the effects of Social Security by marital status because marriage might be affected by father's Social Security.

differences in actual eligibility (i.e., a “first stage” of one, which is a very loose upper bound on the true variation in Social Security coverage associated with our instrument). Given that the average family had more than two sons, the magnitudes of these estimates are suggestive that the total financial gain to recipients’ children from Social Security expansions may have been comparable to or perhaps even greater than the gain to recipients themselves. Hence, contrary to prevailing views, Social Security expansions may not have entailed a large redistribution from younger to older cohorts on net, but rather may have generated considerable financial benefits for the younger generation by improving their labor market outcomes.

To be clear, these calculations incorporate only Social Security benefits and the net financial gain to recipients’ children. They exclude other components of the overall welfare effect of Social Security, including non-pecuniary costs and benefits to recipients’ children, costs and benefits to recipients, and costs and benefits to recipients’ grandchildren and subsequent generations. In terms of recipients, it seems likely that there is important heterogeneity, where some recipients benefit significantly and others are harmed due to losing more from reduced family support than they gain from the net transfers from Social Security.[32] In terms of costs and benefits to recipients’ grandchildren and subsequent generations, our estimates suggest that Social Security led recipients’ children to eventually live in better neighborhoods. This suggests that the intergenerational gains from Social Security may not have stopped at recipients’ children but extended to their grandchildren as well.

6.2 Potential mechanisms and implications

What drove the large gains to recipients’ children? Suggestive evidence highlights the potential role of migration to better-matched labor markets, which may have been associated with the displacement of in-kind family old-age support that required physical proximity.

One piece of evidence is that other, non-migration factors seem unlikely to explain more than a relatively small share of the gains. Arguably the most obvious way in which Social Security would confer gains on recipients’ children—increasing net financial transfers from parents to their children—seems unlikely to match three key patterns of the gains that we estimate. First is the large size of the gains relative to the associated Social Security benefits. Leading theories of family transfers, such as altruism, tend to predict that parents and their children would both benefit from windfall gains, not that children would benefit by much more than the windfall and so, in

[32] Of course, a net windfall from Social Security expands the family’s opportunity set, but nothing guarantees that all family members benefit. This is true even in “unitary” family models if there are realistic constraints (e.g., if family contact and support are more costly when family members live in different locations). For example, a recipient might be harmed if without Social Security they would have lived with their adult child but with Social Security their child moves away and the utility cost of the resulting reduction in family contact and support exceeds the benefit from the Social Security benefits.

a pure cash-transfers model, parents would be harmed. Second is the greater gains of sons than daughters. Downstream financial transfers tend to be shared equally (e.g., Menchik, 1980), so gains from this channel likely would be similar for sons and daughters. Third is the nature of the gains. Our findings suggest that a key proximate factor underlying the gains is that it increased recipients' children's earnings. The estimates of individual-level earnings effects from the OCG are suggestive of important earnings effects, and the finding that that recipients' children lived in ZIP codes with higher income and higher house prices also suggests that they had higher earnings. Our finding that Social Security increased the earnings of recipients' children is suggestive that its key effects on recipients' children came from factors other than an increase in net financial transfers from their parents, since under standard assumptions such an increase would be expected to *decrease* recipients' children's earnings due to income effects of demand for leisure.

Another piece of evidence is that Social Security increased migration by recipients' children to different labor markets, and such migration has been found to lead to large earnings gains in many contexts. For example, we estimate that a ten-year difference in the predicted earliest age of father's Social Security eligibility increased sons' migration to a different census division in 1950 by roughly 3 percentage points (see Figure 7). This migration appears to have been driven not by overall wage levels in a labor market but rather by comparative advantage based on the quality of the match between the household and the labor market. For example, Social Security did not affect the type of location recipients' children moved to in 1950 as measured by urban status or county median income (see Appendix Figure A.12). This is what would be expected in standard spatial equilibrium models (Rosen, 1979; Roback, 1982), and it is consistent with recent evidence on the key role of comparative advantage in driving the gains from migration (e.g., Nakamura et al., 2022).[33]

If the financial gains from Social Security-induced migration were similar to those from many other types of migration that have been studied in the literature, then the financial gains from Social Security-induced migration may have been large both in absolute terms and as a share of the overall financial gain from Social Security to recipients' children. For example, using very different approaches in very different settings, Kennan and Walker (2011) (estimated structural model for young men in the U.S.) and Nakamura et al. (2022) (quasi-experimental evidence from a volcanic eruption in Iceland) both find migration gains on the order of 100% of income. More generally, the broader literature on migration suggests that the financial gains from migration are

[33] Appendix Figure A.13 shows that Social Security caused a reduction in the probability of the joint outcome of staying in parents' 1930 census division and dying in a ZIP code in the first, second or fourth quintile of the ZIP income distribution (though only the impact on the fourth quintile is statistically significant), and a significant increase in the probability of leaving parents' 1930 census division and dying in a ZIP code in the top two income quintiles. This is in keeping with the idea that with little Social Security, some recipients' children stay in their 1930 census division despite being a poor match for that labor market, though it might also reflect that Social Security decreases net upstream transfers of money and time.

often quite large.

As an illustrative benchmark, suppose that Social Security-induced migration to a new labor market conferred a gain of 50% of mean family income in each year from 1950 (son age 40) to 1974 (son age 64) and 25% thereafter. Under this benchmark, if labor markets are defined as census divisions (states), the Social Security-induced increase in migration would account for 74% (62%) of the overall financial gain to recipients' children. Given the high rates of migration during this time period—16% of children left their parents' 1930 census division and 23% left their parents' 1930 state by 1950—even modest increases in the average return to such migration, for example from leading it to occur earlier in the life cycle or over longer distances, could confer large gains. For example, if Social Security increased the average return to migration away from the 1930 census division (state) by 10% of mean family income from age 40–64 and 5% thereafter, that alone would explain 39% (54%) of the overall financial gain to recipients' children. Hence, migration to better-matched labor markets may have accounted for a large share of the substantial financial gains from Social Security to recipients' children.[34]

The patterns of both the inter- and intra-generational effects of Social Security are suggestive that a key driver was the displacement of certain types of in-kind family old-age support that required physical proximity, such as shared housing. In terms of intergenerational incidence, the financial gain from Social Security to recipients' children is, as discussed above, significantly larger than what plausible increases in net downstream financial transfers would produce but is consistent with significant earnings gains from migrating to better-matched labor markets in response to a reduction in demand for in-kind family old-age support. In terms of intra-generational incidence, the larger effects on sons than on daughters is also suggestive of displacement of family old-age support rather than an increase in net downstream financial transfers. Whereas an increase in net downstream financial transfers might be expected to confer a similar benefit on sons and daughters, a Social Security-induced reduction in upstream in-kind transfers from children to their elderly parents would likely benefit sons more than daughters given that sons appear to have been more likely to provide the type of old-age support more readily replaced by Social Security (e.g., financial support in the form of shared housing rather than personal care).[35]

[34]Our findings are suggestive that Social Security, by displacing transfers from children to parents that required spatial proximity, reduced the effective migration costs of recipients' children and thereby may have led to a more efficient matching of recipients' children to labor markets. Social Security may have smoothed the financial component of old-age support across individuals and over time, in a way that may have been hard to achieve otherwise due to incomplete markets. Such a smoothing, to the extent it occurred, would tend to reduce the opportunity cost of migrating away from one's parents by reducing the value of the economies of scale in old-age support from close proximity.

[35]The greater effects of fathers' Social Security on sons than on daughters does not imply that Social Security redistributed from women to men. Many women and men are part of couples, so to the extent that couples share resources, women and men both gain from Social Security transfers to either set of their parents. Our finding that Social Security led sons and their families to move to different census divisions is consistent with the idea that the type of old-age support provided by sons and their spouses (daughters-in-law) is more substitutable with the cash support

Of course, the extent to which our results generalize to other types of government old-age support or other contexts depends on many factors. Perhaps the most fundamental is the extent to which government old-age support displaces family old-age support. Certain types of family old-age support are less prevalent in the U.S. today (e.g., shared housing), but other types remain substantial (e.g., informal care for elderly parents with chronic health conditions; see Barczyk and Kredler, 2018). Moreover, the expansion of markets for certain types of old-age support, such as formal home care, may make it easier for families to substitute away from family-provided care in response to expansions of even monetary government old-age support. These differences might suggest that whereas we find evidence that Social Security reduced family support mainly from sons to their elderly parents (likely often in the form of shared housing), expansions of government old-age support today, especially of in-kind long-term care, might reduce family support mainly from daughters and mainly in the form of informal care.[36]

7 Conclusion

The effects of Social Security on behavior and welfare depend on its effects not only on direct recipients but also on their children. We explore this question by combining a novel empirical approach, based on the fact that Social Security coverage was not universal at first but later expanded, with a newly-built dataset linking parents to the long-run outcomes of their children. We find that sons whose parents were more likely to have received Social Security, or received it earlier, lived farther from their childhood location in adulthood, earned more, and lived in ZIP codes of higher socioeconomic status near the end of their lives. Consistent with both historical and modern survey evidence that daughters tend to provide care that is not typically purchased on the market, we find no evidence of effects among daughters.

A variety of evidence suggests that a key driver of the financial gains for sons was migration to better-matched labor markets, consistent with Social Security displacing old-age support from children to their parents that required physical proximity. The resulting gains are large enough that the total financial gains to recipients' children may even have exceeded the gains to recipients themselves.

that Social Security provides than is the support provided by daughters and their spouses (sons-in-law), which is consistent with daughters providing more time-intensive, care-based support and sons providing more money-intensive support. That we can separately identify the effects of a father's Social Security coverage on his sons versus daughters is an advantage of the fine-grained policy variation we exploit. Common types of more aggregate policy variation, by contrast, would not separately identify the effects of the Social Security coverage of one's father versus one's father-in-law since those would not vary independently of each other.

[36]Other potentially relevant factors include norms about supporting one's parents in their old age, which may have weakened, and the amount of government old-age support, which has increased dramatically. But while these would likely reduce the level of family old-age support, it is not obvious how they would affect its responsiveness to changes in government old-age support.

References

Abramitzky, Ran, Leah Boustan, Katherine Eriksson, James Feigenbaum, and Santiago Pérez, "Automated Linking of Historical Data," *Journal of Economic Literature*, 2021, *59* (3), 865–918.

_ **, Roy Mill, and Santiago Pérez**, "Linking individuals across historical sources: A fully automated approach," *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 2020, *53* (2).

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

_ , "Social Security for State and Local Government Employees," *Social Security Bulletin*, 1945, *8* (4), 4–7.

American Hospital Association, "For Flexibility: Changes in Pension Plan By-Laws," *Hospitals: the Journal of the American Hospital Association*, 1949, *23* (2), 61.

Ardington, Cally, Anne Case, and Victoria Hosegood, "Labor Supply Responses to Large Social Transfers: Longitudinal Evidence from South Africa," *American Economic Journal: Applied Economics*, 2009, *1* (1), 22–48.

Attanasio, Orazio P. and Agar Brugiavini, "Social Security and Households' Saving," *The Quarterly Journal of Economics*, 2003, *118* (3), 1075–1119.

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.

_ **, Hilary W. Hoynes, Maya Rossin-Slater, and Reed Walker**, "Is the Social Safety Net a Long-Term Investment? Large-Scale Evidence from the Food Stamps Program," *Review of Economic Studies*, 2024, *91* (3), 1291–1330.

_ **, Peter Z. Lin, A. R. Shariq Mohammed, Paul Mohnen, Jared Murray, Mengying Zhang, and Alexa Prettyman**, "The creation of LIFE-M: The Longitudinal, Intergenerational, Family Electronic Micro-Database project," *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 2023, *56* (3), 138–159.

Barczyk, Daniel and Matthias Kredler, "Evaluating Long-Tterm-Care Policy Options, Taking the Family Seriously," *The Review of Economic Studies*, 2018, *85* (2), 766–809.

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

Bau, Natalie, "Can Policy Change Culture? Government Pension Plans and Traditional Kinship Practices," *American Economic Review*, 2021, *111* (6), 1880–1917.

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*, 1988, *96* (2), 308–338.

Blau, Peter M., Otis Dudley Duncan, David L. Featherman, and Robert M. Hauser, *Occupational Changes in a Generation, 1962 and 1973 (ICPSR 6162)*, Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 1994. https://doi.org/10.3886/ICPSR06162.v1.

Bugbee, George, "Pensions Plan Is Now Open for Enrollment," *Hospitals: the Journal of the American Hospital Association*, 1946, *20* (9), 59–60.

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, 2006.

Chen, Xi, Karen Eggleston, and Ang Sun, "The impact of social pensions on intergenerational relationships: Comparative evidence from China," *The Journal of the Economics of Ageing*, 2018, *12*, 225–235.

Chetty, Raj, John N. Friedman, Nathaniel Hendren, Maggie R. Jones, and Sonya R. Porter, "The Opportunity Atlas: Mapping the Childhood Roots of Social Mobility," Working Paper 25147, National Bureau of Economic Research 2018.

Clark, Robert L., Lee A. Craig, and Jack W. Wilson, *A History of Public Sector Pensions in the United States*, University of Pennsylvania Press, 2003.

_ , _ , and John Sabelhaus, *State and Local Retirement Plans in the United States*, Edward Elgar, 2011.

Coile, Courtney C., "Economic Determinants of Workers' Retirement Decisions," *Journal of Economic Surveys*, 2015, *29* (4), 830–853.

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.

Edmonds, Eric V., Kristin Mammen, and Douglas L. Miller, "Rearranging the Family?: Income Support and Elderly Living Arrangements in a Low-Income Country," *Journal of Human Resources*, 2005, *40* (1), 186–207.

Eggleston, Karen, Ang Sun, and Zhaoguo Zhan, "The Impact of Rural Pensions in China on Labor Migration," *The World Bank Economic Review*, 2018, *32* (1), 64–84.

Engelhardt, Gary V., Jonathan Gruber, and Cynthia D. Perry, "Social Security and Elderly Living Arrangements Evidence from the Social Security Notch," *Journal of Human Resources*, 2005, *40* (2), 354–372.

Fang, Chichun, "Health and Retirement Study Imputation of Lifetime Earnings," Working paper, Institute for Social Research, University of Michigan 2021.

Feigenbaum, James, "Automated Census Record Linking: A Machine Learning Approach," Working paper 2016.

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.

Ferrie, Joseph P., "A New Sample of Males Linked from the Public Use Microdata Sample of the 1850 US Federal Census of Population to the 1860 US Federal Census Manuscript Schedules," *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 1996, *29* (4), 141–156.

Fetter, Daniel, Lee Lockwood, and Paul Mohnen, "The Intergenerational Incidence of Government Old-Age Support: Evidence from the Early Social Security Era," in Mohnen, Paul Hong-Lum, *Essays in Labor Economics*, Northwestern University PhD Dissertation 2018.

Folbre, Nancy, ed., *For Love or Money*, Russell Sage Foundation, 2012.

Friedman, Jerome, Trevor Hastie, and Robert Tibshirani, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed., Springer, 2009.

Goldstein, Joshua and Casey Breen, "Berkeley Unified Numident Mortality Database: Public administrative records for individual-level mortality research," *Demographic Research*, 2022, *47* (5), 111–142.

Gratton, Brian, "The Poverty of Impoverishment Theory: The Economic Well-Being of the Elderly, 1890–1950," *The Journal of Economic History*, 1996, *56* (1), 39–61.

Grigoryeva, Angelina, "Own Gender, Sibling's Gender, Parent's Gender: The Division of Elderly Parent Care among Adult Children," *American Sociological Review*, 2017, *82* (1), 116–146.

Gruber, Jonathan, *Public Finance and Public Policy*, 4th ed., Macmillan, 2013.

Hamoudi, Amar and Duncan Thomas, "Endogenous Coresidence and Program Incidence: South Africa's Old Age Pension," *Journal of Development Economics*, 2014, *109* (C), 30–37.

Hayhow, Edgar C., "Pensions and Life Insurance Plans for Hospital Employees," *Hospitals: the Journal of the American Hospital Association*, 1940, *14* (11), 73–77.

Health and Retirement Study, *(Imputation of Lifetime Earnings Records) restricted dataset. Produced and distributed by the University of Michigan with funding from the National Institute on Aging (grant number NIA U01AG009740)*, Ann Arbor, MI, 2021.

_ , *(Cross-Wave Geographic Information (Detail) [1992–2020]) restricted dataset. Produced and distributed by the University of Michigan with funding from the National Institute on Aging (grant number NIA U01AG009740)*, Ann Arbor, MI, 2023.

_ , *(Cross-Wave Tracker File [1992–2020]) public use dataset. Produced and distributed by the University of Michigan with funding from the National Institute on Aging (grant number NIA U01AG009740)*, Ann Arbor, MI, 2024.

Helgertz, Jonas, Nesile Ozder, Steven Ruggles, John Robert Warren, Catherine A. Fitch, J. David Hacker, Matt A. Nelson, Joseph P. Price, Evan Roberts, and Matthew Sobek, *IPUMS Multigenerational Longitudinal Panel: Version 1.2 [dataset]*, Minneapolis, MN: IPUMS, 2024.

Horowitz, Amy, "Sons and Daughters as Caregivers to Older Parents: Differences in Role Performance and Consequences," *The Gerontologist*, 1985, *25* (6), 612–617.

Huang, Wei and Chuanchuan Zhang, "The Power of Social Pensions: Evidence from China's New Rural Pension Scheme," *American Economic Journal: Applied Economics*, 2021, *13* (2), 179–205.

Jensen, Robert T., "Do private transfers 'displace' the benefits of public transfers? Evidence from South Africa," *Journal of Public Economics*, 2003, *88* (1-2), 89–112.

Kennan, John and James R. Walker, "The Effect of Expected Income on Individual Migration Decisions," *Econometrica*, 2011, *79* (1), 211–251.

Kestenbaum, Bert, "State and Local Government Employees Covered Under Social Security, 1977-81," *Social Security Bulletin*, 1982, *45* (12), 11–13.

Leibowitz, George J., "Old-Age and Survivors Insurance: Coverage Under the 1950 Amendments," *Social Security Bulletin*, 1950, *13* (12), 3–10, 21.

Leimer, Dean R., "Cohort-Sepcific Measures of Lifetime Net Social Security Transfers," Technical Report, Social Security Administration 1994. Office of Research and Statistics Working Paper Number 59.

Manson, Steven, Jonathan Schroeder, David Van Riper, Katherine Knowles, Tracy Kugler, Finn Roberts, and Steven Ruggles, *IPUMS National Historical Geographic Information System: Version 19.0 [dataset]*, Minneapolis, MN: IPUMS, 2024.

Marquis, James E., "Old-Age and Survivors Insurance: Coverage Under the 1954 Amendments," *Social Security Bulletin*, 1955, *18* (1), 3–10.

McCamman, Dorothy, "State and Local Employees Covered by Government Retirement Systems," *Social Security Bulletin*, 1943, *6* (7), 31–41.

_ , "Higher Educational Institutions and Social Security," *Social Security Bulletin*, 1948, *11* (1), 5–11.

_ , "Retirement Protection for State and Local Employees: Ten Years of Growth," *Social Security Bulletin*, 1953, *16* (5), 3–10,24.

_ , "State and Local Government Employment Under OASI, October 1955," *Social Security Bulletin*, 1956, *19* (2), 15–17.

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

Menchik, Paul L., "Primogeniture, Equal Sharing, and the U.S. Distribution of Wealth," *The Quarterly Journal of Economics*, 1980, *94* (2), 299–316.

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, A. R. Shariq and Paul Mohnen, "Black Economic Progress in the Jim Crow South: Evidence from Rosenwald Schools," *Journal of Economic History*, forthcoming.

Mukherjee, Anita, "Intergenerational Altruism and Retirement Transfers: Evidence from the Social Security Notch," *Journal of Human Resources*, 2022, *57* (5), 1466–1497.

Murray, Merrill G. and Ilse M. Smith, "Higher Educational Institutions and the Social Security Act," *Social Security Bulletin*, 1940, *3* (12), 21–28.

Nakamura, Emi, Jósef Sigurdsson, and Jón Steinsson, "The Gift of Moving: Intergenerational Consequences of a Mobility Shock," *The Review of Economic Studies*, 2022, *89* (3), 1557–1592.

National Education Association, *Statutory Provisions for Statewide Retirement Systems*, Washington, D.C.: National Education Association of the United States, 1946.

Nelson, William, "Employment Covered Under the Social Security Program, 1935-84," *Social Security Bulletin*, 1985, *48* (4), 33–39.

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

Pérez, Santiago, "The (South) American Dream: Mobility and Economic Outcomes of First- and Second-Generation Immigrants in Nineteenth-Century Argentina," *The Journal of Economic History*, 2017, *77* (4), 971–1006.

Roback, Jennifer, "Wages, Rents, and the Quality of Life," *Journal of Political Economy*, 1982, *90* (6), 1257–1278.

Rosen, Sherwin, "Wage-Based Indexes of Urban Quality of Life," in "Current Issues in Urban Economics," Johns Hopkins Univ. Press, 1979, pp. 74–104.

Ruggles, Steven, "Linking Historical Censuses: A New Approach," *History and Computing*, 2006, *14* (1–2), 213–224.

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

_ , Matt A. Nelson, Matthew Sobek, Catherine A. Fitch, Ronald Goeken, J. David Hacker, Evan Roberts, and J. Robert Warren, *IPUMS Ancestry Full Count Data: Version 4.0 [dataset].*, Minneapolis, MN: IPUMS, 2024.

_ , Sarah Flood, Matthew Sobek, Daniel Backman, Annie Chen, Grace Cooper, Stephanie Richards, Renae Rogers, and Megan Schouweiler, *IPUMS USA: Version 15.0 [dataset]*, Minneapolis, MN: IPUMS, 2024.

Schieber, Sylvester J. and John B. Shoven, *The Real Deal: The History and Future of Social Security*, New Haven: Yale University Press, 1999.

Schorr, Alvin Louis, *Filial responsibility in the modern American family: an evaluation of current practice of filial responsibility in the United States and the relationship to it of social security programs*, Washington, D.C.: U.S. Department of Health, Education, and Welfare, Social Security Administration, Division of Program Research., 1960.

Schottland, Charles I., "Social Security Amendments of 1956: A Summary and Legislative History," *Social Security Bulletin*, 1956, *19* (9), 3–31.

Shanas, Ethel, *Family Relationships of Older People: Living Arrangements, Health Status, and Family Ties of Those Aged 65 and Over, As Reportd by the Aged, the Persons to Whom They Would Turn in a Health Crisis, and the General Public*, New York: Health Information Foundation, 1961.

Sharp, Harry and Morris Axelrod, "Mutual Aid Among Relatives in an Urban Population," in Ronald et al. Freedman, ed., *Principles of Sociology: A Text with Readings*, New York: Henry Holt and Company, 1956, pp. 433–439.

Social Security Administration, "Social Security Bulletin, Annual Statistical Supplement 1977-79," 1980. U.S. Government Printing Office.

_ , *Current Population Survey, 1973, and Social Security Records: Exact Match Data (ICPSR 7616)*, Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2005. https://doi.org/10.3886/ICPSR07616.v1.

_ , *Social Security Longitudinal Earnings Public Use File, 1937-1975 (ICPSR 7617)*, Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2008. https://doi.org/10.3886/ICPSR07617.v1.

SSA Office of the Chief Actuary, "OASDI Trustees Report," 2013.

Steiner, Peter Otto and Robert Dorfman, *The Economic Status of the Aged*, Berkeley: University of California Press, 1957.

Streib, Gordon F., "Family Patterns in Retirement," *Journal of Social Issues*, 1958, *14* (2), 46–60.

Subcommittee on Social Security, "Background Material on Social Security Coverage of Governmental Employees and Employees of Nonprofit Organizations," in "Coverage and Termination of Coverage of Government and Nonprofit Organization Employees under the Social Security System: Public Hearings before the Subcommittee on Social Security of the Committee on Ways and Means, U.S. House of Representatives, Ninety-Fourth Congress, Second Session," U.S. Government Printing Office, 1976.

Sussman, Marvin B., "The Help Pattern in the Middle Class Family," *American Sociological Review*, 1953, *18* (1), 22–28.

_ , "Relationships of Adult Children with their Parents in the United States," in Ethel Shanas and Gordon F. Streib, eds., *Social Structure and the Family: Generational Relations*, Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1965, pp. 62–92.

United States Bureau of the Census, *Retirement Systems for State and Local Government Employees: 1941*, U.S. Department of Commerce, Bureau of the Census, 1943.

_ , *Retirement Coverage of State and Local Government Employees*, Washington, D.C.: U.S. Department of Commerce, Bureau of the Census, 1953.

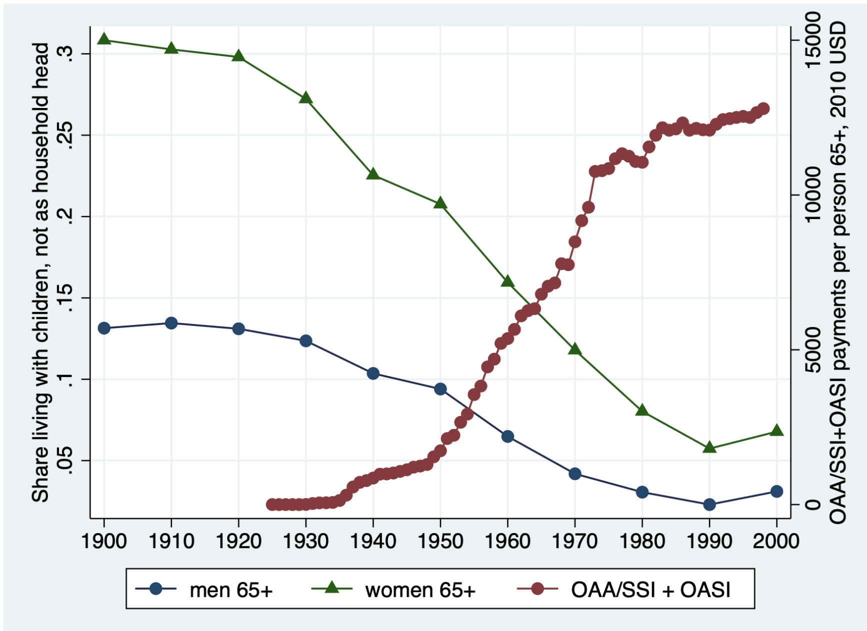
_ , *Current Population Survey: Annual Demographic File, 1973 (ICPSR 7564)*, Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 1992. https://doi.org/10.3886/ICPSR07564.v1.

_ , "National Intercensal Tables, 1900–1990: Population by age, sex, and race," `https://www.census.gov/data/tables/time-series/demo/popest/pre-1980-national.html` 2004. Accessed: 2020-08-10.

Zimran, Ariell, "Sample-Selection Bias and Height Trends in the Nineteenth-Century United States," *The Journal of Economic History*, 2019, *79* (1), 99–138.

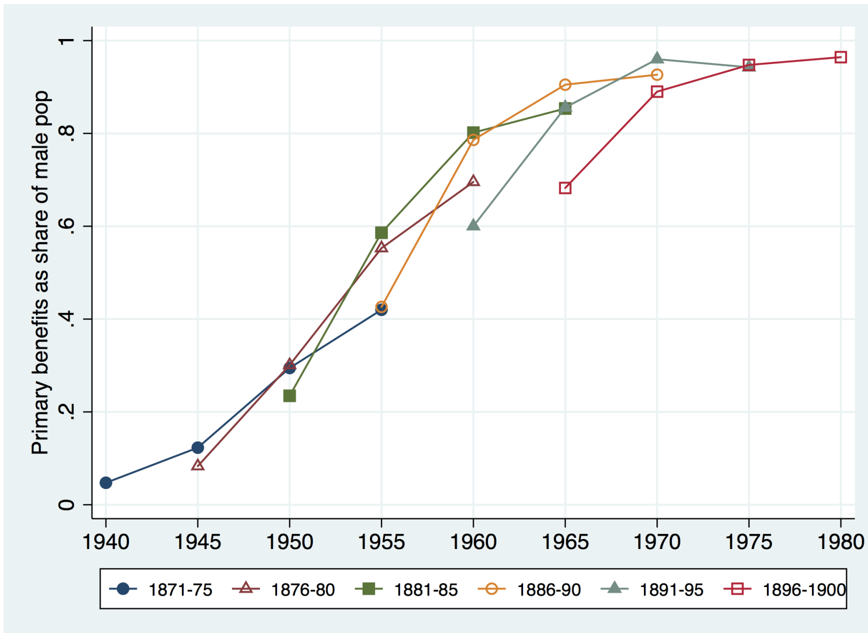
Figures and Tables

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 one of their children or children-in-law, 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). Data on co-residence come from Ruggles et al. (2024b).

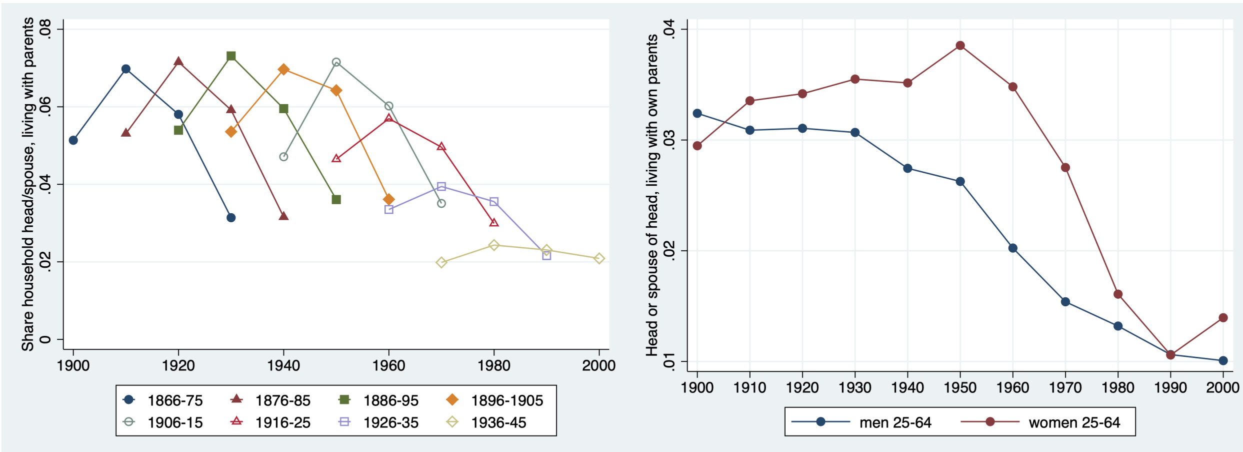
Figure 2: Receipt of retirement benefits by birth cohort and year



Notes: Figure shows share of men of each birth cohort receiving retirement benefits in the specified year, at ages 65–69, 70–74, 75–79, and 80–84. Data on payments are from the *Social Security Bulletin*, data on male population by age and year are from United States Bureau of the Census (2004).

Figure 3: Co-residence from children's perspective

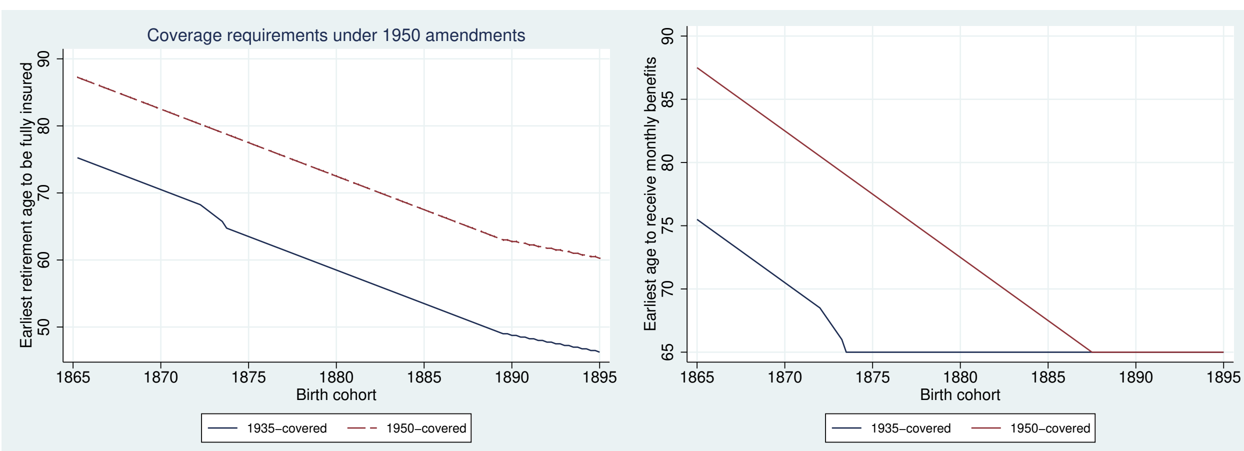
(a) All children, by birth cohort (b) Co-residence rates with own parents, by sex



Notes: Panel (a) shows share of US-born men and women of each birth cohort who are household head or spouse of the head, and living with a parent or parent-in-law. Each cohort is shown at ages 25–34, 35–44, 45–54, and 55–64. Panel (b) shows share of US-born men 25–54 and US-born women 25–54 who were household head or spouse of the head and lived with their own parents in the household in each year. Data come from Ruggles et al. (2024b).

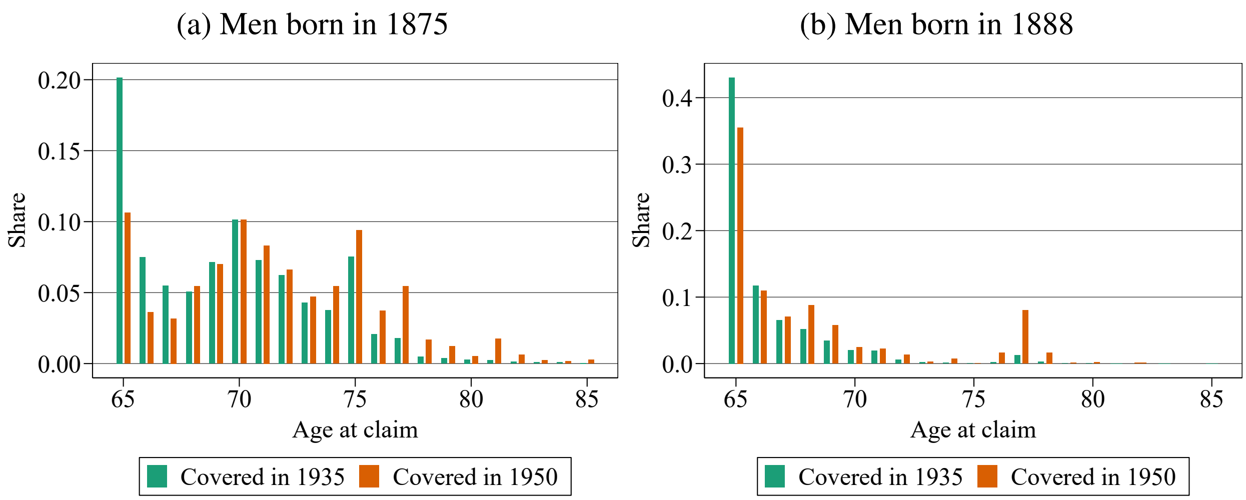
Figure 4: Earliest age when permanently insured and earliest age for benefits

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



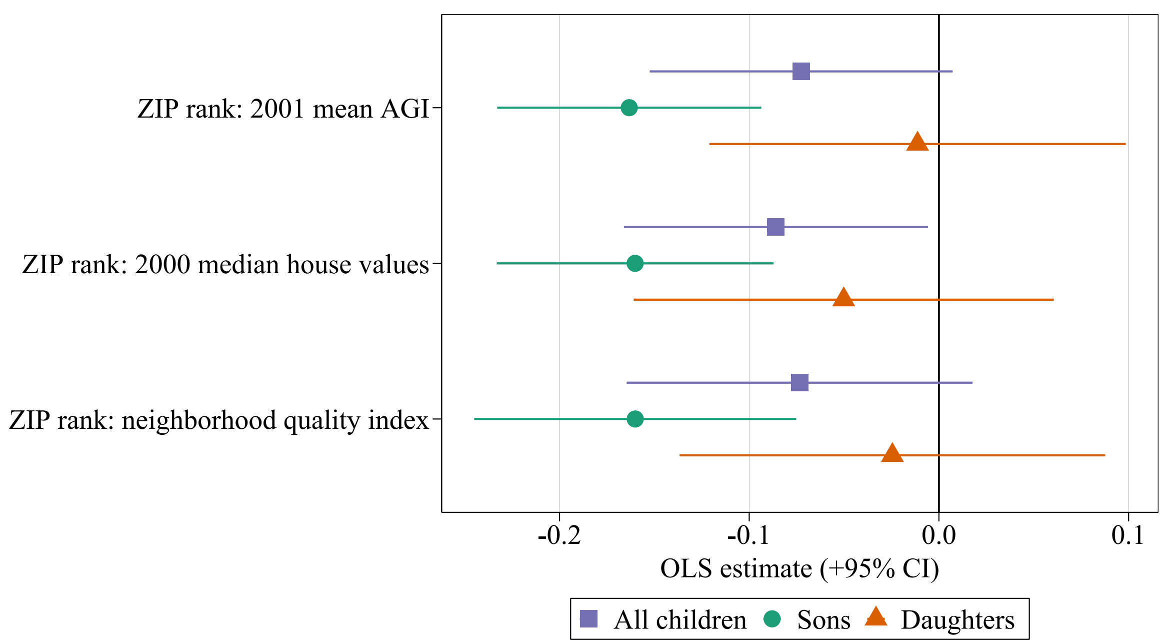
Notes: Minimum ages based on Social Security Act of 1935, 1939 Amendments, and 1950 Amendments, assuming that a worker in 1950-covered employment has no employment that was covered in 1935. The minimum ages shown are those under the 1950 Amendments.

Figure 5: Age at claim distribution by Social Security coverage, as implied by 1930 employment



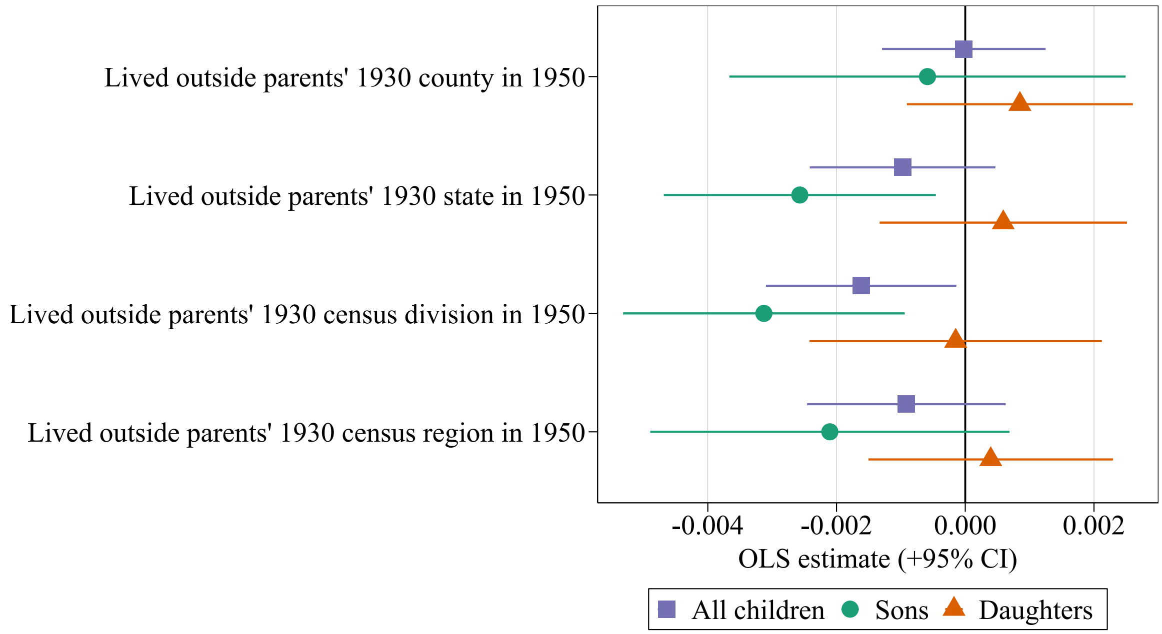
Notes: These graphs are based on the Numident claims-census linked sample, and plot the age at claim distribution for men whose 1930 employment was covered by Social Security in 1935 vs. 1950 (excluding the self-employed), separately for men born in 1875 and men born in 1888. Observations are weighted using IPWs.

Figure 6: The impact of father's later Social Security coverage on children's death ZIP rank



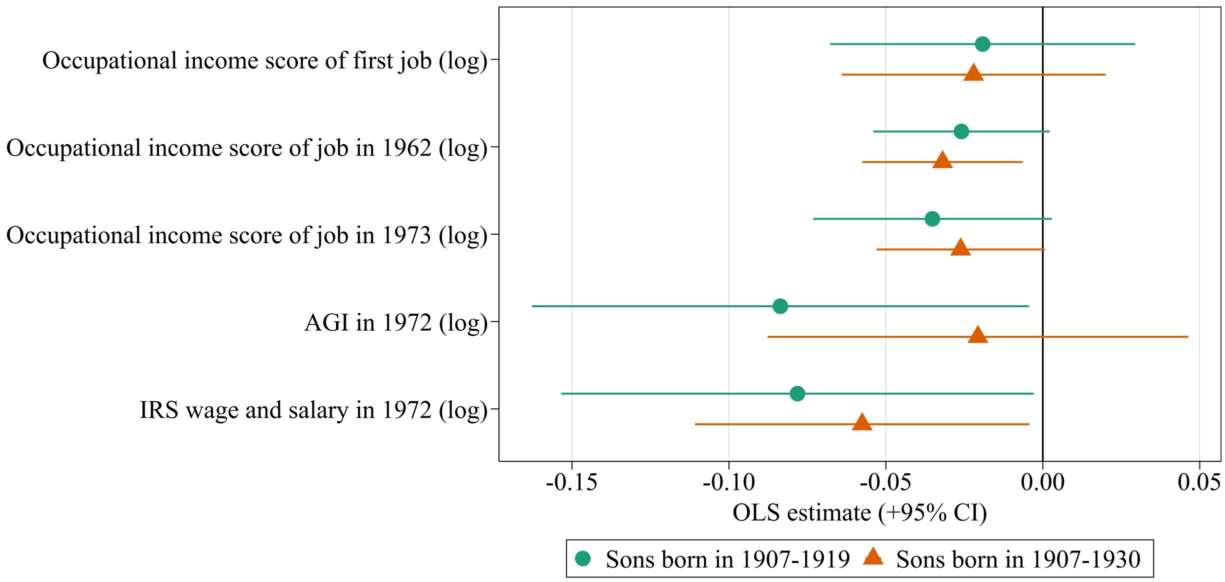
Notes: This figure displays the coefficients on father's minimum age of Social Security eligibility from equation (3), where the outcome is the average percentile rank of his children's ZIP code at death in terms of mean AGI, median house values or the neighborhood quality index. Colors/shapes indicate separate estimates for all children, sons, or daughters. All regressions include fixed effects for sex × cohort and father's occupation × year of birth, Social Security coverage category, race, and state/country of birth. Observations weighted using normalized IPWs. 95% confidence intervals based on standard errors clustered at the level of father's industry of employment.

Figure 7: The impact of father's later Social Security coverage on children's propensity to live outside their parents' 1930 place of residence in 1950



Notes: This figure displays the coefficients on father's minimum age of Social Security eligibility from equation (3), where the outcome is an indicator for living outside parents' 1930 county/state/census division/census region of residence in 1950. Colors/shapes indicate separate estimates for all children, sons, or daughters. All regressions include fixed effects for sex × cohort and father's occupation × year of birth, Social Security coverage category, race, and state/country of birth. Observations weighted using normalized IPWs. 95% confidence intervals based on standard errors clustered at the level of father's industry of employment.

Figure 8: The impact of father's later Social Security coverage on sons' occupational standing and income in adulthood



Notes: This figure displays the coefficients on father's minimum age of Social Security eligibility from equation (3), where the outcome is indicated in the row title. Colors/shapes indicate separate estimates for sons born in 1907–1930 and sons born in 1907–1919. All regressions include fixed effects for father's occupation × year of birth, Social Security coverage category, and son's year of birth, race, and state/country of birth. Observations weighted using OCG sampling weights. 95% confidence intervals based on standard errors clustered at the level of father's industry of employment.

Table 1: The impact of later Social Security coverage on subsequent claiming behavior

	Dependent variable:	
	Linked to claims data	Age at claim
	(1)	(2)
Min. age of Social Security eligibility	-0.0035***	0.0845***
	(0.00)	(0.01)
Mean of dep. var.	0.1086	67.1935
R^2	0.023	0.157
N	4,133,214	448,439

Notes: Unit of observation is a head of household man in 1930 born in 1875–1888 and covered by Social Security in 1935 or 1950 based on his employment information, excluding self-employed workers. All regressions include fixed effects for occupation × year of birth, Social Security coverage category, 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 2: Selection of 1930 households and children into our linked sample

	Dependent variable: Any/number of children linked					
	All children		Sons		Daughters	
	Any	Number	Any	Number	Any	Number
	(1)	(2)	(3)	(4)	(5)	(6)
Min. age of Social Security eligibility	-0.0000	-0.0002	-0.0002	-0.0004	0.0002	0.0002
	(0.0003)	(0.0007)	(0.0003)	(0.0005)	(0.0002)	(0.0003)
Mean of dep. var.	0.2153	0.3576	0.1352	0.1764	0.1353	0.1812
R^2	0.017	0.016	0.012	0.011	0.011	0.011
N	4,127,539	4,127,539	4,127,539	4,127,539	4,127,539	4,127,539

Notes: Unit of observation is a 1930 household headed by a man born in 1875–1888 and covered by Social Security in 1935 or 1950 based on his employment information, excluding self-employed workers. The dependent variable is either an indicator equal to one if at least one child of a specific sex was linked to the 1930 household, or a discrete variable capturing the number of children of a specific sex linked to the 1930 household. All regressions include fixed effects for head of household occupation × year of birth, Social Security coverage category, race, and state/country of birth. Robust standard errors in parentheses, clustered at the head of household industry level. *** 1%, ** 5%, * 10% significance.

Table 3: Placebo tests: 1930 Census household characteristics

	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									
Min. age of Social Security	0.0009*	0.0101*	0.0001	-0.0001	-0.0003	-0.0017	0.0004	0.0002	0.0001
eligibility	(0.0005)	(0.0052)	(0.0010)	(0.0008)	(0.0002)	(0.0020)	(0.0002)	(0.0002)	(0.0003)
Mean of dep. var.	0.486	4.0752	0.4654	0.7297	0.0562	2.1763	0.9404	0.9341	0.1097
R^2	0.059	0.082	0.213	0.226	0.219	0.104	0.149	0.192	0.058
N	4,118,719	4,017,290	4,127,539	4,127,539	4,127,539	4,127,539	4,127,539	3,781,040	3,781,042
Panel (b): All families in linked sample									
Min. age of Social Security	0.0001	0.005	0.0000	-0.0000	-0.0007	-0.0030	0.0003	0.0002	0.0006
eligibility	(0.0006)	(0.0067)	(0.0011)	(0.0007)	(0.0005)	(0.0027)	(0.0004)	(0.0005)	(0.0007)
Mean of dep. var.	0.4846	4.0116	0.4581	0.7185	0.0741	3.8998	0.9414	0.9376	0.0878
R^2	0.072	0.102	0.214	0.246	0.241	0.128	0.14	0.181	0.058
N	888,031	870,465	888,759	888,759	888,759	888,759	888,759	870,884	870,884

Notes: Unit of observation is a 1930 household headed by a man born in 1875–1888 and covered by Social Security in 1935 or 1950 based on his 1930 employment information, excluding self-employed workers. All regressions include fixed effects for head of household occupation × year of birth, Social Security coverage category, race, and state/country of birth. Observations weighted using IPWs in panel (b). Robust standard errors in parentheses, clustered at the head of household industry level. *** 1%, ** 5%, * 10% significance.

A Appendix

A.1 Details on linkages

A.1.1 Linking families in SS-5 records to households in the 1920 and 1930 Censuses

The first step involves reconstituting families in the SS-5 data by linking siblings together primarily based on parent names.[37] The sibling linkages are then used to create a crosswalk between SSNs and family identifiers. In practice, we borrow this crosswalk from Mohammed and Mohnen (forthcoming). The sibling linkages were made using the LIFE-M supervised ML approach described in this paper, except that linking siblings together is an instance of one-to-many rather than one-to-one linking (i.e., individuals can have multiple siblings in the SS-5 data). As a result, trainers were allowed to select multiple links, and a random forest model was used to model their decisions. In particular, each primary-potential sibling pair was treated as a separate linking decision, even though trainers were technically allowed to select multiple siblings simultaneously. Further details can be found in Section B.2 from the online appendix of Mohammed and Mohnen (forthcoming). The crosswalk contains 5.9 million families with at least two siblings, starting from 40 million unique SSNs. Alternatively, 35 percent of SSNs are linked to at least one other SSN in the SS-5 data.

The next step involves linking reconstituted families in SS-5 records to households in the 1930 Census. Although we are mainly interested in families in which the father was born in 1875–1888, since SS-5 records lack parental year of birth information we link all families with at least one child aged under 25 at the time of the census, and later restrict the sample based on age information from the census. Note that families can be a single SS-5 record if no sibling was found in the previous step. To generate the set of candidate census households for each family, we block on father last name initial and the *union* of children's state/continent of birth.[38] Blocking on the union of children's place of birth implies that the only requirement we impose is that at least one of the potential children's place of birth matches one of the primary children's place of birth. Within each block, we select the top 20 potential households based on a Jaro-Winkler string similarity score between primary and potential parent names.

We generate training data for a random sample of 500 primary families using the process described in Section 3. In particular, we display the following information to trainers: father full name, mother first and middle name, and children's first and middle name, age in 1930, and state/country of birth (we also display the modal race among children). We only display primary children that would have been alive at the time of the census, and mask potential children in the census if there is little chance that they match any of the primary children (based on age and place of birth considerations). We limit the number of children displayed to a maximum of eight to avoid overcrowding the screen during the training process (the eight youngest ones).

We are able to manually link 62 percent of families to a household in the 1930 Census in the training data. To train the two-stage model from Bailey et al. (2023) to make similar linking decisions, we first split the training data into a "training set" and a "test set" using a 70/30 percent split. The training set is used as input data to train the algorithm, while the test set can be used

[37] Race, place of birth, and year of birth information is also used in the linking decision, as siblings tend to be born in the same state/country a few years apart (and have matching race).

[38] Blocking on a subset of characteristics is common practice in the historical linking literature, and is necessary for computational reasons as string similarity scores are expensive to compute.

to assess the out-of-sample performance of the model (which typically lines up quite well with the in-sample performance). The set of features used to train the model is shown in Appendix Table A.9, and probabilities are generated using 10-fold cross-validation as described in Section 3. Panel (a) in Appendix Figure A.14 shows the resulting precision-recall curve. The model is able to achieve a recall rate of 85 percent at 97 percent precision. In other words, fixing the error rate of model-based links relative to human trainers at 3 percent, the model is able to replicate 85 percent of the links made by human trainers.

Using the trained model and cross-validated probability threshold, we scale-up linking to the full sample of SS-5 families. After doing so, it is not uncommon for the same census household to be linked to two (or more) different families in the SS-5 data. While some of these cases are true conflicts, a significant share of them are not. The reason is that we were not able to make all possible sibling linkages in the SS-5 data since parent names are not always consistently reported across siblings' SS-5 records. Although we might not be able to link two SS-5 records together solely based on parent name information, we still might be able to (correctly) link them to the same census household since the latter additionally exploits information on children. As a result, we end up with apparent conflicts that simply reflect the incompleteness of our family identifiers.

Rather than throwing away all these conflicts, we manually examined a random sample of them (200), and either chose one link over the others, concluded all links were correct, or chose none of them if we were not confident. We then used this data to assess the performance of different conflict resolution rules in terms of their ability to retain all correct links and drop all incorrect links.[39] The rule we ultimately settled on is able to retain 80 percent of correct links and drop 75 percent of incorrect links in the random sample of cases we manually inspected. The final number of SS-5 families linked to a household in the 1930 Census at 97 percent precision after resolving these conflicts is 6.9 million. Using an analogous procedure and recycling the training data and model, we are able to link 4.1 million SS-5 families to a household in the 1920 Census.

A.1.2 Linking households in the 1930 Census to households in the 1920 Census

The process of linking households in the 1930 Census to households in the 1920 Census is very similar to the one used to link SS-5 families to households in the 1930 Census. One slight difference is that the censuses additionally contain information on parents' age and place of birth. Therefore, when generating sets of potential candidates, we block on fathers' state/continent of birth and a +/-5-year window around fathers' age in 1920 (in addition to father last name initials). We also display father age in 1920, mother age in 1920, and mother place of birth information during the training process. We are able to manually link 53 percent of households headed by a man born in 1860–1905 in the 1930 Census to a household in the 1920 Census in our training data (500 cases). The model recall rate at 97 percent precision is 90 percent (see panel (b) in Appendix Figure A.14). Note that after scaling up the model it is possible for a household in the 1920 Census to be linked to multiple households in the 1930 Census. But in contrast to when we link SS-5 families to census households, these are true conflicts that primarily reflect name commonality. In some cases, we selected one link over the others based on perceived quality (as measured by the match probability). In the remaining cases, we dropped all conflicting links. The final number of

[39] The two extreme rules are to keep all conflicts or drop all conflicts, which exactly achieve one objective but not the other. The optimal rule lies somewhere in between.

households headed by a man born in 1860–1905 in the 1930 Census linked to the 1920 Census at 97 percent precision is 7.6 million.

A.1.3 Linking men and women in SS-5 records to themselves in the 1950 Census

To link men and women in SS-5 records to themselves in the 1950 Census, we used the linking process described in Mohammed and Mohnen (forthcoming) to link SS-5 records to the 1940 Census (including the training data and ML model). Here we provide a brief overview of this process, starting with how we link men and then describing how we link women.

For male primary SS-5 records, we search for potential candidates in the 1950 Census that match on sex, first or last name initial, state/continent of birth, and a +/-3 year window around the primary record's age in 1950.[40] We then compute Jaro-Winkler name similarity scores between the primary record and all potential candidates that satisfy those conditions, and keep the top 25 candidates. The original training data for linking men in SS-5 records to the 1940 Census was generated using a similar process to the one described in Section 3. It was based on a random draw of 1,500 records, and the following information was displayed to trainers: full names (including middles names, if available), age in 1940, country of birth (if born outside the U.S.), and race.[41] The match rate in this training data was 43 percent, and the recall rate for the model trained on this data was 75% at 95% precision. We used this trained ML model and the cross-validated probability thresholds to generate 95%-precision links to the 1950 Census for men.

The process of linking female SS-5 records is analogous, except for a few key differences. First, two versions of each female record is generated, one with the married name as the last name and one with the birth name as the last name (when both are available). Birth and married names are determined by comparing the last name field with the father last name field. Second, we additionally block on marital status when generating potential candidates. Women linked using their birth name are linked to single women in the census, while women linked using their married name are linked to non-single women in the census (i.e., married, divorced, or widowed). The original training data for linking women was based on a random sample of 500 SSNs, corresponding to around 1,500 primary SS-5 records (in this case, a primary record is a particular name variant of the same woman). For individuals that had multiple links in the training data, we manually resolved those cases by either choosing one link over the other or deciding there was too much ambiguity and discarding both links. The SSN-level match rate in this training data was 52% and the SSN-level recall rate was 74% at 95% precision. Again, we used this trained ML model and the cross-validated probability thresholds to generate 95%-precision links to the 1950 Census for women (for model-based links we simply discard both links when there is a within-SSN conflict).

Similar to household linking, while the model links each SSN to a unique census record, it is possible, though less common, for the same census record to be linked to two different SSNs. However, in the case of individual linking, this reflects true conflicts due to name commonality. To resolve those cases, we prioritize links that have a higher precision rate, and discard the remaining conflicts.

[40] Blocking on first or last name initials implies that potentials candidates only need to satisfy one of those two conditions.

[41] Race and country of birth was displayed as optional pieces of information to take into account, knowing that they are not always accurately reported.

A.1.4 Linking individuals in claims records to individuals in the 1930 Census

Claims records for men born between 1865 and 1899 are linked to the 1930 Census using individual-level information. Given that our objective is to explore claiming behavior for regular retirement benefits, we focus on the subset of life claims.[42] We separately link claims records with non-missing place of birth information (40 percent of records) and claims records with missing place of birth information (60 percent of records). Claims records with non-missing place of birth information are linked using name, year of birth, and place of birth information. To generate sets of potential candidates, we block on state/continent of birth, last name initial, and a +/-3-year window around age in 1930. We then compute a string similarity score that accounts for the possibility of inverted first and middle names, and select the top 20 candidates. We manually linked a random sample of 1,000 cases, displaying name, age, and country of birth information (when applicable).[43] We also display the state associated with the first 3 digits of a primary's SSN and the state of residence of potential candidates in the 1930 Census.[44] Since 1930 is relatively close in time to the period in which these cohorts would have applied for an SSN, this optional piece of information is useful when choosing between multiple plausible candidates.

We are able to link 53 percent of claims records with non-missing place of birth information to an individual in the 1930 Census, and the model achieves a recall rate of 67 percent at 97 percent precision (see panel (c) in Appendix Figure A.14). We use an analogous procedure to link claims records with missing place of birth information, except that we cannot block on state/continent of birth, and instead block on the first two last name initials for computational reasons (in addition to the +/-3-year age window). This results in larger blocks and more ambiguity in the set of potential candidates. Consequently, we are only able to manually link 34 percent of claims records with missing place of birth information to an individual in the 1930 Census. The model recall rate at 97 percent precision is also slightly lower at 65 percent (see panel (d) in Appendix Figure A.14). After scaling up linking to the full set of claims records and resolving conflicts in which the same census individual was linked to two different claims records, we end up with a sample size of 1.3 million.

A.1.5 Linking OCG records to other survey/administrative records

Our goal here is to link four sets of records: (1) the 1973 wave of the Occupational Changes in a Generation survey (OCG), (2) the Social Security Records: Exact Match Data file (EMD), (3) the 1937–1975 Social Security Longitudinal Earnings File (SSLEF), and (4) the 1973 Current Population Survey (CPS). Our starting point is the OCG survey as it contains information on fathers' employment and year of birth. The other three datasets contain additional outcomes for sons. The EMD file contains tax record information from the Internal Revenue Service for the 1972 fiscal year, the SSLEF contains Social Security earnings records covering the period 1937–1975, and the CPS contains more granular 1973 place of residence information than in the OCG.

Unfortunately, there is no unique identifier to merge these datasets together. Instead, we proceed in three steps. First, we merge the EMD file with the SSLEF based on row order. More

[42]Life claims capture retirement and disability benefits. Death claims, which we exclude, capture survivor and death benefits. Around 77 percent of records in the claims data are life claims.

[43]For individuals born outside the U.S., only continent of birth matches, so an exact match on country of birth has additional informational content.

[44]Until 1972, the first 3 digits of an SSN was tied to the field office in which the application was filed.

specifically, after restricting the EMD file to individuals that should be found in the SSLEF, we end up with the same number of observations. Moreover, the observations seem to have been ordered in the same way in the raw data, as evidenced by the fact that demographic variables that appear in both datasets (sex, race, month of birth, and year of birth) match for over 99% of observations. Second, while the EMD file and the CPS file have the same number of rows, they are not ordered in the same way. Instead, we first impose the OCG sample restriction on both datasets (civilian males aged 20–65), and identify 24 variables that appear in both datasets and for which aggregate distributions line up exactly. These 24 variables uniquely identify over 99.9% of observations in both datasets. We therefore restrict the samples to uniquely-identified observations and merge them together. Similarly, we merge the EMD file and OCG file based on 34 variables that are contained in both datasets. The final merged OCG-EMD-SSLEF-CPS file contains 34,101 observations, or 99.9% of the original 34,142 observations in the OCG dataset.

A.2 Classification of employment into OASI coverage categories

A.2.1 Identifying Social Security coverage categories

Our assignment of Social Security coverage year to an individual is based on his reported occupation, industry, and "class of worker" (which includes information on whether a worker was self-employed, a wage or salary earner, or an unpaid family worker). Although our analysis uses only some of the Social Security coverage categories, we classify all combinations of occupation, industry, and class of worker and describe our approach here. We do not offer an exhaustive description of all categories, but cover the largest ones. Our description and classification is based primarily on Nelson (1985), supplemented with information from Marquis (1955) and Social Security Administration (1980), as well as the Social Security Act and subsequent amendments. Additional useful references include Leibowitz (1950) on the 1950 Amendments, Schottland (1956) on the 1956 Amendments, and McCamman (1956) on state and local government workers under the 1954 Amendments.

The original Social Security Act in 1935 included most wage and salary workers in its definition of covered employment. It excluded self-employment, as well as certain types of wage and salary work, the largest of which were agricultural wage work; domestic work in private homes; and work for nonprofit organizations, state and local government, the federal government, and railroads.

Since our analysis compares only wage and salary workers, we first describe the extension of Social Security coverage to the categories of wage and salary workers that were excluded from coverage in the 1935 Act. Regular agricultural wage work and domestic work were both covered under the 1950 Amendments, and the requirements for regularity were relaxed under the 1954 and 1956 Amendments. Employees of nonprofit organizations had been excluded from coverage under the 1935 Act because of the tax-exempt status of nonprofits (McCamman, 1948), but (with the exception of ministers) could be covered on an elective basis starting with the 1950 Amendments. The decision to elect coverage was made by both employers and employees: if the organization waived its exemption from Social Security taxes and employees voted to be covered by Social Security, all new employees and existing employees who voted for coverage would be covered, while existing employees who voted against coverage would not be covered. Although not all organizations elected coverage immediately upon its becoming available, take-up was reasonably rapid: as of 1951, 56% of employees of nonprofit organizations were covered under Social Security, rising to 76% by 1960 (Subcommittee on Social Security, 1976, p. 27). State and local government employees were initially excluded from coverage because of perceived constitutional limitations on levying taxes on state and local governments (see, e.g. Clark et al., 2011), but beginning with the 1950 Amendments, states could voluntarily enter into agreements for Social Security to cover employees who were not covered under an existing retirement system. The 1954 Amendments extended coverage to state and local employees already under a retirement system, with the exception of firefighters and police officers, if the state elected to make coverage available and employees voted in favor of coverage. Amendments in 1956 further liberalized the provisions under which state and local government employees could be covered by Social Security.[45] About 11 percent of all state and local government employees were covered by Social Security at the beginning of 1951 (the first year in which they could be), and by 1958 the majority of state and local government employees (54 percent) were covered by Social Security (Kestenbaum, 1982).

Other significant types of work excluded from coverage under the 1935 Act, but that do not play a role in our analysis, are self-employment and work for railroads or the federal government. The 1950 Amendments covered most nonfarm self-employment, but excluded self-employed farmers and farm managers, as well as certain professional groups (accountants, architects, engineers, funeral directors, lawyers, medical professionals, and veterinarians). Most self-employed farmers and farm managers were covered in 1954, and additional farmers were covered under liberalized provisions in 1956. Professional self-employment, except dentists, medical professionals, and lawyers, were covered under the 1954 Amendments; except for self-employed doctors, the professional groups excluded in the 1954 Amendments were covered under the 1956 Amendments. Self-employed doctors were covered under the 1965 Amendments. Railroad workers with at least ten years of service were largely covered under the Railroad Retirement Act of 1937, a system parallel and fairly similar to Social Security; those with less than ten years of service were covered by Social Security beginning in 1951. Many employees of the federal government were covered by existing retirement programs in 1935, and therefore excluded from Social Security. The 1950 Amendments extended coverage to a large number of federal employees, mostly part-time or temporary employees, who were not under a Federal retirement system. The 1983 Amendments extended coverage to all civilian employees of the Federal government hired in 1984 onwards. Military employment was covered under the 1956 Amendments (veterans also received noncontributory wage credits for military service between 1940 and 1956).

Our main coding of Social Security coverage is based on the 1950 Census classifications for occupation and industry, which IPUMS uses as a common coding across years (Ruggles et al., 2024b), as well as "class of worker" (which includes information on whether a worker was self-employed, a wage or salary earner, or an unpaid family worker). The foregoing discussion noted that there was a regularity of employment condition for coverage in some types of work, such as agricultural wage work and domestic work. A single Census provides information at only one point in time and, in the case of the 1930 Census, does not provide sufficient information on regularity of employment to map into the regularity of employment tests. Hence, our classification implicitly assumes that, for example, a worker who reports his occupation as an agricultural wage worker is employed on a sufficiently regular basis that he would be covered by Social Security in the first

[45] Firefighters and police were treated separately, and do not play a role in our analysis. The 1956 Amendments allowed firefighters and police to be covered in certain states, and the 1967 Amendments allowed firefighters to be covered in all states.

instance (1950 for agricultural wage workers).

We assign 1935 coverage to any wage and salary worker who, based on industry and occupation, was not in one of the original excluded categories. The exclusions were primarily on the basis of industry, although there are a few small categories we exclude on the basis of occupation (such as fishermen and oystermen, or officers and crew members of ships). We assign 1950 coverage to the categories of employment covered on a compulsory basis: agricultural wage workers, domestic workers, and the non-farm self-employed. We also assign 1950 coverage to types of employment that were covered on an elective basis beginning in 1950: nonprofits and state and local government. We treat the following industry codes as implying employment in either nonprofits or state and local government: hospitals, educational institutions, welfare and religious organizations, nonprofit membership organizations, and state and local public administration. Of these, hospitals were one case with a nontrivial number of for-profit entities: based on American Hospital Association (AHA) data, about 30 percent of AHA hospitals were were private in 1935.[46] However, private hospitals tended to be smaller than government or nonprofit hospitals: they held only about 6 percent of hospital beds in 1935, suggesting that they represented only a small share of hospital employment. In the main analysis, we do not attempt to separate out for-profit hospitals, but their presence should tend to attenuate the first stage—which nevertheless is strong—and the size of the reduced form effects. We are aware of no simple way in the Census data of distinguishing nonprofit from government employment in hospitals or education, so we treat them in a parallel way, with elective coverage beginning in 1950. Finally, we categorize police and firefighters separately due to their special coverage under the Social Security Act and exclude them from our analysis. They would not be part of the basis for comparisons anyway, since there would be no variation in Social Security coverage within the same occupation.

If a worker's industry, occupation, or self-employment information is missing or not classified into informative codes, we classify their Social Security coverage status as "unclear" and exclude them from our analysis. Similarly, in some cases a worker's reported industry and occupation may imply different dates of coverage (such as a worker whose occupation is "housekeeper, private household" and industry is "Security and commodity brokerage and investment companies."). These cases we also classify as unclear. In general, we do not exclude, or attempt to correct, apparently inconsistent occupation and industry responses that nevertheless have the same implication for dates of coverage.

To apply our Social Security coverage categories in the OCG sample, which reports occupation and industry according to 1970 Census codes, we first map those codes into 1950 codes, and then apply the Social Security coverage classification described above. Fortunately, despite the changes between 1950 and later years, we still assign a unique Social Security coverage code to each combination of occupation, industry, and self-employment status in the 1970 codes.

A.2.2 Validation of the coverage definition

In addition to the first stage analysis presented in the paper, in Appendix Table A.3 we verify our coverage definition against the 1940 Census, where 5% of the population was asked whether they had obtained a Social Security Number (SSN). The vast majority of those whose 1940 employment is classified as covered in 1935 had obtained an SSN by 1940, while those whose employment is

[46]The AHA data appear in the August issues of *Hospitals: The Journal of the American Hospital Association*. We thank Heidi Williams for generously sharing a digitized version of the data.

classified as not covered until later mostly had not obtained an SSN by 1940. Those whose 1940 employment we do not classify as covered, but who did report having an SSN, may have been employed in a private firm in a largely nonprofit industry; they may have formerly held covered employment; they may have been workers whose employment varied seasonally; or they may have simply registered for an SSN on their own, since coverage of one's employment was not required to get one.

A.3 Pension coverage in the comparison group

Because comparisons are restricted to those within occupation and using variation that arises from exclusions of certain industries from coverage under the original Social Security Act, the comparison group in our empirical analysis is made up largely of certain categories of domestic workers, employees of nonprofits, and employees of state and local governments. To aid interpretation of our main results, it is useful to characterize the prevalence and characteristics of retirement plans for these types of workers. Although we are aware of no systematic information on "retirement plans" for domestic workers (formal or informal), there is enough information on retirement plans for nonprofit and state and local workers to characterize their pension coverage, at least broadly.

A large share of State and local government employees were covered by retirement systems prior to their inclusion in the Social Security system, but within that group, the occupations that form the basis of our main empirical comparisons were relatively less likely to be covered by a retirement system. At the beginning of 1942, when the Census Bureau carried out the first systematic national survey, 46 percent of all State and local government employees were covered by a retirement system; the likelihood of coverage was greater for employees of school systems (58 percent) than for non-school employees (38 percent) (United States Bureau of the Census, 1943, Table 2). This aggregate figure conceals wide disparities across states: in California, New York, Ohio, and Pennsylvania, more than 85 percent of permanent, full time employees were covered, while in a handful of states (Alabama, Idaho, Iowa, Mississippi, Missouri, Nebraska, North Carolina, Oklahoma, Oregon, South Carolina, South Dakota, and Wyoming) less than ten percent of permanent, full-time employees were covered (McCamman, 1943). Coverage under State and local retirement systems grew over the course of the 1940s. In the next nationwide survey, in 1952, the share of State and local government employees covered by a State or a local retirement plan had increased to 67 percent, with a further 9.7 percent covered under Social Security; 74 percent of school employees and 62 percent of non-school employees were covered by State or local systems, with an additional 4.9 percent and 13 percent, respectively, covered by Social Security (United States Bureau of the Census, 1953, Table 2).

Beyond broad categories, there is no systematic information on which types of employees were covered in which localities or states, but the available evidence suggests that the public employees that form part of the basis of our comparison group were, compared to other public employees, less likely to be covered by retirement systems. The State and local employees with the highest rates of coverage by State and local systems were police, firefighters, and teachers (McCamman, 1943; Clark et al., 2003, 2011). Police and firefighters in particular had a long history of coverage and are thought to have had close to universal coverage, typically through localities rather than States, by the end of the 1920s; teachers also had widespread, though far from universal, coverage under either State or local plans (Clark et al., 2003, 2011). We exclude police and firefighters from our analysis entirely, and teachers are used for the comparison only to the extent that they appear in covered industries (private, for-profit entities) as well as uncovered ones (public or nonprofit organizations).

On the other hand, particularly as of the early 1940s, less than a third of states had general plans for nonschool employees (United States Bureau of the Census, 1943; McCamman, 1953). Altmeyer (1945) noted that within retirement systems for school department employees, it was also often the case as of the mid-1940s that only teachers were eligible, not custodial or nonteaching employees. Consistent with this claim, a 1946 National Education Association study that described characteristics of retirement plans for school employees in 45 states indicated that all 45 states had plans that covered instructional staff of public school systems, but only 15 of these had plans in which custodial employees of school districts could have membership, and 27 had plans in which clerical employees were covered (National Education Association, 1946, p. 18). Retirement plans for public employees grew significantly over the 1940s, which likely reduced this gap to some extent; McCamman (1953) reports that 32 states had state-wide systems for nonschool city employees as of 1950, 21 of which had been introduced since the 1942 survey. Nevertheless, it appears that many of the public-sector workers who form the basis of our comparison group (such as custodial employees of school districts) were not as likely to be covered by retirement systems as were public employees as a whole.

The nature of benefits under State and local retirement plans varied widely, but there are a few generalizations that are relevant to considering how existing retirement coverage in the control group would affect the interpretation of our results. First, public retirement systems tended to have less generous survivor's insurance than Social Security. Most relevant to our analysis, many of these plans had no provision for widow's benefits, or required that retirees accept a lower benefit during their lifetime in order to continue a benefit to a survivor if the retiree died first (McCamman, 1953). This is particularly relevant to our analysis in that co-residence, and perhaps other types of transfers as well, were especially common among widows. Second, benefits under public retirement systems tended to favor lower-paid workers to a lesser extent that Social Security did; they also required longer periods of service than Social Security, and unlike Social Security after the 1939 Amendments, heavily weighted benefits in favor of those with longer periods of service (McCamman, 1953).

Retirement plans for employees of nonprofit organizations existed by 1940, but appear to have been less common than for public employees. Altmeyer (1944) noted that there were voluntary plans made available by professional associations as of the early 1940s, but take-up tended to be low. Based on a survey carried out by the American Hospital Association (AHA), Hayhow (1940) estimated that about 8 percent of nonprofit hospitals in the United States and Canada had pension plans. In 1946, the AHA introduced its own pension plan, which all member nonprofit hospitals were eligible to join (Bugbee, 1946). By early 1949, 120 hospitals and allied organizations had joined (American Hospital Association, 1949); this take-up was relatively modest compared to the 3,160 nonprofit institutions at the time. Like many public retirement systems, coverage under the AHA plan differed from Social Security; of particular note is that securing continuing payments to a retiree's widow after his death required election of a smaller benefit during one's retirement.

Private institutions of higher education comprise another major group of nonprofit organizations. In the 1940s, about one-third of the approximately 1,700 institutions of higher learning were public, and two-thirds were private; of those two-thirds, more than 90 percent were nonprofit organizations (Murray and Smith, 1940; McCamman, 1948). McCamman (1948) reported that about 30 percent of private institutions (or 338 total) had retirement plans, 78 percent of which were through contracts with the Teachers Insurance and Annuity Association (TIAA). Faculty in these institutions are not part of our empirical comparisons, since they do not appear in both covered and uncovered industries (and because of the rarity of for-profit higher educational institutions, we classify all institutions as public or nonprofit); instead, the employees most relevant to our comparisons are maintenance and other nonprofessional employees. For the most part, contracts that private institutions held with TIAA did not include these types of employees (McCamman, 1948).[47] Of the smaller share that were not covered through TIAA contracts, some were likely affiliated with religious organizations; it is worth noting that Church pension funds frequently covered all employees (Murray and Smith, 1940).

A.4 The relationship between late-life neighborhood quality and lifetime earnings

To document the statistical relationship between late-life neighborhood quality and measures of lifetime earnings, we leverage data from the Health and Retirement Study, a nationally representative longitudinal survey of Americans aged 50 and older that has taken place every two years since 1992.[48] More specifically, we combine publicly available data on HRS respondents with restricted-access information on: (1) their ZIP code of residence at the time of the surveys (Health and Retirement Study, 2023), and (2) estimates of their lifetime earnings based on linkages to administrative earnings records from the SSA (Health and Retirement Study, 2021).

The sample construction proceeds in two steps. Starting with the HRS Cross-Wave Tracker File covering all waves between 1992 and 2018 (Health and Retirement Study, 2024), we first identify respondents' last known spouse (if available), and merge in estimates of lifetime earnings for respondents and their spouses. This allows us to generate measures of lifetime earnings at both the individual and household level. Household lifetime earnings is likely a better measure for women, many of whom had low lifetime labor force attachment. Lifetime earnings are based on administrative earnings records provided by the SSA, which in combination with imputation and projection methods are used to construct estimates of cumulative lifetime earnings through age 65.[49] In a second step, we merge in wave-level ZIP code of residence information. This allows us to observe the quality of respondents' last known ZIP code of residence, as measured by mean AGI in 2001, median house values in 2000, or our neighborhood quality index. The final sample contains around 26,000 HRS respondents aged 65+ at the time of their last interview, around 56% of whom are women.

Appendix Figure A.5 plots moments of the lifetime earnings distribution (in 2016 dollars) by ZIP-level income percentile rank bins. More specifically, HRS respondents are first assigned a percentile rank depending on where their last known ZIP code of residence falls in the national distribution of ZIP codes according to mean AGI in 2001. The 100 ZIP percentile ranks are then aggregated into 50 bins each containing two adjacent percentile ranks. For each of those 50 bins,

[47] Similar to the other types of retirement systems described here, at the time of retirement, TIAA annuity contracts offered the choice between an annuity that ceased at the retiree's death and one that made smaller payments that continued until his widow's death, if the retiree died first.

[48] The HRS (Health and Retirement Study) is sponsored by the National Institute on Aging (grant number NIA U01AG009740) and is conducted by the University of Michigan.

[49] For example, if a respondent only has matched SSA earnings records through age 62, those earnings are projected forward to generate lifetime earnings through age 65. For respondents without matched SSA earnings records, their cumulative earnings at HRS entry and subsequent annual earnings are imputed based on observable characteristics using a nearest neighbor strategy. For more details, see Fang (2021).

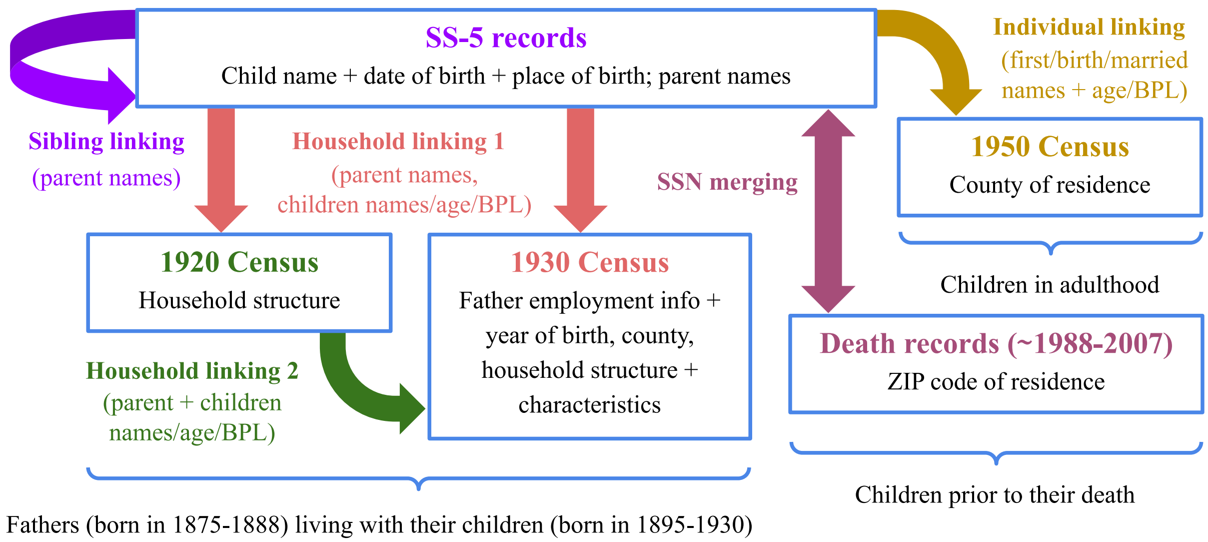
the graphs plot the 10th, 25th, 50th, 75th, and 90th percentiles of the lifetime earnings distribution through age 65 of HRS respondents belonging to those bins. In panel (a), the sample is male HRS respondents and their own lifetime earnings are used. In panel (b), the sample is female HRS respondents whose spouses were identified in the HRS and their household lifetime earnings are used.

Both graphs reveal a strong, positive relationship between the average income in respondents' last known ZIP code of residence and their lifetime earnings. The percentiles imply that the distribution of lifetime earnings "shifts" to the right as late-life ZIP income increases. For example, median own lifetime earnings for men in the bottom ZIP income bin is 707,000 dollars, rises to 1.975 million dollars in the 25th bin (which includes the median), and reaches 3.689 million dollars in the top bin. For women, the corresponding numbers for median household lifetime earnings are 968,000 dollars, 2.76 million dollars, and 4.733 million dollars. The patterns are similar for other percentiles of the lifetime earnings distribution. The differences are generally greater in absolute terms for higher percentiles but greater in relative terms for lower percentiles.

Appendix Table A.10 regresses different measures of lifetime earnings on different measures of late-life ZIP quality (in percentile ranks or in logs), separately for men in panel (a) and for women in panel (b). Consistent with the patterns in Appendix Figure A.5, the coefficients are all positive and statistically significant, regardless of how the relationship is specified. In terms of magnitude, a one percentile rank increase in ZIP-level mean AGI is associated with a 0.35 percentile rank increase in the percentile rank of own lifetime earnings for men. Ranking ZIP codes according to median house values in 2000 or according to the index of neighborhood quality described in Section 5.1 produces similar coefficients (0.25 and 0.28 respectively). Alternatively, the estimates from the log-log specification in column 2 of panel (a) imply that a one percent increase in ZIP-level mean AGI is associated with a 0.62 percent increase in own lifetime earnings for men, while a one percent increase in median house values is associated with a 0.3 percent increase. The estimates in columns 3 and 4 of panel (a) reveal very similar results if we replace own lifetime earnings with household lifetime earnings for men. The patterns for women in panel (b) are very similar to the patterns for men, except that the relationship between late-life ZIP quality and own lifetime earnings is somewhat weaker, as expected. Even so, the relationships between household lifetime earnings and late-life ZIP quality almost mirror the corresponding results for men.

Appendix Figures and Tables

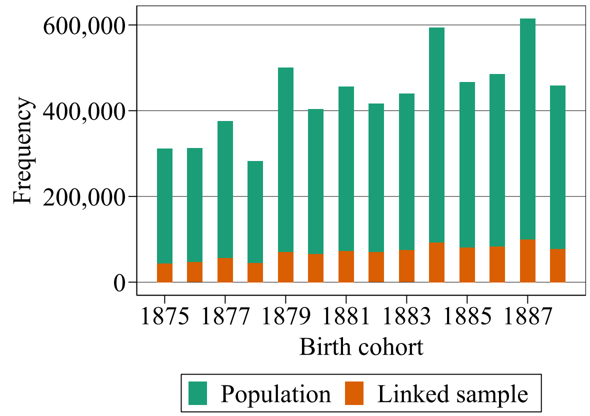
Figure A.1: Overview of Numident death-census linked sample construction



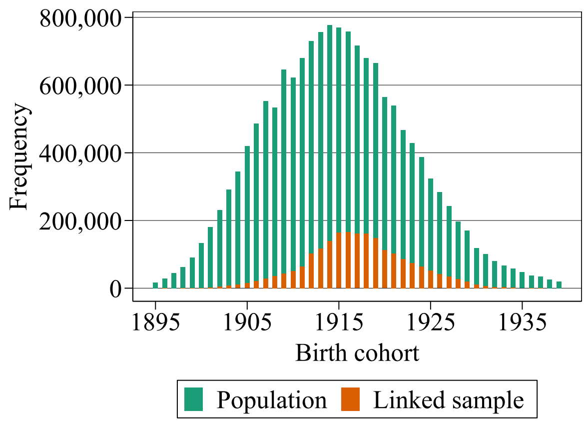
Notes: The key variables from each dataset are indicated in the boxes. The variables used for each type of linkage are indicated in parenthesis.

Figure A.2: Coverage of men born in 1875–1888 and their children born in 1895–1939

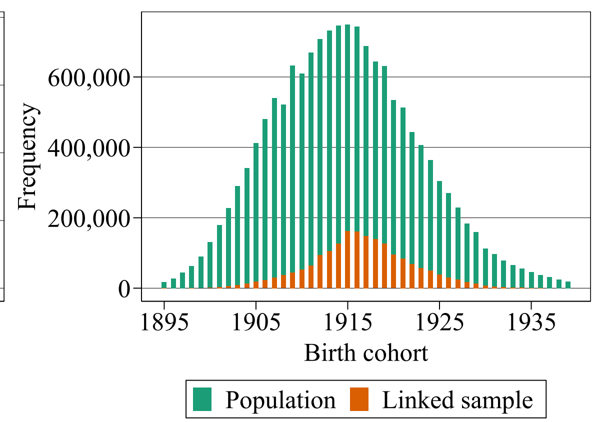
(a) Men born in 1875–1888, covered by Social Security in 1935 or 1950 based on their 1930 employment information



(b) Sons of men born in 1875–1888

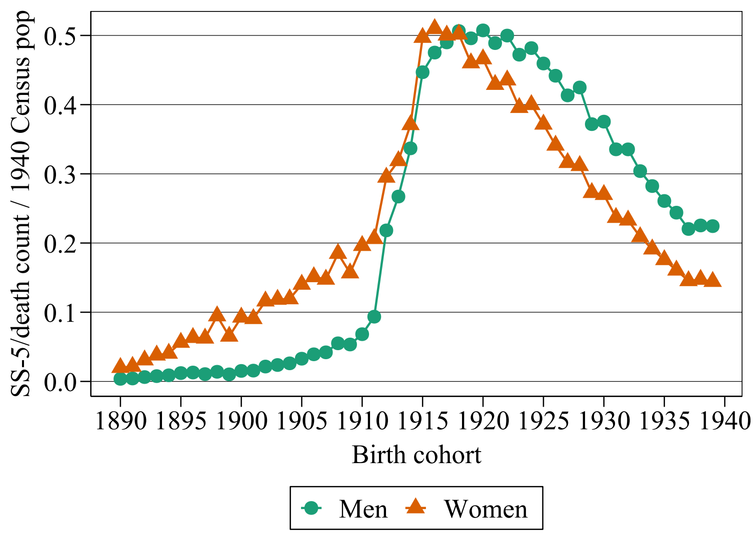


(c) Daughters of men born in 1875–1888



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 information in 1930. Linked sample in panel (a): fathers in linked sample with analogous restrictions. Population in panels (b) and (c): sons/daughters born in 1895–1909/1910–1919/1920–1929/1930–1939 in the 1910/1920/1930/1940 Censuses, whose fathers were born in 1875–1888. Linked sample in panels (b) and (c): sons/daughters born in 1895–1939 in linked sample, whose fathers were born in 1875–1888.

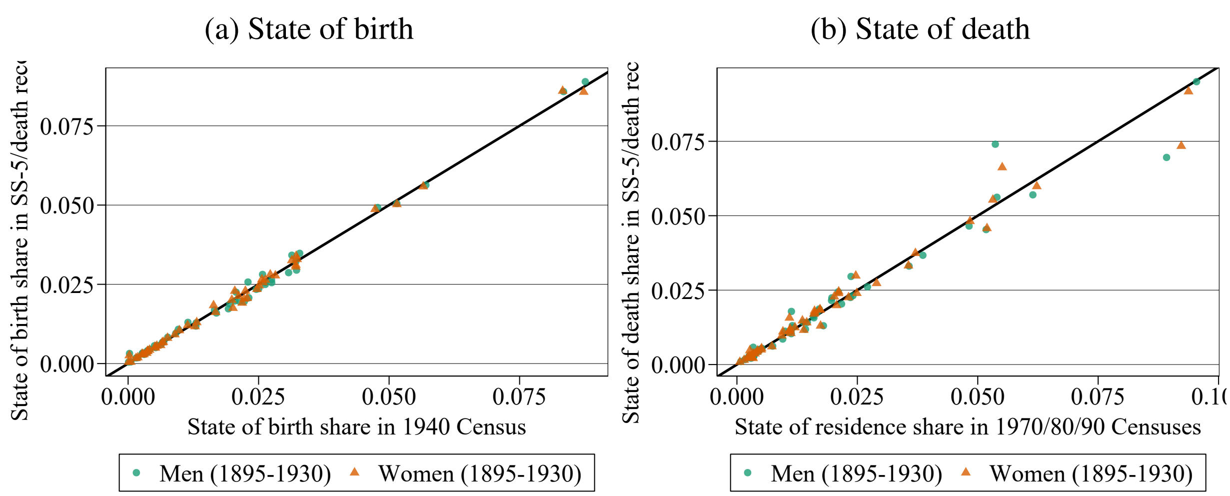
Figure A.3: Cohort coverage of SS-5/death records by sex (1890–1939 cohorts)



Notes: This figure plots the implied cohort coverage rate of SSNs that can be found in both SS-5 and death records with non-missing death ZIP information, defined as the number of SSNs divided by the corresponding number of individuals in the 1940 Census, separately by gender (legend). It therefore captures the share of individuals who survived until 1940 who appear in Numident records.

Source: Numident SS-5 and death records, 1940 US Census.

Figure A.4: State of birth/death coverage of SS-5/death records by sex (1895–1930 cohorts)

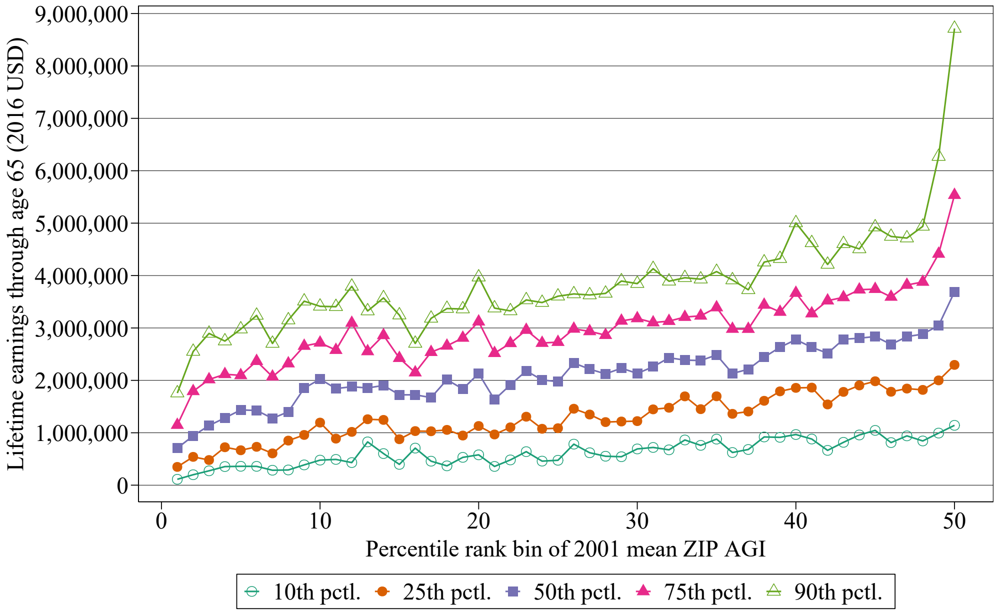


Notes: Panel (a) plots state of birth shares among SSNs that can be found in both SS-5 and death records with non-missing death ZIP information against the corresponding shares in the 1940 Census for individuals born between 1895 and 1930, separately by gender. For the same cohorts and set of SSNs, Panel (b) plots state of death shares in death records against state of residence shares in the 1970 Census (1895–1910 cohorts), 1980 Census (1911–1920 cohorts), and 1990 Census (1921–1930 cohorts), separately by gender.

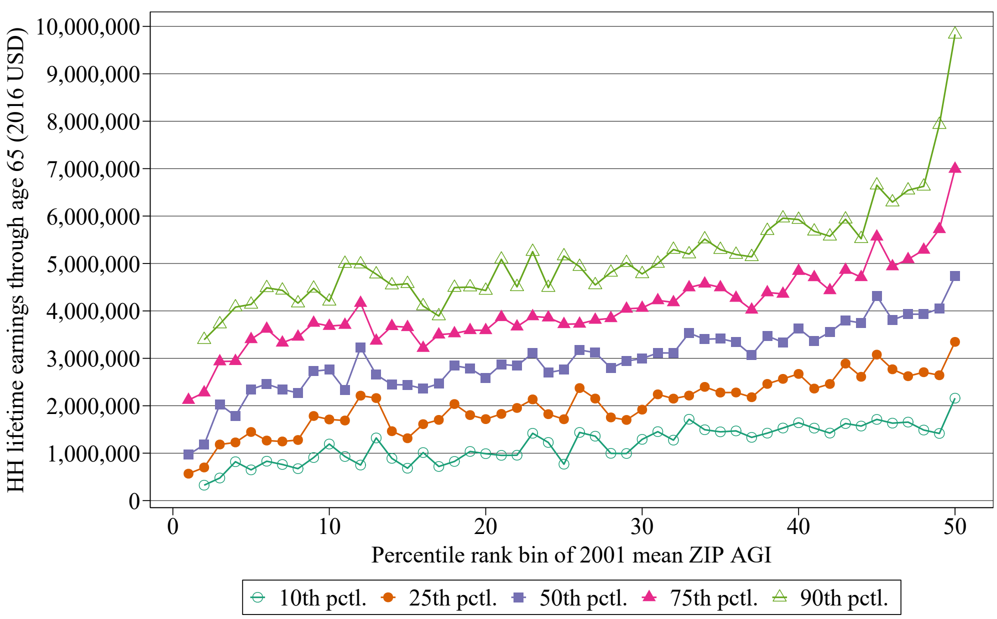
Source: Numident SS-5 and death records, 1940, 1970, 1980 and 1990 US Censuses.

Figure A.5: Moments of lifetime earnings distribution by ZIP-level income percentile rank bin

(a) Men, own lifetime earnings



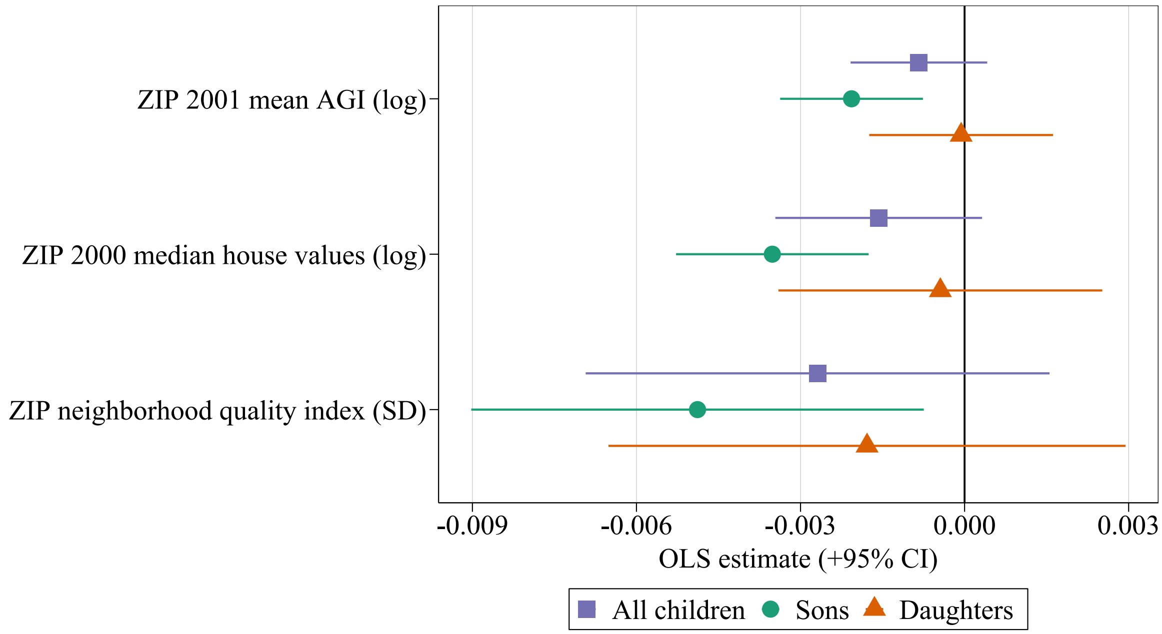
(b) Women, household lifetime earnings



Notes: Panel (a) plots moments of the distribution of own lifetime earnings through age 65 for men, separately by the rank of their last known ZIP code of residence. ZIP codes are ranked nationally according to mean AGI in 2001, and categorized into 50 bins each containing two adjacent percentile ranks (e.g., the 25th bin includes the 49th and 50th percentiles). Panel (b) analogously plots moments of the household lifetime earnings distribution for women with spouses in the HRS.

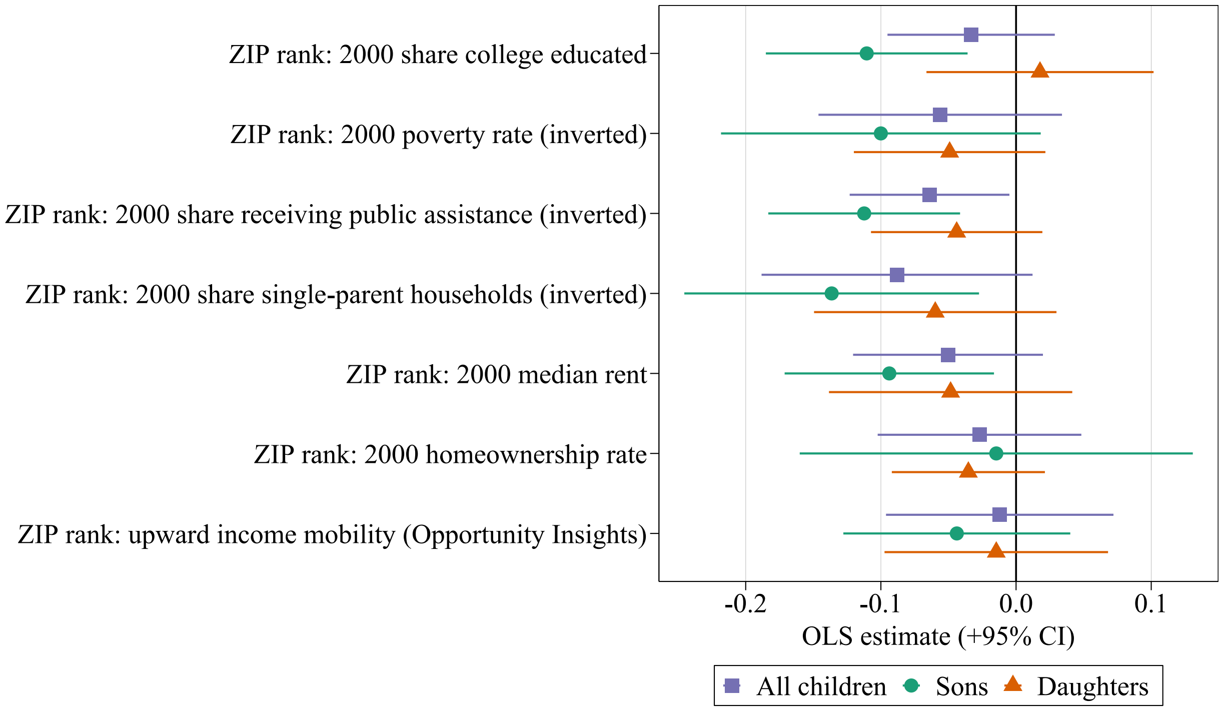
Source: Health and Retirement Study.

Figure A.6: The impact of father's later Social Security coverage on children's death ZIP outcomes (in logs)



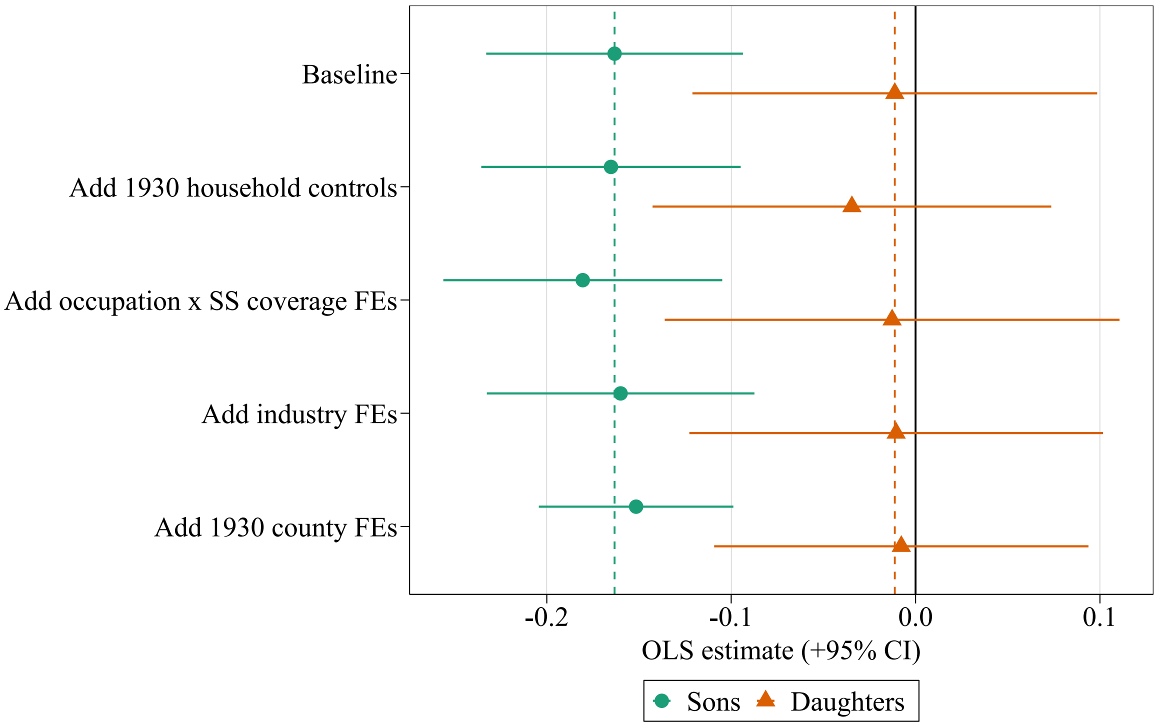
Notes: This figure displays the coefficients on father's minimum age of Social Security eligibility from equation (3), where the outcome is the log of mean AGI or median house values or a neighborhood quality index (normalized to have a mean of 0 and standard deviation of 1) in children's ZIP code at death. Colors/shapes indicate separate estimates for all children, sons, or daughters. All regressions include fixed effects for sex × cohort and father's occupation × year of birth, Social Security coverage category, race, and state/country of birth. Observations weighted using normalized IPWs. 95% confidence intervals based on standard errors clustered at the level of father's industry of employment.

Figure A.7: The impact of father's later Social Security coverage on children's death ZIP rank in terms of 2000 socioeconomic indicators underlying neighborhood quality index (in addition to AGI and house values)



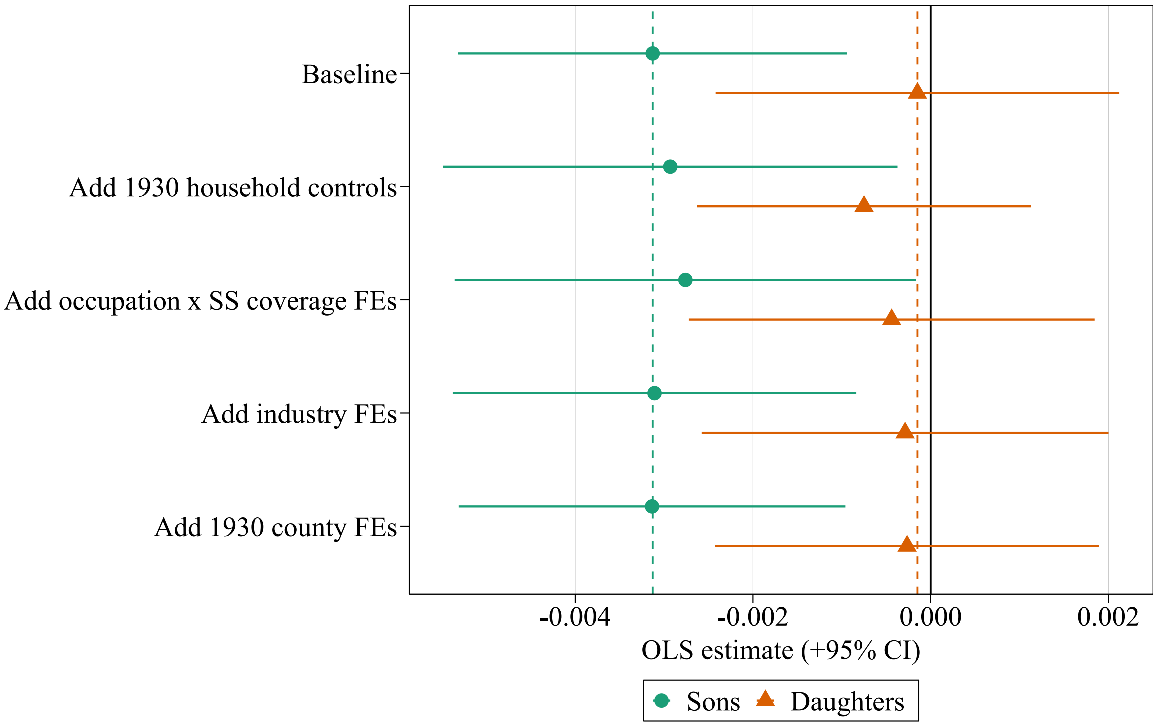
Notes: This figure displays the coefficients on father's minimum age of Social Security eligibility from equation (3), where the outcome is the percentile rank of children's ZIP code at death in terms of various SES indicators. Colors/shapes indicate separate estimates for sons or daughters. All regressions include fixed effects for sex × cohort and father's occupation × year of birth, Social Security coverage category, race, and state/country of birth. Observations weighted using normalized IPWs. 95% confidence intervals based on standard errors clustered at the level of father's industry of employment.

Figure A.8: The impact of father's later Social Security coverage on children's death ZIP rank in terms of 2001 AGI: Robustness checks



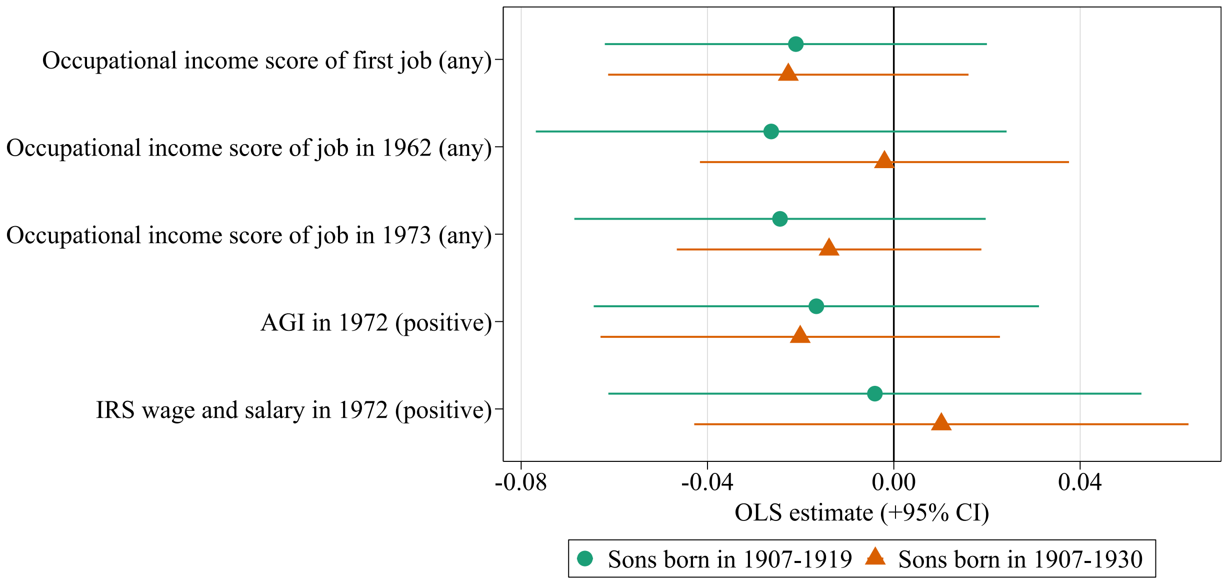
Notes: This figure displays the coefficients on father's minimum age of Social Security eligibility from equation (3), where the outcome is the percentile rank of children's ZIP code at death in terms of mean AGI in 2001. Colors/shapes indicate separate estimates for sons or daughters. All regressions include fixed effects for sex $\times$ cohort and father's occupation $\times$ year of birth, Social Security coverage category, race, and state/country of birth. Observations weighted using normalized IPWs. 95% confidence intervals based on standard errors clustered at the level of father's industry of employment.

Figure A.9: The impact of father's later Social Security coverage on children's propensity to live outside their parents' 1930 census division of residence in 1950: Robustness checks



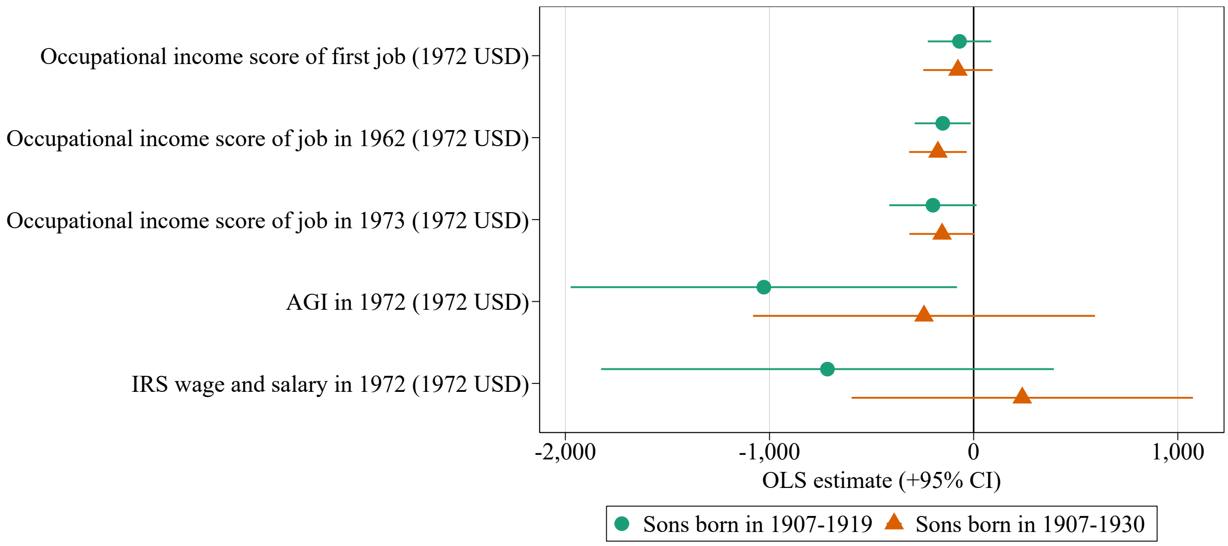
Notes: This figure displays the coefficient on father's minimum age of Social Security eligibility from equation (3), where the outcome is an indicator for living outside parents' 1930 census division of residence in 1950. Colors/shapes indicate separate estimates for sons or daughters. All regressions include fixed effects for sex × cohort and father's occupation × year of birth, Social Security coverage category, race, and state/country of birth. Observations weighted using normalized IPWs. 95% confidence intervals based on standard errors clustered at the level of father's industry of employment.

Figure A.10: The impact of father's later Social Security coverage on sons' probability of having non-missing positive income and a non-missing occupation



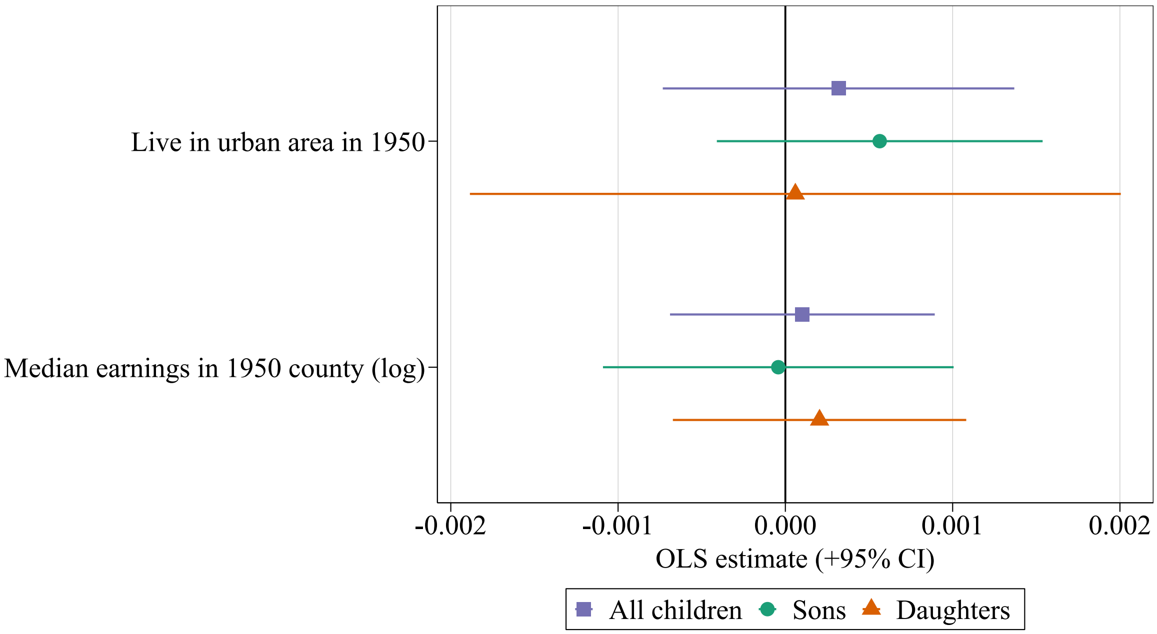
Notes: This figure displays the coefficients on father's minimum age of Social Security eligibility from equation (3), where the outcome is an indicator for a specific outcome being non-missing and positive or simply non-missing. Colors/shapes indicate separate estimates for sons born in 1907–1930 and sons born in 1907–1919. All regressions include fixed effects for father's occupation × year of birth, Social Security coverage category, and son's year of birth, race, and state/country of birth. Observations weighted using OCG sampling weights. 95% confidence intervals based on standard errors clustered at the level of father's industry of employment.

Figure A.11: The impact of father's later Social Security coverage on sons' occupational standing and income (in levels)



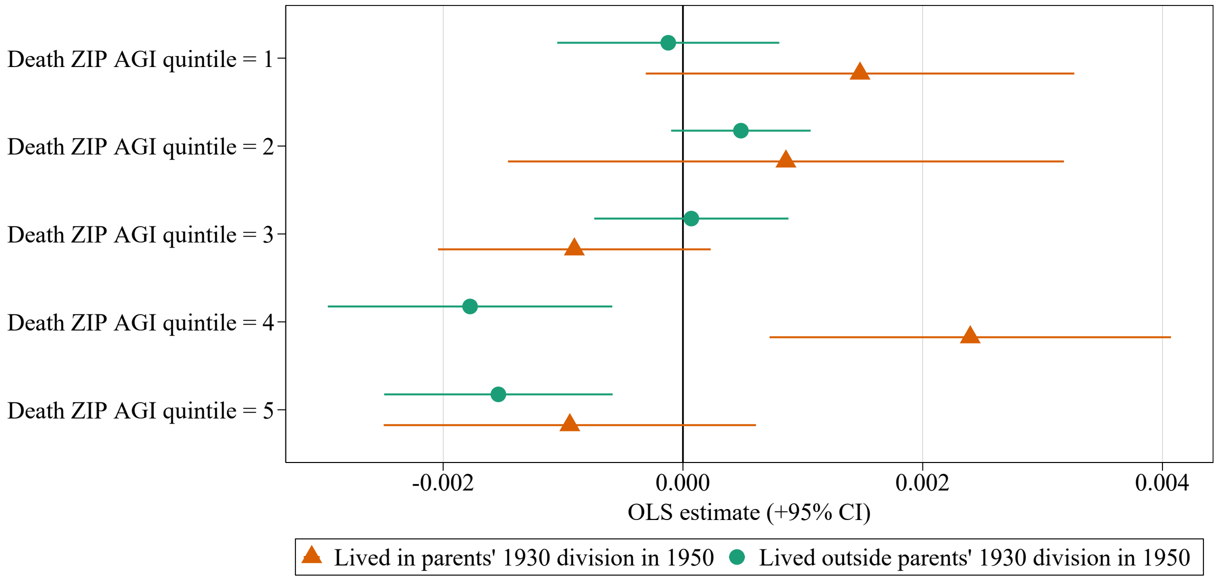
Notes: This figure displays the coefficients on father's minimum age of Social Security eligibility from equation (3), where the outcome is indicated in the row title. Colors/shapes indicate separate estimates for sons born in 1907–1930 and sons born in 1907–1919. All regressions include fixed effects for father's occupation × year of birth, Social Security coverage category, and son's year of birth, race, and state/country of birth. Observations weighted using OCG sampling weights. 95% confidence intervals based on standard errors clustered at the level of father's industry of employment.

Figure A.12: The impact of father's later Social Security coverage on children's location characteristics in 1950



Notes: This figure displays the coefficients on father's minimum age of Social Security eligibility from equation (3), where the outcome is an indicator for living in an urban area in 1950 or median earnings in the 1950 county of residence. Colors/shapes indicate separate estimates for all children, sons or daughters. All regressions include fixed effects for sex × cohort and father's occupation × year of birth, Social Security coverage category, race, and state/country of birth. Observations weighted using normalized IPWs. 95% confidence intervals based on standard errors clustered at the level of father's industry of employment.

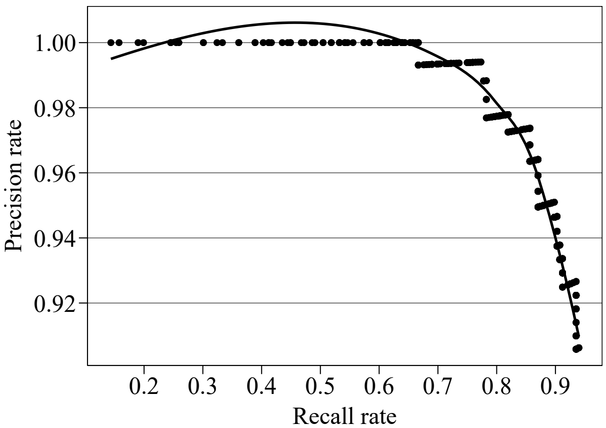
Figure A.13: The impact of father's later Social Security coverage on sons' joint death ZIP rank quintile in terms of 2001 mean AGI and migration status in 1950



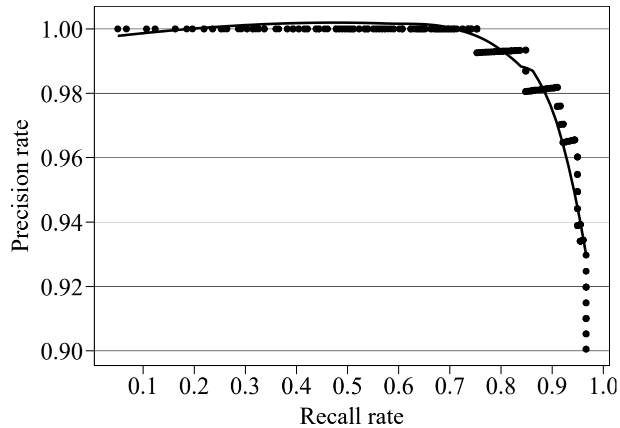
Notes: This figure displays the coefficients on father's minimum age of Social Security eligibility from equation (3), where the outcome is an indicator for the joint outcome of being in a specific death ZIP quintile based on 2001 mean AGI and living in or outside parents' 1930 census division of residence in 1950. All regressions include fixed effects for sex × cohort and father's occupation × year of birth, Social Security coverage category, race, and state/country of birth. Observations weighted using normalized IPWs. 95% confidence intervals based on standard errors clustered at the level of father's industry of employment.

Figure A.14: Model performance: Precision-recall curves

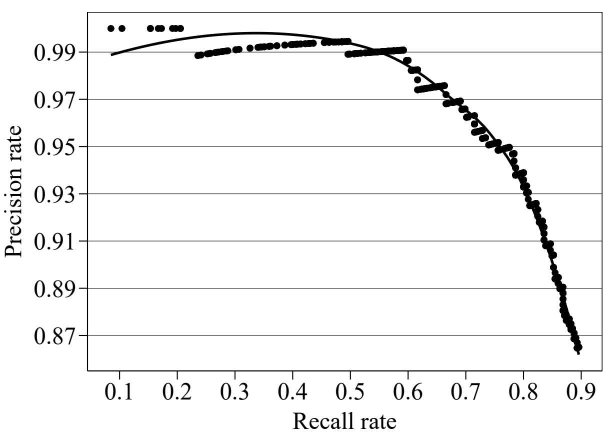
(a) Numident SS-5-1930 Census linking



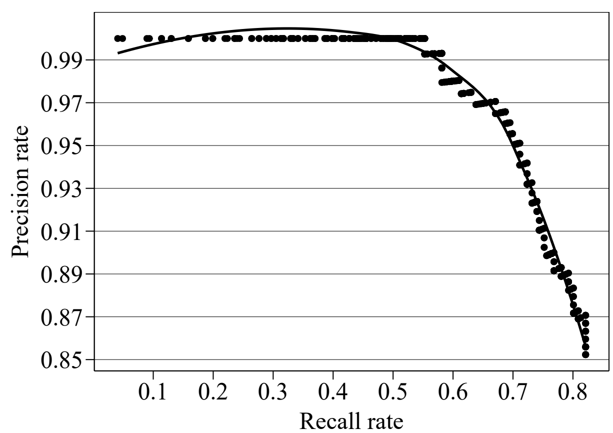
(b) 1930-1920 Census linking



(c) Numident claims-1930 Census linking (primaries with non-missing place of birth)



(d) Numident claims-1930 Census linking (primaries with missing place of birth)



Notes: These figures plot the performance of each model as measured by the cross-validated precision-recall curve in the training set. Each dot corresponds to the precision and recall rate associated with a specific probability threshold.

Table A.1: Representativeness of Numident death-census linked sample

	Mean			Difference (two-sided t-test)	
	Population	Linked sample (unweighted)	Linked sample (weighted)	Linked sample (unweighted)	Linked sample (weighted)
	(1)	(2)	(3)	(4)	(5)
Panel (a): Father characteristics in 1930					
Black	0.079	0.032	0.077	-0.046	-0.001
				[0]	[0.007]
Born in 1875–1879	0.296	0.264	0.296	-0.032	0.001
				[0]	[0.324]
Born in 1880–1884	0.376	0.385	0.375	0.009	0
				[0]	[0.534]
Born in 1885–1888	0.328	0.352	0.328	0.023	0
				[0]	[0.72]
Born abroad	0.3	0.303	0.299	0.003	-0.001
				[0]	[0.144]
Born in Northeast	0.207	0.214	0.207	0.006	0
				[0]	[0.674]
Born in Midwest	0.253	0.269	0.253	0.017	0
				[0]	[0.467]
Born in South	0.216	0.192	0.217	-0.024	0
				[0]	[0.499]
Born in West	0.024	0.022	0.024	-0.002	0
				[0]	[0.485]
Literate	0.94	0.952	0.941	0.012	0.001
				[0]	[0]
Covered by Social Security in 1935	0.862	0.887	0.873	0.025	0.012
				[0]	[0]
Covered by Social Security in 1950	0.138	0.113	0.127	-0.025	-0.012
				[0]	[0]
Farmer	0.01	0.012	0.013	0.002	0.003
				[0]	[0]
Unskilled worker	0.296	0.238	0.268	-0.058	-0.028
				[0]	[0]
Skilled/semi-skilled worker	0.447	0.497	0.48	0.051	0.033
				[0]	[0]
White-collar worker	0.247	0.253	0.239	0.006	-0.008
				[0]	[0]
(table continues on next page)					
N	4,876,245	890,186			

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

Table A.1 (cont.): Representativeness of Numident death-census linked sample

	Mean			Difference (two-sided *t*-test)	
	Population	Linked sample (unweighted)	Linked sample (weighted)	Linked sample (unweighted)	Linked sample (weighted)
	(1)	(2)	(3)	(4)	(5)
Panel (b): Household characteristics in 1930					
Live in Northeast	0.362	0.379	0.374	0.017	0.012
				[0]	[0]
Live in Midwest	0.321	0.34	0.325	0.019	0.004
				[0]	[0]
Live in South	0.201	0.187	0.205	-0.014	0.004
				[0]	[0]
Live in West	0.115	0.093	0.095	-0.022	-0.02
				[0]	[0]
Live on farm	0.072	0.067	0.074	-0.004	0.002
				[0]	[0]
Live in urban area	0.722	0.701	0.719	-0.021	-0.004
				[0]	[0]
Own home	0.48	0.547	0.484	0.067	0.004
				[0]	[0]
Own a radio	0.457	0.495	0.458	0.038	0.001
				[0]	[0.027]
Number of co-resident children	2.028	3.85	3.9	1.823	1.872
				[0]	[0]
N	4,876,245	890,186			

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

Table A.2: Representativeness of Numident claims-census linked sample

	Mean			Difference (two-sided *t*-test)	
	Population	Linked sample (unweighted)	Linked sample (weighted)	Linked sample (unweighted)	Linked sample (weighted)
	(1)	(2)	(3)	(4)	(5)
Panel (a): Own characteristics in 1930					
Black	0.079	0.027	0.08	-0.052	0.001
				[0]	[0.21]
Born in 1875–1879	0.296	0.332	0.288	0.036	-0.008
				[0]	[0]
Born in 1880–1884	0.376	0.388	0.375	0.013	-0.001
				[0]	[0.437]
Born in 1885–1888	0.328	0.236	0.337	-0.092	0.008
				[0]	[0]
Born abroad	0.3	0.261	0.3	-0.038	0
				[0]	[0.807]
Born in Northeast	0.207	0.243	0.207	0.036	0
				[0]	[0.977]
Born in Midwest	0.253	0.311	0.252	0.058	-0.001
				[0]	[0.152]
Born in South	0.216	0.157	0.217	-0.059	0.001
				[0]	[0.308]
Born in West	0.024	0.028	0.024	0.003	0
				[0]	[0.777]
Literate	0.94	0.97	0.94	0.029	0
				[0]	[0.896]
Covered by Social Security in 1935	0.862	0.924	0.902	0.062	0.041
				[0]	[0]
Covered by Social Security in 1950	0.138	0.076	0.098	-0.062	-0.041
				[0]	[0]
Farmer	0.01	0.006	0.009	-0.004	-0.001
				[0]	[0]
Unskilled worker	0.296	0.206	0.258	-0.09	-0.038
				[0]	[0]
Skilled/semi-skilled worker	0.447	0.512	0.494	0.065	0.047
				[0]	[0]
White-collar worker	0.247	0.276	0.239	0.029	-0.008
				[0]	[0]
(table continues on next page)					
N	4,876,245	504,258			

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

Table A.2 (cont.): Representativeness of Numident claims-census linked sample

	Mean			Difference (two-sided *t*-test)	
	Population	Linked sample (unweighted)	Linked sample (weighted)	Linked sample (unweighted)	Linked sample (weighted)
	(1)	(2)	(3)	(4)	(5)
Panel (b): Household characteristics in 1930					
Live in Northeast	0.362	0.37	0.359	0.008	-0.003
				[0]	[0]
Live in Midwest	0.321	0.362	0.331	0.041	0.009
				[0]	[0]
Live in South	0.201	0.143	0.19	-0.058	-0.011
				[0]	[0]
Live in West	0.115	0.124	0.121	0.009	0.005
				[0]	[0]
Live on farm	0.072	0.046	0.073	-0.026	0.002
				[0]	[0.002]
Live in urban area	0.722	0.761	0.72	0.039	-0.002
				[0]	[0.013]
Own home	0.48	0.546	0.48	0.066	0
				[0]	[0.917]
Own a radio	0.457	0.551	0.455	0.094	-0.002
				[0]	[0.038]
Number of co-resident children	2.028	2.074	2.198	0.047	0.17
				[0]	[0]
N	4,876,245	504,258			

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

Table A.3: Responses to 1940 Social Security Number question by Social Security coverage year

	Share of male workers in 1940 (%)	
	Without an SSN	With an SSN
	(1)	(2)
Panel (a): All employment categories, male workers aged 25–64		
Covered in 1935 based on 1940 employment	11.9	88.1
Covered in 1950 or later based on 1940 employment	72.1	27.9
Panel (b): Subset of employment categories used in the analysis, male workers aged 51–64		
Covered in 1935 based on 1940 employment	16.4	83.6
Covered in 1950 based on 1940 employment	73.2	26.8

Notes: Table shows the share of male workers reporting having an SSN vs. not in the 1940 Census (sample-line respondents only), separately for workers whose employment was covered by Social Security in 1935 vs. 1950 or later. Sample restricted to male workers aged 25–64 in 1940 in panel (a). Panel (b) further restricts the sample to the population of interest in the analysis: male workers aged 51–64 whose employment was covered in 1935 or 1950, excluding the self-employed.

Table A.4: Top 20 occupations with some Social Security coverage year variation in 1930

	Summary statistics among men born in 1875–1888 in 1930 Census						
	Population			Death-census linked sample		Claims-census linked sample	
	Number	Share covered in 1935 (%)	Share (%)	Share covered in 1935 (%)	Share (%)	Share covered in 1935 (%)	Share (%)
Occupation	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Laborers (n.e.c.)	815,690	89.6	15.89	90.9	14.77	93	14.54
Operative and kindred workers (n.e.c.)	435,605	99	8.48	99.1	9.18	99.4	9.83
Managers, officials, and proprietors (n.e.c.)	328,324	97.3	6.39	97.1	7.22	98.1	6.46
Carpenters	245,313	98.5	4.78	98.4	5.35	98.6	5.18
Foremen (n.e.c.)	133,092	98.2	2.59	98.2	3.17	99	3.15
Clerical and kindred workers (n.e.c.)	97,906	92.2	1.91	92.1	1.7	95.3	1.98
Painters, construction and maintenance	92,578	97.2	1.8	97.3	1.6	97.4	1.66
Truck and tractor drivers	82,226	95.5	1.6	95.3	1.79	96.6	1.71
Stationary engineers	73,678	92.5	1.44	92.8	1.61	95.4	1.6
Janitors and sextons	72,020	58.6	1.4	51.9	1.47	67.8	1.35
Plumbers and pipe fitters	48,628	96.8	0.95	96.6	0.97	97.7	0.98
Cooks, except private household	43,448	93.3	0.85	93.2	0.36	93.4	0.64
Brickmasons, stonemasons, and tile setters	43,349	99	0.84	99.1	0.89	99.2	0.92
Electricians	42,196	95.4	0.82	95	0.85	97.2	0.93
Mechanics and repairmen (n.e.c.)	35,930	97.2	0.7	97.3	0.75	98.1	0.8
Stationary firemen	34,659	92.9	0.68	93.5	0.77	95.4	0.73
Guards, watchmen, and doorkeepers	35,093	87.7	0.68	88	0.69	91.1	0.63
Teachers (n.e.c.)	34,224	4.5	0.67	3.4	0.85	5.7	0.43
Gardeners, except farm and groundskeepers	30,760	56.5	0.6	57.8	0.55	59.7	0.45
Bookkeepers	30,077	96.2	0.59	96.2	0.57	97.2	0.64

Notes: This table displays summary statistics for the top 20 occupations with some Social Security coverage year variation in the 1930 Census (share covered in 1935 between 1 percent and 99 percent). Sample restricted to men born in 1875–1888 and whose current employment was covered in 1935 or 1950 (excluding the self-employed). Columns 4 and 5 restrict to the subset of men in our Numident death-census linked sample. Columns 6 and 7 restrict to the subset of men of our Numident claims-census linked sample. Individuals in columns 4–7 weighted using IPWs.

Table A.5: Selection of 1930 households and children into our linked sample, by children cohorts

	Dependent variable: Any/number of children linked					
	Children born in 1895–1910		Children born in 1911–1920		Children born in 1921–1930	
	Any	Number	Any	Number	Any	Number
	(1)	(2)	(3)	(4)	(5)	(6)
Panel (a): All children						
Min. age of Social Security eligibility	0.0003	0.0004	-0.0001	-0.0002	-0.0003	-0.0003
	(0.0002)	(0.0002)	(0.0003)	(0.0006)	(0.0003)	(0.0004)
Mean of dep. var.	0.0355	0.0386	0.173	0.2383	0.064	0.0807
Panel (b): Sons						
Min. age of Social Security eligibility	0.0000	-0.0000	-0.0001	-0.0002	-0.0001	-0.0001
	(0.0001)	(0.0001)	(0.0003)	(0.0003)	(0.0002)	(0.0002)
Mean of dep. var.	0.0155	0.0162	0.0999	0.1168	0.0384	0.0435
Panel (c): Daughters						
Min. age of Social Security eligibility	0.0003*	0.0004**	0.0000	0.0000	-0.0002	-0.0002
	(0.0002)	(0.0002)	(0.0002)	(0.0003)	(0.0002)	(0.0002)
Mean of dep. var.	0.0212	0.0224	0.1015	0.1215	0.033	0.0373
N	4,127,539	4,127,539	4,127,539	4,127,539	4,127,539	4,127,539

Notes: Unit of observation is a 1930 household headed by a man born in 1875–1888 and covered by Social Security in 1935 or 1950 based on his employment information, excluding self-employed workers. The dependent variable is either an indicator equal to one if at least one child of a specific sex and cohort range was linked to the 1930 household, or a discrete variable capturing the number of children of a specific sex and cohort range linked to the 1930 household. All regressions include fixed effects for head of household occupation × year of birth, Social Security coverage category, race, and state/country of birth. Robust standard errors in parentheses, clustered at the head of household industry level. *** 1%, ** 5%, * 10% significance.

Table A.6: Children in death-census linked sample: summary statistics

	All children (1895–1930)	Sons (1895–1930)	Daughters (1895–1930)
	(1)	(2)	(3)
Death ZIP percentile rank based on 2001 mean AGI	63.1	63.2	62.9
Death ZIP percentile rank based on 2000 median house values	67.9	68.1	67.7
Percentile rank based on neighborhood quality index	53.3	53.7	52.9
Death ZIP mean 2001 AGI (2001 dollars)	46,934	47,013	46,858
Death ZIP median 2000 house values (2000 dollars)	140,160	140,650	139,691
Number of children	1,478,374	729,312	749,062
Number of families	890,186	558,865	559,309

Notes: This table displays summary statistics for children in Numident death-census linked sample. Observations weighted using normalized IPWs.

Table A.7: Children in death-census linked sample and additionally linked to 1950: summary statistics

	All children (1895–1930) (1)	Sons (1895–1930) (2)	Daughters (1895–1930) (3)
Live outside parents' 1930 county of residence in 1950	0.434	0.426	0.442
Live outside parents' 1930 state of residence in 1950	0.225	0.228	0.222
Live outside parents' 1930 census division of residence in 1950	0.158	0.163	0.153
Live outside parents' 1930 census region of residence in 1950	0.121	0.126	0.116
Co-reside with father in 1950	0.122	0.124	0.121
Co-reside with mother in 1950	0.164	0.162	0.165
Co-reside with any parent in 1950	0.188	0.184	0.193
Married in 1950	0.795	0.796	0.794
Any children in 1950	0.644	0.624	0.666
In labor force in 1950	0.639	0.939	0.324
Occupational income score in 1950 (1949 dollars)	1,592	2,457	684
Number of children	731,751	378,603	353,148
Number of families	547,230	320,139	304,284

Notes: This table displays summary statistics for children in Numident death-census linked sample and additionally linked to 1950. Observations weighted using normalized IPWs.

Table A.8: OCG sample: summary statistics

	Sons (1907–1930) (1)	Sons (1907–1919) (2)
AGI in 1972 (1972 dollars)	14,813	14,157
IRS wage and salary in 1972 (1972 dollars)	12,341	11,298
Occupational income score of first job (1949 dollars)	2,561	2,510
Occupational income score of job in 1962 (1949 dollars)	3,074	3,076
Occupational income score of job in 1973 (1949 dollars)	3,162	3,132
N	3,020	1,739

Notes: This table displays summary statistics in our OCG sample. Observations weighted using OCG sampling weights.

Table A.9: Numident death-census linking: Stage 1 and stage 2 features

Stage 1 set-level features

Type	Description
Continuous	Gap between highest and second-highest JW score
Continuous	Max JW score
Continuous	Mean JW score
Discrete	Absolute difference in number of matching kids (method 1) between highest and second-highest JW score potential
Continuous	Gap between highest and second-highest dad last JW score
Continuous	Max dad last JW score
Continuous	Mean dad last JW score
Discrete	Absolute difference in number of matching kids (method 1) between highest and second-highest dad last JW score potential
Binary	Unique potential with 1+ matching kids (method 1)
Binary	Unique potential with 2+ matching kids (method 1)
Binary	Unique potential with max JW score or max dad last JW score & 1+ matching kids (method 1)
Binary	Unique potential with max JW score or max dad last JW score & 2+ matching kids (method 1)
Binary	Unique potential with max JW score or max dad last JW score & dad or mom middle initial match & 1+ matching kids (method 1)
Binary	Unique potential with max JW score or max dad last JW score & dad or mom middle initial match & 2+ matching kids (method 1)
Binary	Unique potential with 1+ matching kids (method 1) & 1 matching kid (method 2)
Binary	Unique potential with 1+ matching kids (method 1) & 2+ matching kids (method 2)
Binary	Unique potential with max JW score or max dad last JW score & 1+ matching kids (method 1) & 1 matching kid (method 2)
Binary	Unique potential with max JW score or max dad last JW score & 1+ matching kids (method 1) & 2+ matching kids (method 2)
Binary	Unique potential with max JW score or max dad last JW score & dad or mom middle initial match & 1+ matching kids (method 1) & 1 matching kid (method 2)
Binary	Unique potential with max JW score or max dad last JW score & dad or mom middle initial match & 1+ matching kids (method 1) & 2+ matching kids (method 2)
Binary	Unique potential with max JW score or max dad last JW score & 1+ matching kids (method 1) or 1+ matching kids (method 2) & dad or mom middle initial match & 1+ matching kids (method 3)
Binary	Unique potential with max JW score or max dad last JW score & 1+ matching kids (method 1) or 1+ matching kids (method 2) & dad or mom middle initial match & 1+ matching kids (method 3)
Binary	Unique potential with max JW score or max dad last JW score & 1+ matching kids (method 1) or 1+ matching kids (method 2) & dad or mom middle initial match & 1+ matching kids (method 4)
Binary	Unique potential with max JW score or max dad last JW score & 1+ matching kids (method 1) or 1+ matching kids (method 2) & dad or mom middle initial match & 1+ matching kids (method 4)

Stage 2 pair-level features

Type	Description	Type	Description
Continuous	JW score	Binary	Modal race match
Continuous	Dad last JW score	Binary	Dad middle initial match
Discrete	Number of matching kids (method 1)	Binary	Mom middle initial match
Discrete	Number of matching kids (method 2)	Binary	Max JW score or max dad last JW score & 1+ matching kids (method 1)
Discrete	Number of matching kids (method 3)	Binary	Max JW score or max dad last JW score & 1+ matching kids (method 2)
Discrete	Number of matching kids (method 4)	Binary	Max JW score or max dad last JW score & 1+ matching kids (method 3)
Discrete	Rank	Binary	Max JW score or max dad last JW score & 1+ matching kids (method 4)
Binary	Rank = 1		

Notes: "JW score" is the average of three Jaro-Winkler scores comparing primary and potential: (1) father first names, (2) father last names, and (3) mother first names. Two kids match according to method 1 if three conditions hold: (1) the Jaro-Winkler score between their first names is above 0.8 or if the Levenshtein distance between their first names is within 2, (2) their absolute age difference is within 2 years, and (3) their birthplaces match. Method 2 allows for inverted first and middle names, method 3 allows for matches on initials (both first and middle name initials must match), and method 4 is a combination of methods 2 and 3.

Table A.10: Regressing measures of lifetime earnings on measures of late-life ZIP quality

	Lifetime earnings through age 65			
	Own lifetime earnings		Household lifetime earnings	
	Rank-rank	Log-log	Rank-rank	Log-log
	(1)	(2)	(3)	(4)
Panel (a): Men				
2001 mean AGI	0.348***	0.617***	0.369***	0.569***
	(0.01)	(0.024)	(0.011)	(0.021)
2000 median house values	0.249***	0.298***	0.296***	0.31***
	(0.011)	(0.018)	(0.013)	(0.016)
Neighborhood quality index	0.284***	—	0.287***	—
	(0.009)		(0.01)	
Panel (b): Women				
2001 mean AGI	0.151***	0.49***	0.353***	0.534***
	(0.01)	(0.036)	(0.012)	(0.022)
2000 median house values	0.164***	0.365***	0.285***	0.287***
	(0.011)	(0.027)	(0.013)	(0.017)
Neighborhood quality index	0.119***	—	0.276***	—
	(0.008)		(0.01)	

Notes: Each entry corresponds to the estimated slope coefficient from a linear regression of a measure of lifetime earnings through age 65 on a constant and a measure of late-life ZIP quality, where both measures are either specified in percentile ranks or in logs (see row and column titles). The regressions are run separately for men in panel (a) and for women in panel (b). Household lifetime earnings are based on men and women with spouses in the HRS. Rounded sample sizes range from 8,100 to 9,900 for men and from 7,600 to 12,300 for women, depending on the combination. *** 1%, ** 5%, * 10% significance.