# 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

March 26, 2026

**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 observation, 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 is 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, Jessica Van Parys, Heidi Williams, as well as many seminar/conference participants. 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 benefit 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, nor does it tell us about how any changes in family connections impacted children in other ways. To that end, we introduce a novel empirical approach for estimating the causal impact of Social Security on children. 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 only covered 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 late-life socioeconomic status. The 1950 Census allows us to study children's migration patterns by observing their location 10 years after the first monthly Social Security benefits were issued, 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 so 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 better neighborhoods near the end of their lives (around the late 20th century or beginning of the 21st century), as measured by ZIP-level income or home values. 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. Using the 1973 OCG, we also show that sons had better-paying jobs in adulthood, consistent with the interpretation that residence in higher-income ZIP codes late in life reflects higher late-career earnings for sons.

The results suggest that Social Security led to large, long-lasting financial gains for recipients' children. Under our baseline 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—led to full-career average earnings gains for sons of roughly 13–15 percent, with an expected present value of about $140,000–$160,000 (in 2020 dollars). Although there is, of course, uncertainty over these assumptions as well as statistical uncertainty in our estimates, our results strongly suggest that the total net financial gain of the average Social Security recipient's children was quite large, and potentially significantly larger than the corresponding difference in Social Security benefits.

A variety of evidence suggests that the key driver was migration to better-matched labor markets, consistent with Social Security displacing location-dependent family support. 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. We find no evidence of such effects for daughters, consistent with the absence of late-life neighborhood quality gains. Second, several features of the migration response point to productive sorting: Sons moved to locations offering higher education-specific earnings premiums for their education level and were less likely to move after 1950, suggesting that early movers found better matches and stayed put. Third, migration to better-matched labor markets offers a natural explanation for sons' late-career occupational standing gains and can plausibly account for the timing and magnitude of the gains given the high returns to migration in this time period (Ward, 2022). Finally, both late-life neighborhood gains and migration effects were especially large for the sons of Black fathers, consistent with Black workers facing severe location-specific labor market barriers at the time (e.g., Collins and Wanamaker, 2014; Boustan, 2016). In contrast, alternative explanations, such as the possibility that parents passed down some of their Social Security benefits to their children, 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 for sons versus daughters.

Our paper makes several contributions to the literature. Social Security's large statutory redistribution to "initial old" cohorts (Moffitt, 1984; Leimer, 1994; Geanakoplos et al., 1998; Feldstein and Liebman, 2002) is the starting point of an important literature on the theoretical and quantitative effects of pay-as-you-go programs in overlapping generations economies (e.g., Samuelson, 1958; Diamond, 1965; Feldstein, 1974; Auerbach and Kotlikoff, 1987). A foundational result is that expansions of Social Security and other pay-as-you-go programs confer a windfall gain on

the initial old, often at the considerable expense of subsequent generations due to crowd out of the capital stock. Our results, by contrast, suggest that in terms of economic incidence, the gains from early Social Security expansions were much less concentrated on the initial old. Indeed, our results suggest that Social Security's large net transfers to early recipients led to long-run net financial gains for their children that may have been even larger.

While the theoretical literature on Social Security has long recognized the importance of potential changes in family relationships for understanding the effects of Social Security (Barro, 1974; Becker, 1974; Bernheim and Bagwell, 1988), much of the focus has implicitly been on the special case of frictionless monetary transfers. In contrast, our findings highlight the importance of in-kind transfers that require physical proximity. That Social Security and other old-age transfer programs displaced in-kind transfers within families in the form of shared housing and caregiving has been well-documented (Costa, 1997, 1999; McGarry and Schoeni, 2000; Engelhardt et al., 2005; 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 our novel linkages which allow us to follow children of early recipients over time until their death, our work builds on this literature and provides new evidence on lifetime impacts and the role of location choices.[1]

Beyond the implications for intergenerational incidence, our findings also have implications for the efficiency and welfare effects of Social Security. A key efficiency cost of Social Security is thought to be its negative impact on labor supply, given its large effect on leading recipients to retire earlier (see Coile, 2015, for a review). Our findings suggest, however, that once one accounts for the effects on recipients' children, these major Social Security expansions did not decrease *total* family labor earnings, but rather increased them. So, at least for these early expansions, what is widely thought to be a major efficiency cost is alleviated, and potentially outweighed, by a previously-underemphasized and considerable efficiency benefit. The earnings gains also significantly reduce, and potentially transform, the fiscal cost of the program. Our point estimates suggest these Social Security expansions more than paid for themselves—the additional government outlays on Social Security benefits and the foregone tax revenue from recipients' labor supply reductions were more than offset by increased tax revenue from recipients' children's higher earnings—though the confidence interval ranges from substantial savings to modest costs, so the evidence does not conclusively establish fiscal self-financing.

By introducing a new research design to estimate the causal impact of eligibility for Social

[1] Various studies have documented long-run effects of US safety net programs such as the Mothers' Pension program (Aizer et al., 2016), the Earned Income Tax Credit (Bastian and Michelmore, 2018), Medicaid (Goodman-Bacon, 2021), and Food Stamps (Bailey et al., 2024). One key distinguishing feature of our paper is the channel underlying long-run gains: Whereas the usual channel in these studies is early childhood investments in education and health, in our setting the key channel is a relaxation of location constraints in adulthood.

Security among early recipients, our paper adds to the quasi-experimental literature on Social Security. A fundamental challenge in studying Social Security is its near-universal nature and the fact that the timing of receipt and benefit amounts are functions of endogenous decisions that workers make. Therefore, previous studies have exploited features of the program and policy changes that generate discontinuities to isolate the causal impact of Social Security on various outcomes of interest, such as labor supply, savings behavior, and mortality. This includes the Social Security "Notch" (e.g., Krueger and Pischke, 1992; Engelhardt et al., 2005; Gelber et al., 2016; Mukherjee, 2022), the early eligibility age of 62 (Fitzpatrick and Moore, 2018), the eligibility age of 60 for survivors benefits (Coyne et al., 2024), the introduction of the early eligibility age (Engelhardt et al., 2022), and staggered increases in the full retirement age (Mastrobuoni, 2009; Bryant et al., n.d.). Our new research design, which exploits the nature of the roll-out of Social Security, makes use of significantly larger differences in treatment of otherwise similar individuals. Moreover, studying an era with minimal government old-age support and considerable family old-age support helps shed new light on the role of government support in displacing family support.

Similar themes on the introduction of pension programs and spillover effects on the next generation through family ties have been explored in the context of modern-day developing countries (e.g., Jensen, 2003; Edmonds et al., 2005; Ardington et al., 2009; Huang and Zhang, 2021; Bau, 2021). Separately, a large literature has examined the welfare effects of promoting internal migration in developing countries, predominantly from rural to urban areas (e.g., Munshi and Rosenzweig, 2016; Lagakos et al., 2023; Gai et al., 2025; Bau et al., 2026). In addition to providing a long-run perspective, our findings show that alleviating location constraints can yield large benefits in more developed economies as well, and not only for families living in rural areas.

Overall, our paper provides a new perspective on the introduction and early expansion of Social Security. Our findings suggest that it was not so much a large net transfer to the initial recipients at the expense of future generations as a shift away from family-based old-age support, which enabled significant long-run benefits for recipients' children by facilitating migration to better-matched labor markets. This accords with work by historians such as Gratton (1996), who alludes to the significant costs of a family-based old-age support system and notes that in public opinion surveys, Social Security was as popular with the young as with the old. Our findings may also hold important lessons for other contemporary old-age support policies such as long-term care programs, given that informal caregiving to the elderly by family members remains prevalent in the US today (Barczyk and Kredler, 2018).

## 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.

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]

[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.

[3]See Appendix B.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,

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 provisions 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

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.

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

[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.

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 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 analysis relies on two data sources: the full count US Censuses from 1920, 1930 and 1950 (Ruggles et al., 2024), 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. These records include information on individuals' last known ZIP code of residence, which we use to measure late-life socioeconomic status. Our goal is to link individuals in death records 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, 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 can be linked 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 families to households in the 1930 Census, improving both the quality and the number of links we are able to make. Since some children may have left their parents' household by 1930, we augment our sample by combining analogous links to the 1920 Census and links between households in the 1920 and 1930 Censuses. In order to test our hypothesis that Social Security relaxed location constraints for children of recipients, we also link children in SS-5 records to the 1950 Census, which allows us to observe their location in early-to-mid adulthood.[7]

Appendix Figure A.1 provides an overview of how we construct our linked sample. We start by leveraging linkages between siblings in SS-5 records based on parent names from Bailey et al. (2026), which allows us to reconstruct families in the SS-5 data. We then link these families to households in the 1920 and 1930 Censuses based on parent names and information on children (first and middle name, age, and birthplace). 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 birthplace. Lastly, we link children in SS-5 records to themselves in the 1950 Census based on name, age, and birthplace information. Crucially, because we can observe both married and birth names for women in SS-5 records, we are able to find them in 1950 regardless of their marital status at the time.[8]

[7]The Numident Files have been independently linked to census data using a variety of approaches, including Goldstein et al. (2023), Green (2024), Althoff et al. (2025), Ruggles et al. (2025a), and Bailey et al. (2026).

[8]We can infer women's birth names from their father's last name. For women who were already married when they first filed Form SS-5, the last name they list is their married name in most cases. Women who subsequently got married typically filed another Form SS-5 to update their last name information for administrative purposes.

Each of the linkages described above involves linking records primarily based on names, which is notoriously challenging. 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 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 for each record we want to link 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 select or concluding there is too much ambiguity is not something that can easily be automated using deterministic rules. One valid concern is that the quality of the model links is 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 since 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 an algorithm to mimic human decisions. We use the two-stage model from Bailey et al. (2023). 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). 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 that satisfies the property that probabilities sum to one within each set. These probabilities are then multiplied to obtain the unconditional match probability for every pair of records. Model-based links are determined using a threshold rule. In general, lowering the threshold raises the share of human links that the model is able to reproduce (recall rate), but lowers the share of model links that agree with human links (precision rate), creating a natural trade-off between making more links and fewer false links. Training data allows us to control this trade-off, and in practice we select the threshold associated with a 97 percent precision rate. Finally, using the trained model and the chosen threshold, linking is 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.

[9] Other supervised ML approaches to historical record linking include the IPUMS Linked Representative Samples (Ruggles, 2006), Feigenbaum (2016), the Census Tree (Buckles et al., 2025), and the IPUMS Multigenerational Longitudinal Panel (Ruggles et al., 2025a). A common alternative approach is to use automated linking methods (Ferrie, 1996; Abramitzky et al., 2020, 2021, 2025).

[10] Following the LIFE-M process, trainers first complete a trial module to familiarize themselves with each type of link. Two trainers then make linking decisions for a random sample of records. Cases where they disagree are shown to three additional trainers (mixed in with a random sample of other cases). Only links chosen by the two original trainers or by four out of five trainers are considered links for the purpose of training the model.

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 born between 1895 and 1930 whose fathers were born between 1875 and 1888 and covered by Social Security in 1935 or 1950 based their employment information in 1930, excluding self-employed workers.[11] The resulting linked sample contains around 1.6 million children belonging to around 860,000 unique families. Appendix Figure A.2 plots the cohort distribution of fathers and children in the population and in our linked sample. The implied coverage rate of fathers in our population of interest is 22 percent while the coverage rate of their children born by 1930 is 18 percent.[12] We are also able to link around 600,000 of children in our main sample to the 1950 Census. Importantly, our dataset covers sons and daughters at similar rates. This is because men and women have similar coverage in SS-5 records, and because household linking is gender-neutral while the availability of married names enables us to follow women over time.[13]

Two key factors contribute to the incomplete coverage of our death-census linked sample. First, only a subset of children appear in the public version of the Numident released by NARA, which only covers individuals who ever applied for a Social Security Number, 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, 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 from those families appear in our sample. Appendix Table A.1 assesses the representativeness of our linked sample in terms of 1930 household characteristics. Overall, our sample is fairly representative of our population of interest, though as is common in historical record linking, Black and lower-socioeconomic status families are underrepresented as they tend to have more common names. To account for these differences, we re-weight families in our linked sample using inverse propensity score weights generated using the procedure described in Bailey et al. (2020).[14] The resulting weighted means in column 3 are very close to the population

[11] We focus on fathers born in 1875–1888 for two reasons. First, including men born earlier would lower our sample's coverage since their children are more likely to have left the household by 1920. Including later cohorts would have a similar effect since their children are more likely to be born after 1930. Second, as will be evident in Section 4, the variation in Social Security coverage we leverage is most salient for the 1875–1888 birth cohorts.

[12] We cannot compute the true coverage rate for children since our population of interest conditions on fathers' employment status in 1930 (and not all children live with their fathers in 1930). Instead, 18 percent is the coverage rate of children whose fathers were born in 1875–1888 regardless of their Social Security coverage status, which is based on the number of children aged under 10 (or 15) with fathers from those cohorts in the 1910–1930 Censuses.

[13] 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.

[14] The weighting variables are fixed effects for father year of birth, race, place of birth, and literacy status, as well as

means in column 1, even for characteristics we did not explicitly target.

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 is 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 into our linked sample matters for internal validity only to the extent that it is correlated with the Social Security policy variation we exploit. In Section 4.4, we show that this is not the case.

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 will use this sample to test whether fathers' predicted eligibility for Social Security based on their 1930 employment information is correlated with their subsequent claiming behavior, both in terms of the probability of ever claiming and the timing of claiming. These linkages are made based on name, age, birthplace, and place of residence information using the same supervised ML approach described above (see Appendix A.4 for more details). Ultimately, we are able to link around 360,000 men covered by Social Security in 1935 or 1950 at a 97 percent precision rate, or around 10 percent of the population. This lower coverage rate reflects both the coverage of claims records and the the more limited information we can use for linking. We re-weight this sample using the same procedure used to re-weight our death-census linked sample (see Appendix Table A.2 for summary statistics).[15]

### 3.2 1973 Survey of Occupational Changes in a Generation

While our core analysis is based on the Numident death-census linked sample we just described, we also present complementary evidence based on the 1973 Survey of Occupational Changes in a Generation. The 1973 OCG was a supplement to the 1973 Current Population Survey (CPS) that focused on civilian males aged 20–65 and collected detailed information on respondents themselves but also their parents and spouses (Blau et al., 1994). Crucially, it contains retrospective information on respondents' father's year of birth as well as their occupation, industry, and self-employment status when respondents were aged 16. This information allows us to predict fathers'

fixed effects for household home ownership, radio ownership, farm status, and urban status.

[15]Using the same methodology, we also linked sons in claims records to the 1920 Census and combined this with links between households in the 1920 and 1930 Censuses to augment our death-census linked sample. This helped expand coverage since 57 percent of SSNs in claims records are found in death records but not SS-5 records.

likely Social Security eligibility age. The OCG has two key advantages relative to our Numident death-census linked sample: (1) it provides direct measures of occupational standing in mid-to-late adulthood for sons (including information on past occupations), and (2) it is a nationally representative sample. The main drawback are small sample sizes: 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, we are left with only 2,819 observations.[16]

# 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.[17] Although some workers surely changed from 1950-covered employment to 1935-covered employment, we will be able to assess the degree to which employment in 1930 is predictive of later claiming of Social Security.

Figure 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

[16] 1907 is the earliest birth cohort in the OCG (aged 65 when surveyed but turning 66 in 1973), while 1930 is chosen as an upper bound for consistency with our Numident death-census linked sample and to measure sons' occupations when they are at least in their 30s. We slightly expanded the set of father cohorts to increase sample sizes.

[17] "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 "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.

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.[18]

### 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 a year when this type of employment was covered by Social Security. Full details are provided in Appendix B.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 constant across birth cohorts, one could net these differences out in a specification of the form

$$y_f = \alpha_{b(f)} + \delta_{g(f)} + \beta \cdot (\text{Social Security exposure})_{b(f)g(f)} + \Lambda' \mathbf{x}_f + \varepsilon_f, \tag{1}$$

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

where $y$ is an outcome for family $f$, $\alpha_{b(f)}$ and $\delta_{g(f)}$ are fixed effects for father's birth cohort $b$ and father's coverage category $g$, and $\mathbf{x}_f$ a vector of controls (fixed effects for father's race and state or country of birth). For expositional purposes, we define Social Security exposure as the additive inverse of the earliest age fathers could receive Social Security benefits based on their employment information in 1930 so that the coefficient of interest, $\beta$, captures the impact of fathers being predicted to be eligible for Social Security benefits one year *earlier*.

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

$$y_f = \alpha_{b(f)o(f)} + \delta_{g(f)} + \beta \cdot (\text{Social Security exposure})_{b(f)g(f)} + \Lambda' \mathbf{x}_f + \varepsilon_f, \tag{2}$$

where, relative to equation (1), we include fixed effects for the interaction of father's birth cohort and father's 1930 3-digit occupation (subscripted by $o$) in order to focus on variation in earliest age of eligibility arising from work in different industries. In this setup, the three coverage categories in $\delta_{g(f)}$ are: covered in 1935, covered in 1950 via domestic work, and covered in 1950 via work in the nonprofit or state and local government sector. Observations are weighted using the inverse propensity score weights from Appendix Table A.1, and standard errors are clustered at the 3-digit industry level. Equation (2) serves as our preferred empirical specification when analyzing family or parental outcomes. When analyzing the impact of father's Social Security eligibility on

children's outcomes, we estimate a variant of equation (2)

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

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 $c$). In our main analysis, these fixed effects account for any average differences in late-life neighborhood quality that may result from the uneven cohort coverage of Numident records (Appendix Figure A.3). When analyzing outcomes in the 1973 OCG or in the 1950 Census, they additionally account for age effects since each cohort is observed at a different age.[19]

As described in detail in Appendix Section B.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.[20] 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 B.1 and C. 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.[21] 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

[19] In equation (3), individuals are weighted by our baseline family-level weights normalized by the number of children that appear in our linked sample. This avoids mechanically placing more weight on larger families.

[20] 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. The results are robust to treating all such workers as being covered in 1954.

[21] 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, they 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 miscellaneous types of laborers, managers and clerical workers that appear in many industries, but also carpenters, 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 C, 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 the private sector. 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 record. 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 leads to a 0.34 percentage point increase in the likelihood of being linked to a claim for monthly benefits, relative to a baseline mean of 9.3 percent. Column 2 shows that, conditional on being linked to a claims record, eligibility one year earlier leads to claiming monthly benefits 0.11 years earlier.

In Figure 6, we combine both predictions by estimating equation (2) with dependent variables $\mathbf{1}(\text{linked to claim at age } a)$, for ages $a$ ranging from 65 to 95, and $\mathbf{1}(\text{linked to claim made in year } t)$, for $t$ ranging from 1940 to 1970. The pattern of results is encouragingly consistent with the predictions that underlie our empirical strategy. In particular, panel (a) shows that men we predict to be eligible at earlier ages have a greater likelihood of claiming, with the largest effects at earlier ages (between 65 and 70). Furthermore, since men working only in 1950-covered employment would have started to become eligible starting in 1952, it is consistent with our predictions that those who were eligible earlier had a higher likelihood of filing retirement claims during the 1940s, but a *lower* likelihood of filing claims in years from 1952 through 1954.

Note that we do not formally treat the estimates from Table 1 or Figure 6 as a first stage—and in particular do not use them to scale our reduced-form estimates. Because only a subset of actual claims are linked to the Census, these estimates tend to understate the true difference in probability of claiming. 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' earliest 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.[22]

### 4.5 Placebo checks 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 fact that all the coefficients are close to zero and statistically insignificant confirms 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.

[22] 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 or 1921–1930. As we will show, our main findings are driven by sons so we do not view this as a major concern.

# 5 Results

## 5.1 Children's late-life neighborhood characteristics

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 socioeconomic status using the "quality" of the ZIP code they last resided in prior to their death, which has also been used in other work (Althoff and Reichardt, 2024; Green, 2024). 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 income mobility from Chetty et al. (2026).[23]

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 net resources. Here we provide novel empirical evidence to support this claim 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 survey as well as information on their lifetime earnings from SSA administrative records (see Appendix A.5 for more details). This allows us to examine the relationship between lifetime earnings and late-life ZIP quality. Appendix Figure A.5 provides a visual representation of this relationship by plotting moments of the distribution of lifetime earnings through age 65 (in 2016 dollars) by ZIP-level income percentile rank bin.[24] In panel (a), the sample consists of men and their own lifetime earnings are used. In panel (b), the sample consists of women whose spouses also responded to the HRS and the couple's lifetime earnings are used (since many women from the HRS cohorts had low lifetime labor force attachment).

[23] Data on ZIP-level mean AGI comes from the Internal Revenue Service, while data on other ZIP-level indicators comes from the National Historical Geographic Information System (Schroeder et al., 2025). Upward income mobility is defined as the average income percentile rank among children with parents at the 25th percentile of the parental income distribution, and is provided at the census tract level and mapped into ZIP codes based on population shares. 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 sub-indices, and (3) normalize the resulting composite index.

[24] HRS respondents are first assigned a percentile rank depending on where their last known ZIP code of residence falls within 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.

Both figures 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 lifetime earnings for men in the bottom ZIP income bin is 707,000 dollars, rises to 2 million dollars in the 25th bin (which includes the median ZIP code), and reaches 3.7 million dollars in the top bin. The patterns are qualitatively similar for women and other percentiles of the lifetime earnings distribution. In our sample, the correlation between late-life ZIP income and lifetime earnings is 0.276 for men, while for women the corresponding correlation with household lifetime earnings is 0.288. Overall, this suggests that late-life neighborhood characteristics are a valid proxy for long-term socioeconomic status.

Figure 7 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). Throughout the analysis, we estimate separate models pooling all children or restricting the sample to sons or daughters. As discussed in Section 2.2, sons and daughters differed in the types of support they provided to their parents, and therefore may have been differentially affected by parents' eligibility for Social Security. The positive coefficients imply that children whose parents were more exposed to Social Security—i.e., whose fathers were predicted to be eligible for Social Security at an earlier age—lived in better neighborhoods towards the end of their life.[25] Importantly, we find that the these effects are primarily driven by sons, for whom the coefficients are twice as large in magnitude and highly statistically significant. In contrast, the coefficients for daughters are closer to zero and insignificant at conventional levels.

Table 4 presents the full regression results from Figure 7, including alternative specifications where ZIP-level AGI and home values are expressed in logs and the neighborhood quality index is expressed in levels (instead of percentile ranks). In terms of magnitudes, a ten-year earlier predicted Social Security eligibility for fathers—roughly the difference between being covered in 1935 vs. 1950 for workers born in 1877—leads to a 2-percentile increase in the rank of sons' late-life ZIP code, which is around 3 percent of the in-sample means. Alternatively, the same 10-year difference leads to a 2.4-percent increase in ZIP-level income, a 4.3-percent increase in ZIP-level home values, and a 0.054-standard deviation increase in neighborhood quality.

The results are robust to a wide range of alternative specifications. Appendix Figure A.7 shows that the ZIP income rank results for sons and daughters are robust to using alternative measures of exposure to Social Security (a variant of our baseline measure that assumes that 1950-covered

[25] Appendix Figure A.6 shows results based on ranking ZIP codes according to the other sub-indices underlying the neighborhood quality index. We find that the higher neighborhood quality difference is associated with more-educated and lower-poverty areas, but we see little difference in home ownership rates or rates of upward income mobility.

workers become eligible in 1954 instead of 1950 to account for the liberalization of coverage provisions for state and local government employees over the early 1950s, and a dollar-based measure capturing expected average monthly Social Security benefits between ages 65 and 84) and including additional controls (fixed effects for fathers' 3-digit industry, fixed effects for fathers' 3-digit occupation-by-Social Security coverage category, controls for household characteristics in 1930, and fixed effects for parents' 1930 county of residence). In particular, the fact that our estimates are robust to the inclusion of industry and occupation-by-coverage fixed effects highlights that our results are not driven by average differences between fathers covered in 1935 vs. 1950, but rather by fathers' differential exposure to Social Security based on the interaction of cohort and coverage status. The county fixed effects alleviate any concern that our findings might reflect differences in the characteristics of parents' 1930 locations that are correlated with our Social Security eligibility variation. This rules out differential place-based exposure to the Great Depression, for example.

Finally, Appendix Table A.6 uses alternative samples linking parents to the 1940 Census to provide evidence that the results we see for sons are not due to changes in outcomes for their families over the 1930s, prior to eligibility for Social Security. We find little evidence of differences in wage income, labor force participation, or employment in 1940 for men predicted to be eligible for Social Security earlier. Since education is reported in the 1940 Census, we additionally test for differences in years of education for both fathers and mothers. We find no evidence that mother's education differed across families with different Social Security exposure, and although we find a statistically significant, positive relationship between Social Security exposure and father's education in one sample, it implies that a 10-year difference in predicted eligibility is associated with only a 0.1-year difference in father's education, relative to a mean of 7.4, too small to account for the magnitude of the gains we find for sons; in addition, this result does not appear in the other sample.

### 5.2 Sons' occupational standing in adulthood

The findings in the previous section show that sons whose fathers were predicted to be eligible for Social Security at an earlier age lived in better neighborhoods towards the end of their lives. We now provide complementary evidence from the 1973 OCG survey, and show that sons whose fathers were more exposed to Social Security also had higher-paying jobs in adulthood. This serves two purposes. First, while we have shown that late-life neighborhood quality is positively correlated with lifetime earnings, the OCG provides more direct evidence on sons' economic standing in adulthood. Second, the fact that Social Security had a positive impact on sons' labor market outcomes sheds light on the channels through which Social Security led to long-term gains for sons.

Figure 8 displays the estimates from equation (3), where the dependent variables are the occu-

pational income scores associated with sons' first job, the job they held in 1962, and their job at the time of the survey in 1973 (in logs).[26] While the OCG also contains other income measures, we focus on occupational income scores since occupations are generally more stable than income in a given year, which can fluctuate due to transitory shocks. Occupational income scores are obtained by first mapping OCG occupation codes into 1950 Census occupation codes and using the race-and-region-specific occupational income scores from Collins and Wanamaker (2022). These scores account for the fact that workers in the same occupation can have very different earnings depending on their location and the fact that non-white workers tend to earn less than White workers. We use 1940-based scores for sons' first job, 1960-based scores for their job in 1962, and 1970-based scores for their job in 1973. Since OCG respondents were asked about their father's employment information when they were 16 years old, which might have endogenously responded to the introduction of Social Security in 1935 for fathers of sons born after 1919, we show separate results restricting the sample to sons born in 1907–1919.

The estimates in the top row of Figure 8, which are close to zero and statistically insignificant, imply that fathers' Social Security eligibility did not have a meaningful effect on sons' first job. This likely reflects the fact that for many sons, their first job may have preceded their parents' Social Security receipt. In contrast, the estimates for sons' occupational income score in 1962 and 1973 are larger in magnitude and significant at the 5 or 10 percent level. Although the confidence intervals are wide due to small sample sizes in the OCG, the point estimates imply that a ten-year earlier predicted eligibility for Social Security increased sons' occupational income scores in adulthood by over 30 percent. Table 5 shows that these findings are not sensitive to specifying occupational income scores in levels rather than in logs, and although occupational information is missing for around 10 percent of respondents we find no evidence that this is correlated with our Social Security variation. Appendix Figure A.8 shows that the estimates for sons' occupational standing in 1962 are also robust to alternative ways of measuring fathers' exposure to Social Security and the inclusion of additional controls.

At first glance, the over-30-percent increase in sons' late-career occupational income scores may seem difficult to reconcile with the comparatively modest 2.4 percent increase in ZIP-level income and 4.3 percent increase in ZIP-level home values. However, these magnitudes are not directly comparable: The former is an individual-level measure of late-career earnings potential, whereas the latter are neighborhood-level averages observed near the end of sons' lives. To assess whether the two sets of estimates are quantitatively consistent, we estimate in the HRS the cross-sectional gradients relating the log of individual late-career earnings to the log of characteristics of the respondent's late-life ZIP code (see Appendix D for details). The estimated gradients are

[26] Relative to the empirical setup for the late-life neighborhood analysis, here the race and birthplace fixed effects are based on sons' information and observations are weighted using OCG sampling weights.

about 0.075. Thus, 10-percent higher late-career earnings is associated with only about 0.75-percent higher ZIP-level income, reflecting substantial within-ZIP heterogeneity in resources. A 35-percent earnings effect yields a predicted ZIP income effect of around 2.6 percent, very close to our direct estimate of 2.4 percent. The analogous calculation for ZIP-level home values yields about 2.7 percent, somewhat below our direct estimate of 4.3 percent but within its confidence interval. We therefore view the earnings and neighborhood estimates as quantitatively consistent once individual-level gains are translated into neighborhood-level averages.

### 5.3 Children's migration patterns

As we have emphasized, one potential channel through which Social Security may have benefited children of recipients in the long run is migration: To the extent that Social Security displaced in-kind transfers that require physical proximity, it may have induced some children to move to places where they had better job opportunities than in their parents' location. We now test this hypothesis using the subset of children in our Numident death-census linked sample that we can observe in the 1950 Census, 15 years after the Social Security Act and 10 years after the first monthly benefits were issued.

Figure 9 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) by 1950.[27] 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 care for them. Consistent with the notion that Social Security relaxed location constraints, we find that sons whose fathers were predicted to be eligible for Social Security at an earlier age were more likely to live in a different state or census division by 1950. In terms of magnitude, a ten-year earlier predicted eligibility for Social Security increases sons' likelihood of moving to a different state by 2 percentage points and their likelihood of moving to a different census division by 2.5 percentage points, relative to baseline means of 23 and 16 percentage points, respectively. These findings are robust to various sensitivity checks, as shown in Appendix Figure A.9 for cross-division migration.

Note that the coefficient for moving to a different county is close to zero and statistically insignificant, which means that the cross-state and cross-division migration effects are largely offset by a decline in sons' propensity to move to a different county within the same state, as indicated by the negative coefficient in the second row of Figure 9. In other words, Social Security seems to have prompted some sons to migrate farther away than they otherwise would have.[28] Since we

[27] Since county boundaries have changed over time, we use the crosswalk from Eckert et al. (2020) to map counties to 2,833 county groups with stable boundaries 1930 and 2010.

[28] Table 6 reinforces this point by showing that sons were more likely to move to places over 500 miles away from

can infer children's late-life location from their ZIP code in death records, we can also examine their post-1950 migration patterns. Appendix Table A.7 shows analogous results comparing children's location in 1950 to their location at death, and reveals that sons whose fathers were more exposed to Social Security were *less* likely to move to a different state, census division, or census region after 1950. Therefore, Social Security not only induced sons to migrate farther away than they otherwise would have, but also earlier. The fact that Social Security affected the timing of sons' migration rather than their probability of ever migrating may reflect the fact that location constraints are eventually relaxed once parents pass away.

In contrast to sons, we find no evidence of any significant migration response among daughters. This is consistent with our hypothesis that caregiving duties prevented them from out-migrating to the same extent as their brothers. This is also in line with recent evidence showing that daughters played an important role in supporting their parents in the early 20th century, prior to the introduction of Social Security (Aizer et al., 2025). The absence of migration effects for daughters provides a rationale for the lack of effects on late-life neighborhood quality, reinforcing the notion that migration was a key determinant of the impact of Social Security on children's long-term outcomes.

We provide two additional pieces of evidence to support the fact that migration was central to sons' long-term gains. First, we show that Social Security exposure shifts the joint distribution of early long-distance migration and late-life neighborhood outcomes in a manner consistent with migration being an important driver of the gains. Appendix Figure A.10 examines the impact of Social Security on the joint outcome of moving to a different census division by 1950 or not *and* ending up in a given quintile of the ZIP income distribution late in life. To the extent that Social Security raised the probability of moving and ending up in a high-income ZIP code at the expense of not moving and ending up in a low-income ZIP code, this would be consistent with migration being an important mediating channel. This is exactly what we find: sons whose fathers were more exposed to Social Security were less likely to remain in their parents' census division and end up in the bottom two quintiles of the national ZIP income distribution and more likely to move to a different census division and end up in the top two quintiles of the ZIP income distribution.

Second, we provide suggestive evidence that the way migration led to long-term gains for sons was by facilitating moves to labor markets that were a better match for their skills, consistent with the positive effects on occupational standing in adulthood we have documented. For this analysis we focus on the subset of children in our linked sample for whom we can observe completed years of education, either in the 1950 Census or via additional links to the 1940 Census from Bailey et al. (2026). Because the education question was only asked to sample-line respondents in 1950 (5 percent of the population), and because many children in our cohorts of interest have yet to

their parents' location at the expense of moves within 100 to 500 miles (panel (b), columns 7 and 8).

complete their education by 1940 (younger than 25 years old or enrolled in school), this sample only contains around 200,000 children.[29]

Education serves as a proxy of children's skills, and we measure the demand for those skills in each state in 1950 using education-specific median earnings among male workers aged 25 to 54. Intuitively, if the demand for workers with 12 years of education is high in a given state (perhaps due to its industrial or occupational mix), then that should be reflected in the earnings of the median worker with that education level. In practice, there is substantial variation in median earnings by education across states, and some states can be a good match for some workers but a poor match for others. This is illustrated in Figure 10, which plots earnings premiums by state for workers with 8 vs. 12 years of education. While there are some similarities across both state rankings (e.g., Southern states generally offer the lowest wages), there are some notable differences. For example, Minnesota offers a positive premium for workers with 12 years of education, but a negative premium for workers with 8 years of education. Moreover, in Virginia the negative premium is four times larger for workers with 8 years of education than workers with 12 years of education. These patterns hold for other education levels as well (see Appendix Table A.10).

Table 7 looks at the impact of Social Security on children's direction of migration as measured by the difference in the education-specific earnings premium in their state of residence in 1950 and the corresponding premium in their parents' 1930 state of residence. The estimate in the first column implies that sons whose fathers were predicted to be eligible for Social Security at an earlier age moved to better-matched locations on average. Quantitatively, a ten-year earlier eligibility is associated with a 20-dollar increase in education-specific median earnings premium, relative to a baseline mean of 15.6 dollars (in 1949 dollars). The second and third columns examine other moments of the earnings distribution and reveal a similar positive effect on the 75th percentile but not the 25th percentile, which suggests that sons moved to places with more upside potential. The last three columns show analogous results based on moments of the overall earnings distribution by state. The lack of corresponding effects highlights that moving to a better-matched location does not necessarily entail moving to a place with higher earnings on average. Collectively, our results suggest that Social Security allowed sons to move to places where they had better labor market prospects, which enabled them to enter higher-paying occupations in adulthood and in turn eventually settle in better neighborhoods.

[29] When we observe two different levels of education across 1940 and 1950, we prioritize the 1940 value due to known coding issues with the 1950 Census education variable. In the initial digitization of the files, the character "C" (for "college") was sometime misread as "S" (for "secondary"), and vice versa.

### 5.4 Heterogeneity by family characteristics

How do our key findings vary across types of families and children? Figure 11 shows heterogeneity in the impact of father's Social Security eligibility on children's late-life ZIP income rank and propensity to migrate across census divisions by select characteristics of families and children. These include race, family size, birth order, siblings' sex composition, proxies for family resources, and characteristics of parents' location. These results are obtained by estimating equation (3) separately for different subsets of families or children, which allows the controls to flexibly vary across subpopulations. Note that family size, birth order, and the sex composition of siblings are not directly observable in the data and are approximated using the subset of children we can observe across the 1920 Census, 1930 Census, and Numident records (see Appendix A.6 for details).

We highlight several key takeaways. First, the heterogeneity in migration effects aligns well with the heterogeneity in late-life neighborhood gains: Subgroups with larger migration responses to fathers' Social Security tend to experience larger improvements in late-life neighborhood quality. While some of the individual heterogeneity estimates are imprecise, the overall pattern across subgroups is broadly consistent with the idea that migration to better-matched labor markets was a key channel through which Social Security improved sons' long-run outcomes.

Second, the patterns of heterogeneity across subgroups are broadly consistent with the idea that Social Security displaced certain types of family old-age support typically provided by sons, which in turn relaxed location constraints and facilitated migration. Sons gained considerably more than daughters, consistent with the idea that sons' traditional support roles were more readily substituted by Social Security's cash benefits. The point estimates suggest that sons from smaller families gained more than sons from larger families, consistent with the average son in a smaller family bearing a larger share of family old-age support in the absence of Social Security, leaving more counterfactual support for Social Security to displace. Similarly, the point estimates suggest that sons whose fathers had below-median occupational income scores, and sons in states with below-median Old Age Assistance, experienced larger gains, consistent with Social Security having a greater marginal impact in families with fewer alternative resources. Sons with at least one sister appear to have gained more than sons with no sisters, consistent with the presence of a sister reducing the share of "hands-on" support sons would provide and thereby increasing the share of his support that was easily substituted by Social Security. We emphasize, however, that many of these individual heterogeneity comparisons are imprecise and should be interpreted as suggestive patterns rather than precisely estimated differences.

Third, the estimated effects are especially large (though imprecise) for sons of Black fathers, with correspondingly larger migration responses.[30] This pattern is consistent with the idea that

[30]To increase sample sizes and precision, the results for Black sons and daughters are based on a slightly expanded sample that includes links with a precision rate of 90 percent precision or higher (rather than 97 percent).

the returns to migration were particularly high for Black individuals at this time, given the severe location-specific labor market distortions they faced, including Jim Crow laws, occupational exclusion, and monopsony in agricultural labor markets.

### 5.5 Children's long-run mortality and outcomes in 1950

While we have focused on late-life neighborhood quality as our primary measure of children's long-term socioeconomic status, another natural outcome that we can observe in death records is long-run mortality. However, one challenge is that the public Numident records from NARA mainly cover deaths that occurred between 1988 and 2007. Therefore, we are much more likely to observe age at death conditional on dying in a specific age range that varies across cohorts. This is shown in panel (a) of Appendix Figure A.11, which plots the age at death distribution by cohort in our linked sample. Take the 1900 birth cohort as an example. Since near-universal coverage of deaths begins in 1988 and 2007 is the latest death year, we are more likely to observe individuals from that cohort who died between the ages of 88 and 107. For the 1901 birth cohort, the corresponding age at death range is 87 to 106, and so on. In contrast, Numident death records have good coverage of the entire range of ZIP income ranks, as shown in panel (b) of Appendix Figure A.11. Intuitively, this reflects the fact that age at death is only weakly correlated with late-life neighborhood quality (0.09).

Given this data limitation, we follow Hollingsworth et al. (2024) and examine the impact of Social Security on long-run mortality for a subset of cohorts for which we have good coverage for a common age at death window. More specifically, we focus on the 1923–1930 birth cohorts and deaths at ages 65 to 77, which occurred in a rolling 13-year window between 1988 and 2007. The first column in Appendix Table A.8 first shows that fathers' Social Security eligibility had no impact on the probability of dying at those ages for children that appear in our data. In the second column, we focus on the roughly 60 percent of children from those cohorts in our linked sample who died at those ages and estimate the impact on age at death. We find that Social Security had a positive and significant impact on sons' long-run longevity, with a 10-year earlier eligibility associated with a 0.47-year increase in sons' conditional age at death. Consistent with our other results, we find no significant effects on daughters' long-run mortality. Given the incomplete coverage of Numident death records, it is unclear whether these patterns would hold if we could observe unconditional age at death for a broader set of cohorts, so we view these results as suggestive. Nevertheless, they are consistent with the notion that parents' Social Security eligibility had meaningful effects on sons' long-run socioeconomic status.

So far, we have used links to the 1950 Census to explore the role of migration, but it also enables us to observe several other outcomes of interest. One such outcome is children's propensity to co-reside with their parents, a key proxy for family connections. The first three columns in

Appendix Table A.9 show the impact of fathers' Social Security eligibility on the probability of living with any parent, one's father or one's mother. When pooling all children in panel (a), we find that father's earlier Social Security eligibility reduced the probability of co-residing with parents in 1950, consistent with our hypothesis that Social Security reduced upstream in-kind transfers. The estimates are of similar magnitude when looking at sons and daughters separately, but not statistically significant. Beyond smaller sample sizes, there are two reasons why co-residence effects may be hard to detect. First, here we are measuring co-residence from children's perspective. Since parents typically live with one of their children at any given time, this makes it harder to identify changes in co-residence patterns. Second, this issue is compounded by the fact that many parents from our cohorts of interest are no longer alive by 1950.

The last four columns in Appendix Table A.9 examine children's labor force participation, occupational standing (as measured by median total income by occupation in 1950), marital status and fertility in 1950. We find no evidence of impacts on any of these margins, neither for sons nor for daughters. Of course, this does not rule out the possibility of effects post-1950. Note that the absence of an effect on sons' occupational standing in 1950 is consistent with what we found for their first job in the OCG analysis. This suggests that the occupational gains we detect in 1962 and 1973 may have taken time to materialize, possibly because climbing the job ladder in a better-matched labor market is a gradual process.

# 6 Implications

## 6.1 Social Security led to large financial gains for recipients' children

Social Security led to substantial, long-lasting financial gains for recipients' children. The most direct evidence of children's *net* financial gains comes from the improvement in sons' late-life neighborhood quality. Unlike earnings gains, which do not incorporate the costs that might have been incurred to secure them, neighborhood quality, like other components of consumption, presumably reflects a household's long-term net financial position. As discussed in Section 5.2, however, the estimated improvements in neighborhood quality align well with the estimated late-career earnings gains once one accounts for the cross-sectional relationship between individual earnings and neighborhood-level income. This is consistent with non-earnings-related financial costs and benefits—such as costs of moving, changes in net transfers from their parents, and foregone economies of scale from shared housing—roughly canceling out, so the earnings gains may approximate the net financial gains. However, the standard errors are large enough that the alignment should be viewed as a rough consistency check rather than a precise accounting (see Appendix D).

We estimate sons' late-career earnings gains using three complementary approaches (see Appendix E for details). The first uses the occupational income score estimates together with the

assumption that occupational income score gains translate one-for-one into earnings gains, yielding a late-career gain of roughly 35 percent (SE $\approx$ 17 percent). The second uses the late-life ZIP income estimate together with the cross-sectional relationship between late-life ZIP income and late-career earnings, yielding an implied late-career gain of roughly 32 percent (SE $\approx$ 11 percent). The third uses the late-life ZIP home price estimate together with the cross-sectional relationship between late-life ZIP home prices and late-career earnings, yielding an implied late-career gain of roughly 57 percent (SE $\approx$ 16 percent). The three approaches, which draw on different data and assumptions, give broadly consistent estimates. The occupational income score approach is most direct but least precise, while the late-life ZIP approaches are more precise but require the additional assumption that the treatment affects late-life neighborhood characteristics only through individual earnings. We focus on the more precise ZIP-income-based estimate as our primary specification for most of the calculations that follow.

As a rough, back-of-the-envelope measure of sons' full-career earnings gains, we linearly interpolate between the estimated occupational income score effects at different career stages: no gain at first job or in 1950, and the full late-career gain in 1962 and 1973, for sons of fathers born around 1877.[31] This yields a full-career average earnings gain of roughly 13–15 percent and an expected present value lifetime earnings gain for the average son of roughly \$140,000–\$160,000 in 2020 dollars (see Appendix E).

These gains are strikingly large relative to the associated Social Security benefits. We estimate that 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 \$38,000.[32] Given that the average family had more than two sons, the magnitudes of these estimates suggest that the total net financial gain to recipients' children from these Social Security expansions was on the order of about 9 to 10 times the associated Social Security benefits (95% CI under our more precise, ZIP-based estimate: roughly 3 to 15 times; see Appendix E).[33]

What about daughters? The vast majority of the children of fathers affected by our instrument

[31] This assumes that the baseline earnings path was median annual earnings for male four-quarter workers from the Social Security Bulletin Annual Statistical Supplements (Social Security Administration, 1965, 1975). We assume sons were born in 1910, and we discount the implied earnings and income gains to reflect time discounting (3% annual real interest rate, discounted to 1950) and mortality (1910 cohort life table for males from the SSA).

[32] We simulate lifetime Social Security benefits for individuals in the 1877 birth cohort who earned the median earnings in covered employment, 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. Note that the difference in the expected present value to the parents associated with a ten-year difference in the *predicted* (as opposed to actual) earliest age would be smaller, except under the extreme assumption that differences in the predicted earliest age of eligibility translate one-for-one into differences in actual eligibility (i.e., a "first stage" of one, which is a loose upper bound on the true variation in Social Security coverage associated with our instrument).

[33] The 95-percent confidence interval reflects only sampling uncertainty in the underlying treatment effects and HRS gradient. Additional uncertainty from the occupational-income-score-to-earnings mapping, the interpolation over the career, and other modeling assumptions is not captured. The true uncertainty is therefore larger.

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.

### 6.2 Potential mechanisms

What drove these large gains? The evidence points to migration to better-matched labor markets as the primary mechanism, facilitated by the displacement of in-kind family old-age support that required physical proximity. We develop this argument in four steps. First, we explain why the most obvious alternative mechanism—increased net downstream financial transfers—is inconsistent with our results. Second, we present evidence that Social Security facilitated migration to better-matched labor markets. Third, we show that Social Security-induced migration can account for the magnitude of sons' gains. Fourth, we discuss complementary investments in career capital.

Arguably the most natural way in which Social Security would benefit recipients' children is through increasing net financial transfers from parents to their children. However, this mechanism is unlikely to account for more than a small share of the gains we estimate, for several reasons. First, the gains are too large relative to the associated Social Security benefits. Leading theories of family transfers, such as altruism, predict that both parents and their children would benefit from windfall gains—not that children would benefit by many multiples of the windfall. For example, Mukherjee (2022) finds that parents passed down 15 percent of Social Security benefits using variation from the Notch cohorts. Second, under standard assumptions a pure increase in net downstream transfers would be expected to *decrease* recipients' children's earnings due to income effects on the demand for leisure, yet we find large earnings *increases*. Third, the gains are highly unequal—much greater for sons than daughters, and apparently concentrated among sons who migrated by 1950 (Appendix Figure A.10)—whereas downstream financial transfers tend to be shared roughly equally across children (e.g., Menchik, 1980). Fourth, the heterogeneity patterns across family characteristics are suggestive of being inconsistent with a transfer-based mechanism: The point estimates suggest larger gains for sons whose fathers had fewer resources and for sons with at least one sister, whereas in a transfer model one would generally expect the reverse, though we note that many of these individual comparisons are imprecise.[34]

Social Security increased migration by recipients' sons to different labor markets, and such

[34] Another natural explanation for long-run gains is greater investments in schooling. However, most sons had completed their education by the time their parents received Social Security. Consistent with this, we find no evidence of a positive impact on sons' educational attainment in the 1940 Census and 1973 OCG (Appendix Table A.11). There is a slight negative and significant effect in the 1950 Census, but this may be due to the coding issue with the education variable we highlighted in Section 5.3. We also find no evidence that sons were more likely to be veterans or unionized, suggesting that differential exposure to WWII or unionization are unlikely to be important factors.

migration has been found to lead to large earnings gains in many contexts. We estimate that a ten-year difference in the predicted earliest age of father's Social Security eligibility increased sons' migration by 1950 to a different state by 2 percentage points and to a different census division by 2.5 percentage points (see Table 6). Several features of the migration response suggest productive sorting to better-matched labor markets rather than mere churning. Social Security reduced cross-state and cross-division migration *after* 1950, suggesting that early movers found better labor market matches and stayed put. The near-zero effect on cross-county migration, combined with significant cross-state and cross-division effects, is consistent with sons substituting shorter moves (staying close enough to help parents) for longer ones (to a different state or division), rather than Social Security simply creating moves from nothing. Appendix Figure A.10, on the joint outcome of being in parents' 1930 census division in 1950 and dying in a ZIP code in each ZIP income quintile, reinforces this interpretation, as the pattern is consistent with Social Security enabling some sons to leave labor markets to which they were poorly matched.

The broader migration literature finds that the financial gains from migration are often very large. Modern evidence from both structural and quasi-experimental settings finds that moving costs can be substantial and that induced movers can experience large lifetime gains, consistent with comparative-advantage-based location matching (Kennan and Walker, 2011; Nakamura et al., 2022). In settings closely aligned with ours, the evidence suggests sizable gains from migration in general and from the Great Migration in particular (Collins and Wanamaker, 2014; Boustan, 2016; Ward, 2022).

At first glance, the magnitudes of the migration and earnings effects may appear difficult to reconcile. Our estimates imply that Social Security caused roughly 2.5–6.3 percent of sons to migrate to a farther "migration bin" by 1950.[35] If 5 percent of sons (near the midpoint of the range) had their cross-bin migration affected and each gained $X$ percent in late-career earnings (and other sons gained nothing), the average gain across all sons would be only $0.05 \times X$. A 35-percent average late-career earnings gain would then require a gain per measured migrant of 700 percent—implausibly high. However, three considerations substantially close this gap.

First, our migration measures likely understate the true effects on sons' migration. We rely on coarse migration measures: movements across county, state, division, and region boundaries as of just two time points (1950 and death). Within-bin moves are invisible, as are moves pulled earlier within the same time bin (e.g., a son who moves in 1942 instead of 1948). Given the high

[35]The lower bound of 2.5 percent comes from observing that 2.5pp more sons leave the 1930 division by 1950 (Table 6). The upper bound of 6.3 percent comes from assuming a "cascading" structure of migration effects, where Social Security leads sons to shift their moves "one level up" the move hierarchy. Social Security causes an additional 0.5pp of sons to leave the 1930 county. If all moved to a different county in the same state (the next "level"), then to match the 1.5pp decrease in different-county-same-state moves, Social Security must have caused 2pp of sons to switch from different-county-same-state moves to different-state-same-division moves, etc.

baseline rates of migration—43 percent of sons left their parents' 1930 county, with 22 percent crossing state lines and 15 percent crossing division lines by 1950—and the evidence that Social Security affected even the migration of sons who would otherwise have moved fairly far (e.g., to a different state), Social Security likely affected the migration of considerably more sons than our coarse measures capture. The substitution pattern reinforces this point: The near-zero effect on cross-county migration combined with significant cross-state and cross-division effects implies that Social Security redirected moves that would have happened anyway. Many sons who appear as "non-movers" in our cross-bin analysis may have changed their destination or timing in response to Social Security. If the observed 2.5–6.3 percentage point increases in cross-bin migration are the visible portion of a broader shift in the migration distance distribution, the true share of sons whose migration was meaningfully affected could plausibly be considerably larger.

Second, the roughly 35-percent earnings gains reflect late-career gains, not average gains over the full career. Both theory and evidence suggest that migration is a long-term investment with back-loaded returns. In terms of theory, a long tradition models migration as an investment decision with potentially large returns (Sjaastad, 1962). In terms of evidence, Nakamura et al. (2022) estimate strongly back-loaded earnings gains, with small gains through migrants' twenties, moderate gains in their thirties, growing-but-still-below-average gains until their mid-to-late forties, and finally over 2.5 times the considerable average gain (of 82 percent) at peak earning ages. Our pattern of occupational income score estimates is qualitatively and quantitatively similar. Applying a ratio of late-career to full-career gains of 2.5—somewhat lower than Nakamura et al.'s (2022) estimates imply—to our late-career occupational income score estimates suggests a full-career average gain of roughly 14 percent per son ($\approx 35\%/2.5$)—close to our back-of-the-envelope estimate of 15 percent in Section 6.1. This is consistent with migration and associated investments in career capital paying off gradually as sons climbed job ladders in better-matched labor markets.

Third, returns to Social Security-induced migration were plausibly very high. This period featured large differences in labor market conditions across locations—across regions, urban versus rural, and farm versus non-farm—which created substantial scope for migration gains (Ward, 2022). Consistent with high returns to migration, the estimated effects for sons of Black fathers are substantially larger than for sons of white fathers, with larger migration responses, though these estimates are imprecise. This accords with the significant location-specific labor market barriers faced by Black workers at this time—including Jim Crow laws, occupational exclusion, and monopsony in agricultural labor—and with the large earnings gains documented in the Great Migration literature (Collins and Wanamaker, 2014; Boustan, 2016). Moreover, that Social Security mainly hastened and lengthened migration (more than increasing extensive margin migration) suggests strong latent demand for migration and significant location constraints, both pointing to high potential gains from Social Security-induced migration.

Combining these considerations, the magnitudes are consistent with migration accounting for a large share of sons' gains, depending on the extent to which migration is understated by our coarse measures. For illustration, suppose roughly 20 percent of sons had their migration meaningfully affected and the full-career average gain per son was 14 percent ($\approx 35\%/2.5$). Then a per-mover average full-career gain of roughly 70 percent ($\approx 14\%/0.20$) would fully account for sons' gains. This is below the 82 percent average return estimated by Nakamura et al. (2022), and it seems plausible in our context given the theoretical reasons to expect, and the empirical evidence of, high returns in our context. If 15 percent of sons were affected, the required per-mover gain rises to roughly 93 percent, still within the range of estimates in the literature. Hence, Social Security-induced migration could plausibly account for most or all of sons' gains, though given the imprecision of the underlying estimates, considerable uncertainty remains.

Still, migration was likely not the only channel. Social Security likely facilitated complementary investments in career capital, such as occupational upgrading and accumulation of job-specific skills, that compounded the returns from migration. By reducing the time sons spent providing or coordinating old-age support, Social Security may have freed up time for labor supply and career investments even independent of migration. There are increasing returns to the utilization of human capital (Rosen, 1983), so greater investments in one dimension (such as migration to a better-matched labor market) tend to increase the returns to investments in other dimensions (such as occupational choice and on-the-job training). The pattern of occupational income scores—no effect on first occupation or 1950 occupation, and large effects in 1962 and 1973—is consistent with sons climbing higher-return job ladders over time in better-matched labor markets, rather than simply receiving a one-time wage premium upon arrival. The earnings effects of such complementary investments among migrants would be included in estimates of the return to migration, which could help explain why such estimates are often very large.

Beyond migrants, the relaxation of current and future location constraints may have benefited sons who did not change their migration in response to Social Security. Even for a son who never physically moved, the knowledge that he would not need to live near his parents to provide old-age support—or a reduced probability of "having to" migrate back to care for aging parents—increases the expected returns to investments in career capital by increasing the option value of future migration. If 43 percent of sons migrated to a new county even without Social Security, it seems plausible that a non-negligible share of the remaining 57 percent would at least consider moving under the right circumstances. A Social Security-induced relaxation of location constraints could increase the returns to career investments for such sons, even if their actual migration was unaffected. Although the suggestive evidence that the neighborhood gains are concentrated among those who migrated weighs somewhat against important gains to non-migrants, that evidence is only suggestive and should not be over-weighted.

Our best assessment is that Social Security-induced migration to better-matched labor markets accounted for most of sons' considerable gains, with complementary investments in labor supply—including occupational upgrading and climbing higher-return job ladders—accounting for much of the remainder. The heterogeneity patterns across family characteristics and children are broadly consistent with the idea that Social Security displaced certain types of family old-age support that required physical proximity, thereby relaxing location constraints and facilitating migration. The broad alignment of migration and neighborhood gains across subgroups (Figure 11), the larger effects for sons than daughters, and the suggestive (though imprecise) patterns of greater effects in families with fewer children or lower parental resources, for sons with at least one sister, and for sons of Black fathers all point in the same direction.[36]

### 6.3 Implications for the distributional and welfare effects of Social Security

Our findings—that Social Security enabled large financial gains for recipients' children, primarily by facilitating migration to better-matched labor markets—have important implications for the distributional and welfare effects of Social Security. In terms of intergenerational incidence, our estimates suggest that the financial gain from Social Security to recipients' children was far larger than the gain to recipients themselves—on the order of 9 to 10 times larger (95% CI: roughly 3 to 15 times). This reverses the conventional view that these major Social Security expansions entailed a large redistribution from younger to older cohorts—the oft-discussed "windfall to the initial old." Instead, these expansions largely replaced one type of old-age support (informal, mostly in-kind support within families) with another (formal, cash support through Social Security taxes and benefits), and this replacement enabled large financial gains for recipients' children by facilitating their migration to better-matched labor markets. Moreover, that Social Security led recipients' children to eventually live in better neighborhoods suggests that the intergenerational gains may not have stopped at recipients' children but extended to their grandchildren as well.

In terms of *intra*-generational incidence, our results suggest that by displacing family old-age support, Social Security systematically redistributed from larger to smaller families. This can be viewed either as redistribution across different "types" or as insurance against risk in the costs of providing old-age support from behind the veil of ignorance. Smaller families benefited more for two reasons. First, because many of the costs of family old-age support are relatively fixed and the lowest-cost provider has a higher average cost in smaller families. Second, because larger families have more working-age taxpayers per retirement-age benefit recipient (an effect not captured by

[36] We caution against over-interpreting individual heterogeneity comparisons, since some are imprecise and since predicting heterogeneity in the effects of Social Security requires predicting not the *level* of family old-age support but the *responsiveness* of such support to changes in Social Security, which depends on the share of individuals near the margin of indifference between various support and migration options.

our estimates, which capture differences across children, not families). These theoretical reasons are in keeping with the suggestive, though imprecise, empirical evidence that children from smaller families tended to benefit more.

In terms of efficiency, our findings suggest that Social Security, by displacing family old-age support, reduced the effective migration costs of recipients' children and thereby facilitated a more efficient matching of workers to labor markets. The associated efficiency gain is an important, heretofore overlooked benefit of Social Security. As discussed in Section 6.1, our rough, back-of-the-envelope measure of recipients' children's collective lifetime earnings gains is roughly $340,000–$370,000 in expected present value, or roughly 9 to 10 times our benchmark estimate of the associated difference in the expected present value of Social Security benefits (95% CI under our more precise, ZIP-based estimate: roughly 3 to 15 times; see Appendix E). In other words, each dollar of Social Security benefits led to an increase in recipients' sons' earnings of roughly $9–$10.

Even conservative estimates of the financial gain from Social Security-induced migration exceed plausible levels of other sources of efficiency gains from Social Security that have been emphasized in the literature, including the consumption-smoothing gain from correcting myopic failure to plan for retirement or from imperfections in private annuity markets. In other words, the efficiency gain from Social Security facilitating a more efficient matching of recipients' children to labor markets appears to be a crucial determinant of its overall efficiency effect.

In fact, the resulting gains are so large that Social Security benefits appear to *increase* the total earnings of recipients and their children, despite significantly decreasing the earnings of recipients themselves. A key efficiency cost of Social Security is thought to be its negative impact on labor supply, given its large effect on leading recipients to retire earlier (see Coile, 2015, for a review). We find, however, that once one accounts for the effects on recipients' children, these major Social Security expansions did not decrease total family labor earnings—they increased them substantially. Under reasonable assumptions about recipients' labor supply reductions (see Appendix E), children's earnings gains significantly exceed recipients' earnings losses.[37] Our estimates suggest that, for these specific Social Security expansions, accounting for intergenerational effects reverses the conclusion that Social Security decreases labor earnings, transforming this supposed efficiency cost into an efficiency benefit.

The apparently large gains from migration and related investments in labor supply raise a natural question: Why didn't recipients' children migrate more even without Social Security? If the potential gains were so great, couldn't families have found a way to capture them while ensuring

[37] Under our primary specification, children's earnings gain per dollar of Social Security is estimated to be roughly $9 (95% CI: [$3, $15]), while the assumed recipients' earnings reduction is about $1 per dollar of Social Security, so the reversal holds even at the lower bound of the confidence interval. See Appendix E for details.

elderly members were supported? Appendix F discusses how missing markets and transaction costs may have limited families' ability to do so, and how Social Security may have facilitated the exploitation of otherwise-foregone gains from trade in old-age support. These considerations suggest that Social Security may have smoothed the financial component of old-age support across individuals and over time, in a way that seems likely to have been desirable but hard to achieve otherwise.

Social Security's large positive effect on total family earnings has important implications for the overall cost of the program. Under our point estimates, these Social Security expansions more than paid for themselves: The additional government outlays on Social Security benefits and the foregone tax revenue from recipients' labor supply reductions were more than offset by increased tax revenue from recipients' children's higher earnings. Specifically, under our point estimates the average dollar of additional Social Security benefits had a net cost to the government of roughly *negative* \$0.90—that is, each dollar of additional Social Security benefits *saved* the government about \$0.90.[38] However, the 95-percent confidence interval for the net fiscal cost ranges from roughly −\$2.40 (substantial savings) to roughly \$0.55 (modest cost), so the evidence is consistent with but does not conclusively establish fiscal self-financing.[39] Had it not been for the intergenerational effects on recipients' children, by contrast, our assumption about recipient labor supply responses to Social Security implies that each dollar of additional Social Security benefits would have cost the government about \$1.20. Under our point estimates, *if* recipients and their children benefited from this Social Security expansion, then the negative cost to the government would imply an infinite marginal value of public funds (MVPF) spent on these Social Security expansions, since they would have helped recipients and their families at negative cost to the government. However, in addition to the uncertainty about the government's cost, it seems likely that at least some recipients or their children were harmed by these expansions (e.g., a recipient for whom the reduction in family old-age support outweighed the increase in cash income from Social Security).

The private and fiscal externality gains from induced effects on the migration and earnings of recipients' children may be a crucial determinant of the overall gain from old-age support programs more generally, perhaps especially long-term care programs. Whereas Social Security seems likely to have mainly displaced family old-age support of everyday living expenses, long-term care continues to be provided mostly by family members, including adult children (see, e.g., Barczyk and Kredler, 2018). Coe et al. (2023) find that children whose parents are exposed to higher subsidies for private long-term care insurance are less likely to co-reside or live nearby their parents and are

[38]This assumes a 25 percent reduction in recipients' labor supply after the Social Security eligibility age and an average marginal tax rate on recipients' and children's earnings of 25 percent. See Appendix E for details.

[39]The confidence interval reflects only sampling uncertainty. Fiscal self-financing requires a ratio of children's gains to Social Security benefits of at least about 5; our point estimate of this ratio is roughly 9, but the lower bound of the 95-percent confidence interval is roughly 3.

more likely to work full-time. This is consistent with the idea that the overall efficiency effects of long-term care-related policies may depend crucially on any induced changes in the migration and earnings of the adult children of the direct recipients. Such intergenerational effects may be increasingly important given reductions in fertility, and so average family sizes, in the US and many other countries. Smaller families tend to experience bigger benefits from displacements of family old-age support, so the gains from smoothing old-age support costs via government old-age support may be considerable even today. Relaxing the location and time constraints of family old-age support by expanding government old-age support programs may have similar long-run benefits for recipients' children to the extent that it puts them on a better career trajectory. Key beneficiaries of government long-term care programs likely would be recipients' adult daughters, who provide the bulk of informal care today. While there is a growing body of research on the short-run benefits in terms of labor supply (e.g., Coe et al., 2023), there is little evidence on the longer-term benefits. This is an important topic for future research.

Of course, the extent to which our results generalize to other types of government old-age support or other contexts depends on many factors. Four conditions were especially favorable to large intergenerational effects in our context. First, our context featured minimal government old-age support prior to the Social Security expansions we study. These expansions may have had outsized effects that would not generalize to expansions on top of already-substantial government support. In the modern US, for example, Social Security and Medicare already provide a baseline of old-age support, so further expansions may displace less family support on the margin.

Second, family old-age support was widespread in our context, and importantly, much of it—particularly shared housing and financial support from sons—was of a type readily substitutable with Social Security's cash benefits. In contexts where family support takes forms less substitutable with cash transfers (e.g., hands-on personal care), government cash transfers may displace less family support and thereby generate smaller intergenerational effects through the migration channel.

Third, this was a period of major economic transformation—structural change from agriculture to industry, rapid urbanization, and highly heterogeneous local labor markets—which created substantial scope for gains from migration. The returns to migration may have been unusually high, especially for Black individuals facing severe location-specific discrimination under Jim Crow. To the extent that the geographic specificity of returns to different occupations has declined over time (see, e.g., Kaplan and Schulhofer-Wohl, 2017), the returns to migration may be lower today, which would attenuate the intergenerational effects. On the other hand, the increase in dual-earner households may increase the return to being in thick labor markets (e.g., Costa and Kahn, 2000), with ambiguous effects on the average return to migration.

Fourth, the gender-specific pattern of our results—large effects for sons, minimal effects for

daughters—reflects the historical pattern of gendered support roles in which sons more often provided financial support while daughters more often provided personal care. In contexts with different gender norms or where the relevant form of family support is care-intensive, the gender pattern of effects could differ substantially. As noted above, expansions of government old-age support today, especially of in-kind long-term care, might affect daughters more than sons.

Despite these reasons to think that our estimated effects may be larger than what one would expect in many other contexts, the same fundamental mechanism—that government old-age support can relax location and time constraints that bind recipients' children—is presumably operative wherever family old-age support is location-specific and substitutable with government support. This mechanism is particularly relevant for contemporary policy debates around long-term care, where family-provided care remains the norm.

# 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.

## 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.

**_ , _ , _ , Myera Rashid, and Santiago Pérez**, *Census Linking Project: Version 3.0 [dataset]*, Harvard Dataverse, 2025.

**_ , 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), 94–111.

**Aizer, Anna, Gabrielle Grafton, and Santiago Pérez**, "Daughters as Safety Net? Family Response to Parental Employment Shocks: Evidence from Alcohol Prohibition," NBER Working Paper 33346, National Bureau of Economic Research 2025.

**_ , Shari Eli, Joseph Ferrie, and Adriana Lleras-Muney**, "The Long-Run Impact of Cash Transfers to Poor Families," *American Economic Review*, 2016, *106* (4), 935–971.

**Althoff, Lukas and Hugo Reichardt**, "Jim Crow and Black Economic Progress After Slavery," *The Quarterly Journal of Economics*, 2024, *139* (4), 2279–2330.

**_ , Harriet Brookes Gray, and Hugo Reichardt**, "Mothers, Schools, and the Making of American Human Capital Mobility," Working paper 2025.

**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.

**Auerbach, Alan J. and Laurence J. Kotlikoff**, *Dynamic Fiscal Policy*, Cambridge: Cambridge University Press, 1987.

**Bailey, Martha J., A. R. Shariq Mohammed, and Paul Mohnen**, "The Evolution of U.S. Educational Mobility over the 20th Century and the Role of Public Education," Atlanta Fed Working Paper 2026-1 2026.

**_ , 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-Term-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.

**Bastian, Jacob and Katherine Michelmore**, "The Long-Term Impact of the Earned Income Tax Credit on Children's Education and Employment Outcomes," *Journal of Labor Economics*, 2018, *36* (4), 1127–1163.

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

_ , **Gaurav Khanna, Corinne Low, and Alessandra Voena**, "Traditional Institutions in Modern Times: Dowries as Pensions when Sons Migrate," *Quarterly Journal of Economics*, 2026, *141* (1), 205–262.

**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.

**Bleakley, Hoyt and Joseph Ferrie**, "Shocking Behavior: Random Wealth in Antebellum Georgia and Human Capital Across Generations," *The Quarterly Journal of Economics*, 2016, *131* (3), 1455–1495.

**Boustan, Leah Platt**, *Competition in the Promised Land: Black Migrants in Northern Cities and Labor Markets*, Princeton University Press, 2016.

**Bryan, Gharad, Shyamal Chowdhury, and Ahmed Mushfiq Mobarak**, "Underinvestment in a Profitable Technology: The Case of Seasonal Migration in Bangladesh," *Econometrica*, 2014, *82* (5), 1671–1748.

**Bryant, Victoria L., Jonathan M. Leganza, Dhiren Patki, and Ellen Stuart**, "The Effects of Increasing the Full Retirement Age on Retirement Savings: Evidence from U.S. Tax Data."

**Buckles, Kasey, Adrien Haws, Joseph Price, and Haley E.B. Wilbert**, "Combining family history and machine learning to link historical records: The Census Tree data set," *Explorations in Economic History*, 2025, *98*, 101717.

**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.

**Chetty, Raj, John N. Friedman, Nathaniel Hendren, Maggie R. Jones, and Sonya R. Porter**, "The Opportunity Atlas: Mapping the Childhood Roots of Social Mobility," *American Economic Review*, 2026, *116* (1), 1–51.

**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.

**Coe, Norma B., Gopi Shah Goda, and Courtney Harold Van Houtven**, "Family spillovers and long-term care insurance," *Journal of Health Economics*, 2023, *90.*

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

**Collins, William J. and Marianne H. Wanamaker**, "Selection and Economic Gains in the Great Migration of African Americans: New Evidence from Linked Census Data," *American Economic Journal: Applied Economics*, 2014, *6* (1), 220–252.

**_ and _** , "African American Intergenerational Economic Mobility since 1880," *American Economic Journal: Applied Economics*, 2022, *14* (3), 84–117.

**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.

**_ and Matthew E. Kahn**, "Power Couples: Changes in the Locational Choice of the College Educated, 1940–1990," *The Quarterly Journal of Economics*, 2000, *115* (4), 1287–1315.

**Coyne, David, Itzik Fadlon, Shanthi P. Ramnath, and Patricia K. Tong**, "Household Labor Supply and the Value of Social Security Survivors Benefits," *American Economic Review*, 2024, *114* (5), 1248–1280.

**Diamond, Peter A.**, "National Debt in a Neoclassical Growth Model," *American Economic Review*, 1965, *55* (5), 1126–1150.

**Eckert, Fabian, Andrés Gvirtz, Jack Liang, and Michael Peters**, "A Method to Construct Geographical Crosswalks with an Application to US Counties since 1790," NBER Working Paper 26770, National Bureau of Economic Research 2020.

**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.

**Engelhardt, Gary V., Jonathan Gruber, and Anil Kumar**, "Early Social Security Claiming and

Old-Age Poverty: Evidence from the Introduction of the Social Security Early Eligibility Age," *Journal of Human Resources*, 2022, *57* (4), 1079–1106.

**_ , _ , 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**, "Social Security, Induced Retirement, and Aggregate Capital Accumulation," *Journal of Political Economy*, 1974, *82* (5), 905–926.

_ **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 K. and Lee M. Lockwood**, "Government Old-Age Support and Labor Supply: Evidence from the Old Age Assistance Program," *American Economic Review*, 2018, *108* (8), 2174–2211.

**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.

**Fitzpatrick, Maria and Timothy Moore**, "The mortality effects of retirement: Evidence from Social Security eligibility at age 62," *Journal of Public Economics*, 2018, *157*, 121–137.

**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.

**Gai, Qingen, Naijia Guo, Bingjing Li, Qinghua Shia, and Xiaodong Zhu**, "Rural Pensions, Labor Reallocation, and Aggregate Income: An Empirical and Quantitative Analysis of China," *Econometrica*, 2025, *93* (5), 1663–1696.

**Geanakoplos, John, Olivia S. Mitchell, and Stephen P. Zeldes**, "Social Security's Money's Worth," in Olivia S. Mitchell, Robert J. Myers, and Howard Young, eds., *Prospects for Social Security Reform*, University of Pennsylvania Press, 1998, pp. 79–151.

**Gelber, Alexander, Adam Isen, and Jae Song**, "The Effect of Pension Income on Elderly Earnings: Evidence from Social Security and Full Population Data," Working paper 2016.

**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.

**Goldstein, Joshua R., Monica Alexander, Casey Breen, Andrea Miranda González, Felipe Menares, Maria Osborne, Mallika Snyder, and Ugur Yildirim**, *CenSoc Mortality File: Version 3.0*, Berkeley: University of California, 2023.

**Goodman-Bacon, Andrew**, "The Long-Run Effects of Childhood Insurance Coverage: Medicaid Implementation, Adult Health, and Labor Market Outcomes," *American Economic Review*, 2021, *111* (8), 2550–2593.

**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.

**Green, Allison**, "Networks and Geographic Mobility: Evidence from World War II Navy Ships," Working paper 2024.

**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.

**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.

**Hollingsworth, Alex, Krzysztof Karbownik, Melissa A. Thomasson, and Anthony Wray**, "The Gift of a Lifetime: The Hospital, Modern Medicine, and Mortality," *American Economic Review*, 2024, *114* (7), 2201–2238.

**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.

**Kaplan, Greg and Sam Schulhofer-Wohl**, "Understanding the Long-Run Decline in Interstate Migration," *International Economic Review*, 2017, *58* (1), 57–94.

**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.

**Krueger, Alan B. and Jörn-Steffen Pischke**, "The Effect of Social Security on Labor Supply: A Cohort Analysis of the Notch Generation," *Journal of Labor Economics*, 1992, *10* (4), 412–437.

**Lagakos, David, Ahmed Mushfiq Mobarak, and Michael E. Waugh**, "The Welfare Effects of Encouraging Rural–Urban Migration," *Econometrica*, 2023, *91* (3), 803–837.

**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.

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

**Mastrobuoni, Giovanni**, "Labor supply effects of the recent social security benefit cuts: Empirical estimates using cohort discontinuities," *Journal of Public Economics*, 2009, *93* (11-12), 1224–1233.

**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.

**Meghir, Costas, A Mushfiq Mobarak, Corina Mommaerts, and Melanie Morten**, "Migration and Informal Insurance: Evidence from a Randomized Controlled Trial and a Structural Model," *The Review of Economic Studies*, 2022, *89* (1), 452–480.

**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, 1984, pp. 327–358.

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

**Munshi, Kaivan and Mark Rosenzweig**, "Networks and Misallocation: Insurance, Migration, and the Rural-Urban Wage Gap," *American Economic Review*, 2016, *106* (1), 46–98.

**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.

**Rosen, Sherwin**, "Specialization and Human Capital," *Journal of Labor Economics*, 1983, *1* (1), 43–49.

**Rosenzweig, Mark R. and Kenneth I. Wolpin**, "Intergenerational Support and the Life-Cycle Incomes of Young Men and Their Parents: Human Capital Investments, Coresidence, and Intergenerational Financial Transfers," *Journal of Labor Economics*, 1993, *11* (1), 84–112.

**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.

_ **, Nesile Ozder, Catherine A. Fitch, Matthew Sobek, Julia A. Rivera Drew, J. David Hacker, Jonas Helgertz, Cheyenne Lonobile, Matt A. Nelson, Evan Roberts, and John Robert Warren**, *IPUMS Multigenerational Longitudinal Panel: Version 2.0 [dataset]*, Minneapolis, MN: IPUMS, 2025.

_ **, Sarah Flood, Matthew Sobek, Daniel Backman, Grace Cooper, Julia A. Rivera Drew, Stephanie Richards, Renae Rogers, Jonathan Schroeder, and Kari C.W. Williams**, *IPUMS USA: Version 16.0 [dataset]*, Minneapolis, MN: IPUMS, 2025.

**Samuelson, Paul A.**, "An Exact Consumption-Loan Model of Interest with or without the Social Contrivance of Money," *Journal of Political Economy*, 1958, *66*, 467–467.

**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.

**Schroeder, Jonathan, David Van Riper, Steven Manson, Katherine Knowles, Tracy Kugler, Finn Roberts, and Steven Ruggles**, *IPUMS National Historical Geographic Information Sys-*

*tem: Version 20.0 [dataset]*, Minneapolis, MN: IPUMS, 2025.

**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.

**Sjaastad, Larry A.**, "The Costs and Returns of Human Migration," *Journal of Political Economy*, 1962, *70* (5, Part 2), 80–93.

**Social Security Administration**, "Social Security Bulletin, Annual Statistical Supplement, 1965," 1965. U.S. Government Printing Office.

_ , "Social Security Bulletin, Annual Statistical Supplement, 1975," 1975. U.S. Government Printing Office.

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

**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.

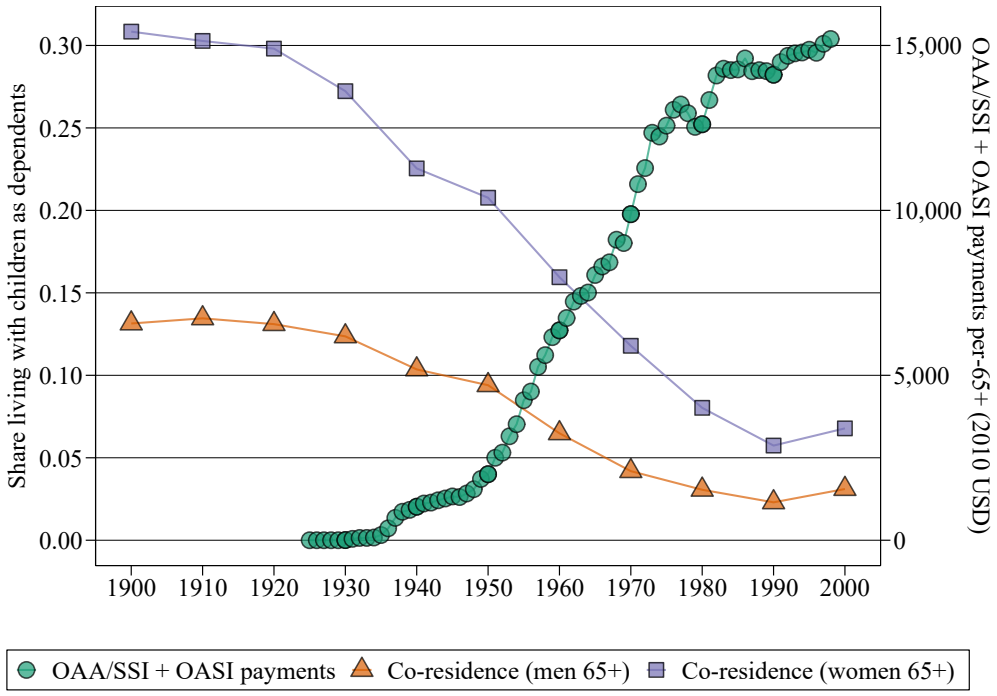
_ , "National Intercensal Tables, 1900–1990: Population by age, sex, and race," https://www.census.gov/data/tables/time-series/demo/popest/pre-1980-national.html 2004. Accessed: 2020-08-10.

**Ward, Zachary**, "Internal Migration, Education, and Intergenerational Mobility: Evidence from

American History," *Journal of Human Resources*, 2022, *57* (6), 1981–2011.

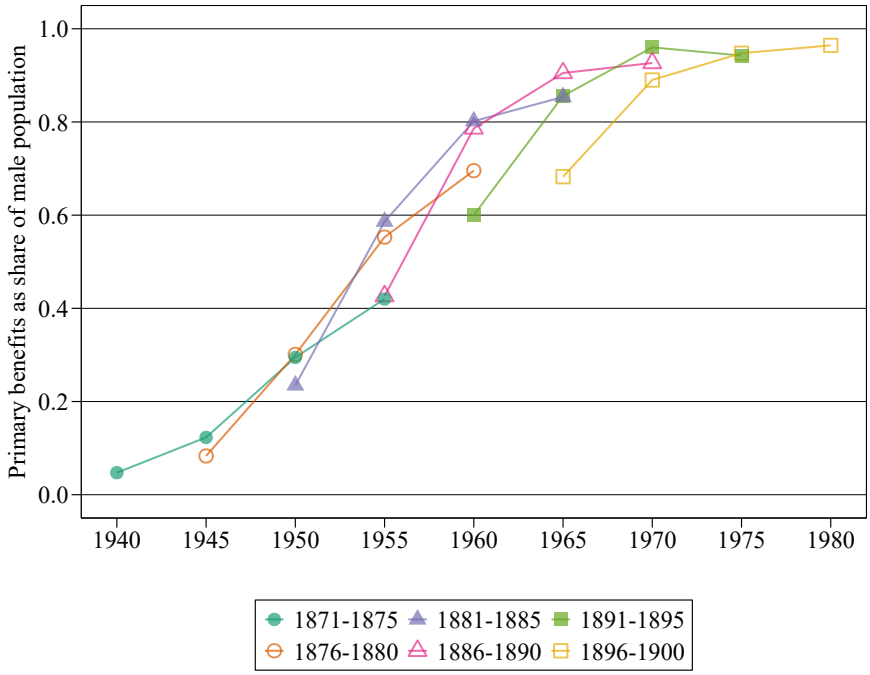
## Figures and Tables

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




*Notes:* This figure 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 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. (2025b).

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

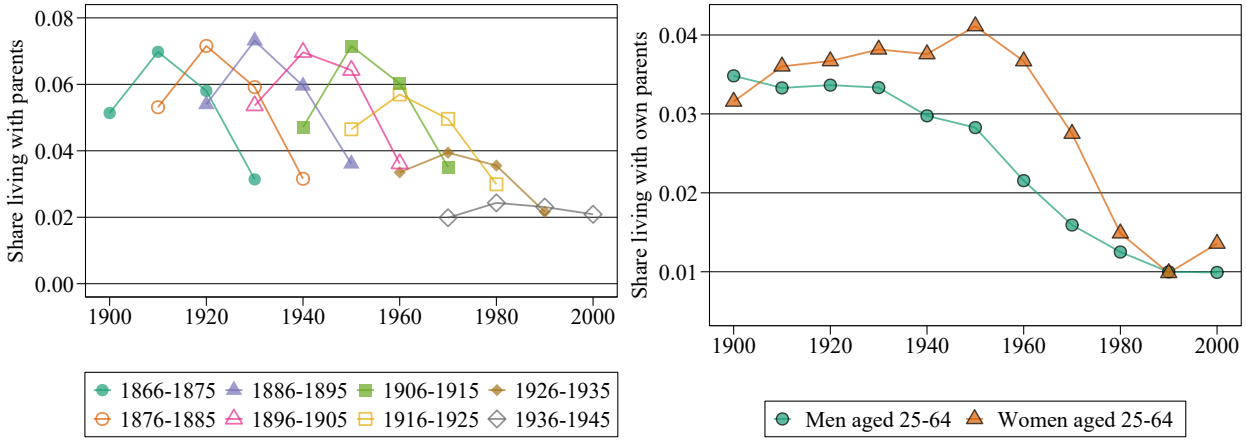



*Notes:* This figure shows share of men receiving retirement benefits in a specified year, separately by 5-year birth cohort. Each cohort is shown at ages 65–69, 70–74, 75–79, and 80–84. Data on payments are from the *Social Security Bulletin*, data on male population by age and year are from United States Bureau of the Census (2004).

Figure 3: Co-residence with parents from children's perspective

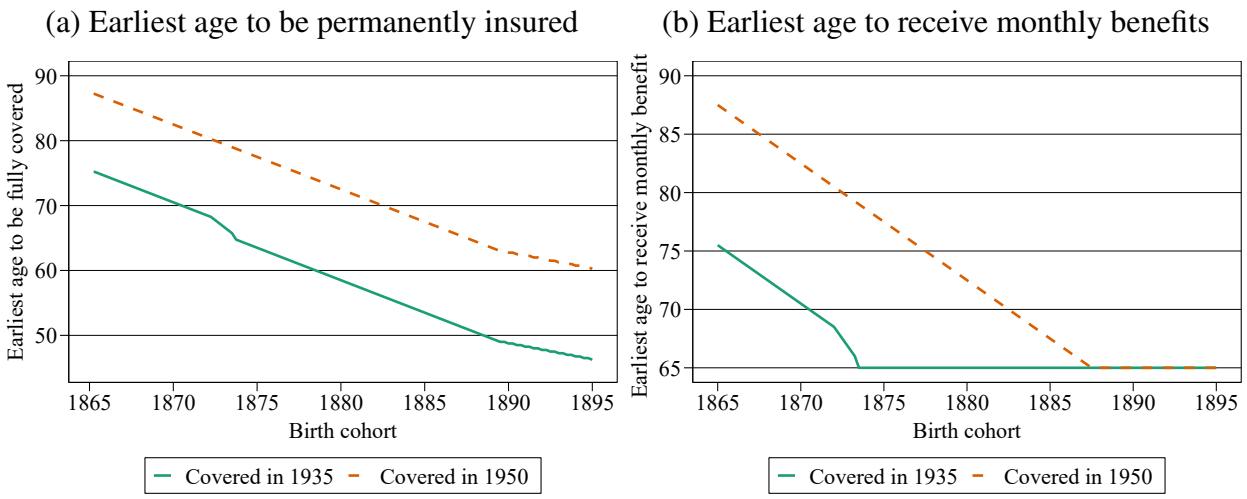
(a) Co-residence with parents, by birth cohort

(b) Co-residence with own parents, by sex



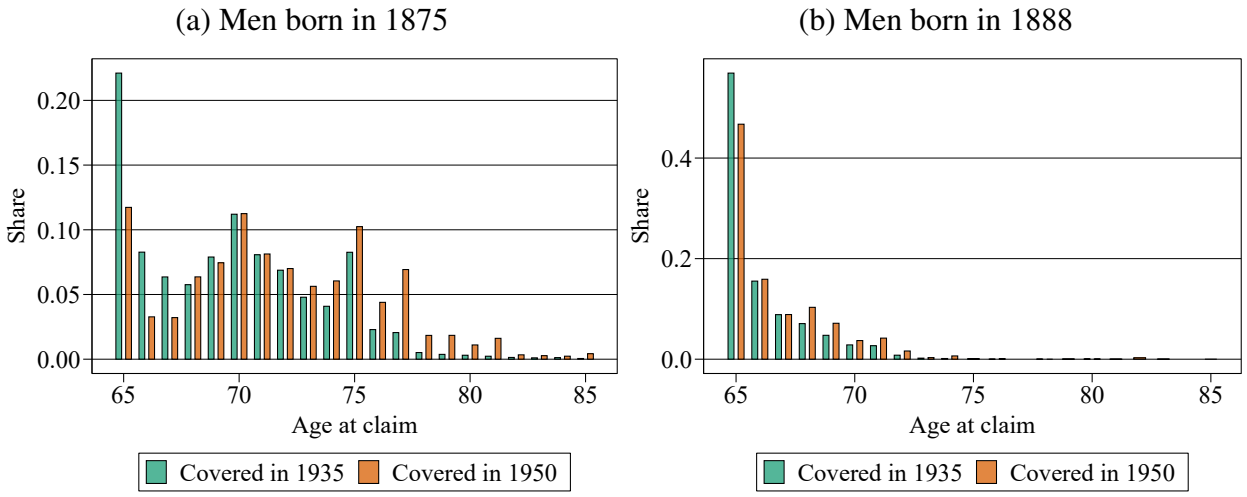

*Notes:* Panel (a) shows the share of US-born men and women who are household head or spouse of the head and co-reside with a parent or parent-in-law in a specified year, separately by 10-year birth cohort. Each cohort is shown at ages 25–34, 35–44, 45–54, and 55–64. Panel (b) shows the share of US-born men and women aged 25–54 who are household head or spouse of the head and co-reside with their own parents in a specified year. Data come from Ruggles et al. (2025b).

Figure 4: Earliest age to be permanently insured and earliest age to receive monthly Social Security benefits



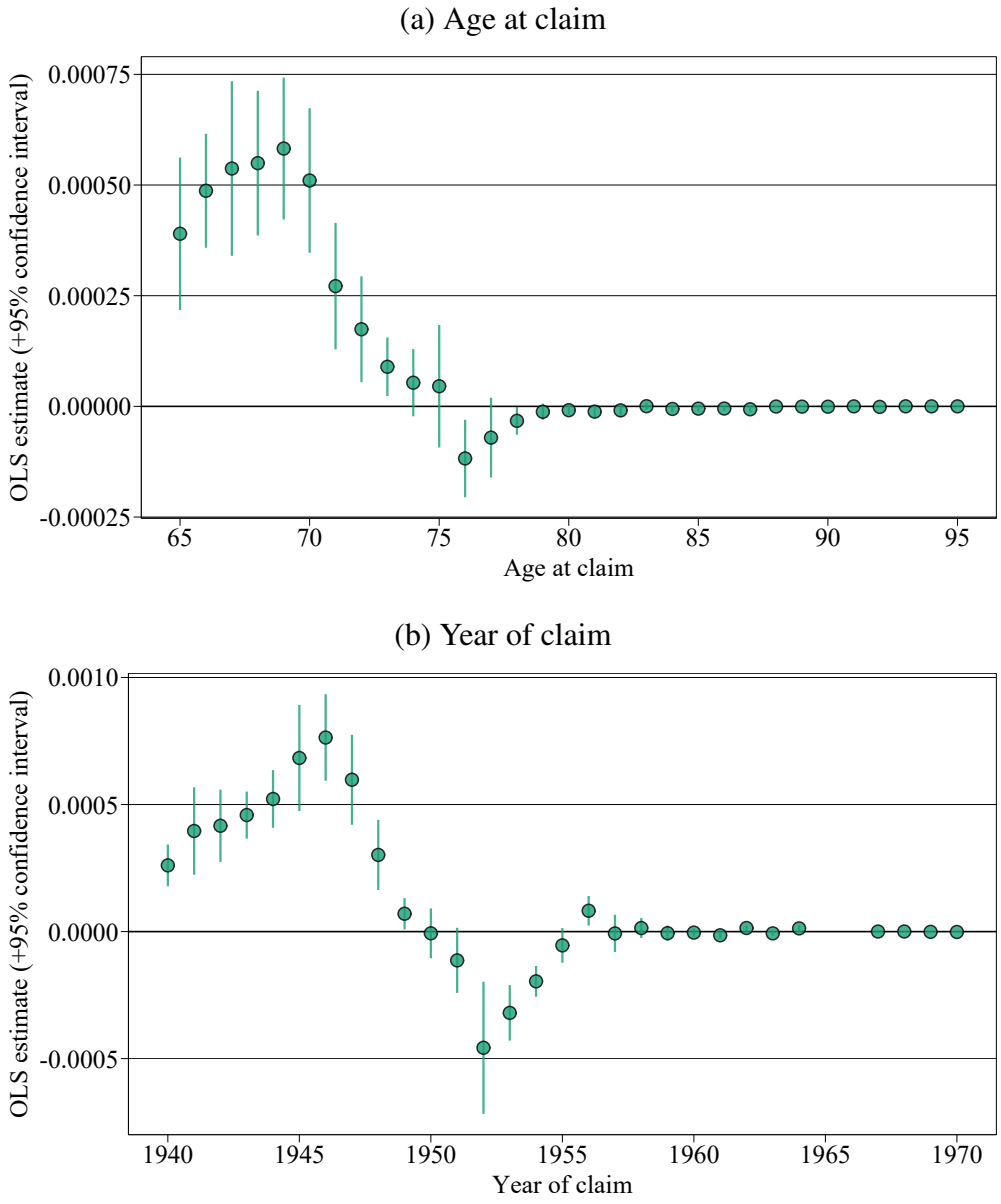

*Notes:* Minimum ages are based on Social Security Act of 1935, 1939 Amendments, and 1950 Amendments, assuming continuous employment in 1935 or 1950-covered employment. The minimum ages shown are those under the 1950 Amendments.

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



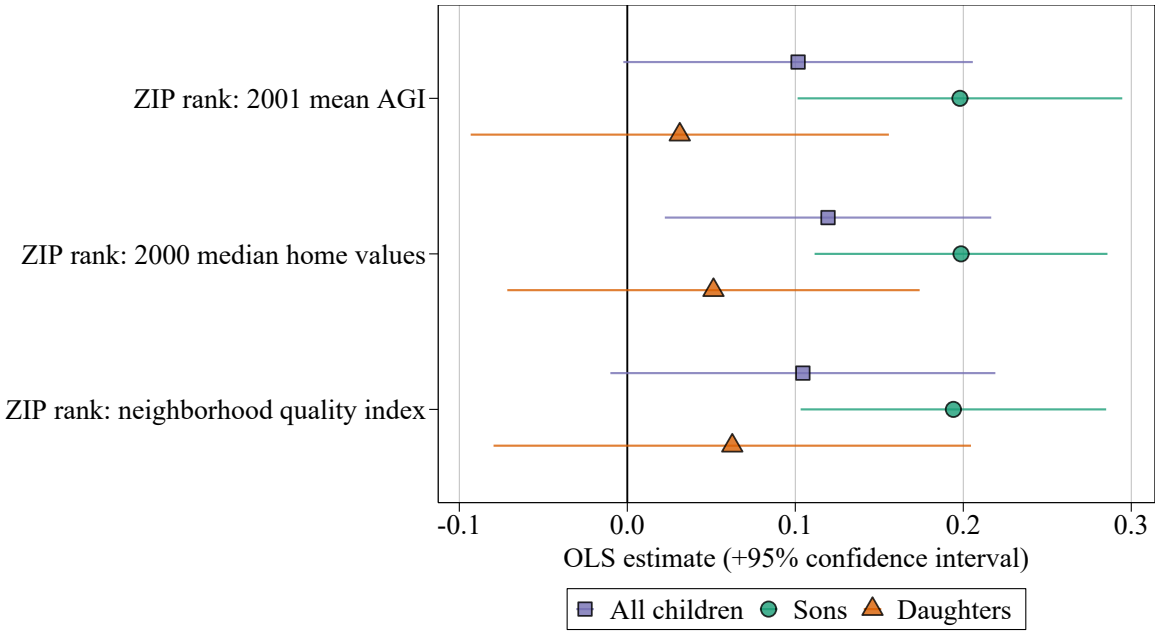

*Notes:* These figures are based on the Numident claims-census linked sample and plot the age at claim distribution for men whose 1930 employment was covered by Social Security in 1935 vs. 1950 (excluding the self-employed), separately for men born in 1875 and men born in 1888. Observations weighted using inverse propensity score weights that re-weight the sample to match the population of interest in 1930.

Figure 6: The impact of exposure to Social Security on age at claim and year of claim



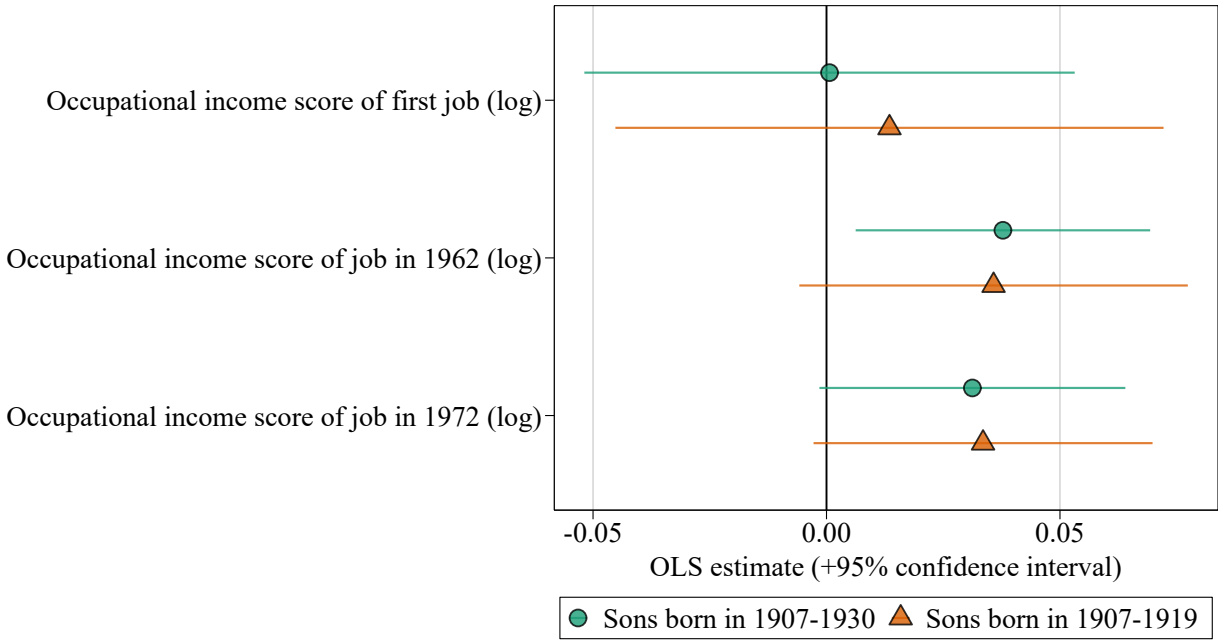

*Notes:* Panel (a) shows the coefficients on Social Security exposure (i.e., negative of earliest age of Social Security eligibility) from equation (2), where the sample consists of all men in our population of interest in the 1930 Census and the dependent variable is an indicator for appearing in the Numident claims-census linked sample and claiming at a specified age. Panel (b) plots analogous estimates for claiming in a specified year. All regressions include fixed effects for men's 3-digit occupation-by-birth cohort, Social Security coverage category, race, and birthplace. 95% confidence intervals based on robust standard errors clustered by men's 3-digit industry. Claims made in 1965 and 1966 are excluded from the sample to avoid capturing Prouty beneficiaries.

Figure 7: The impact of father's exposure to Social Security on children's late-life ZIP rank



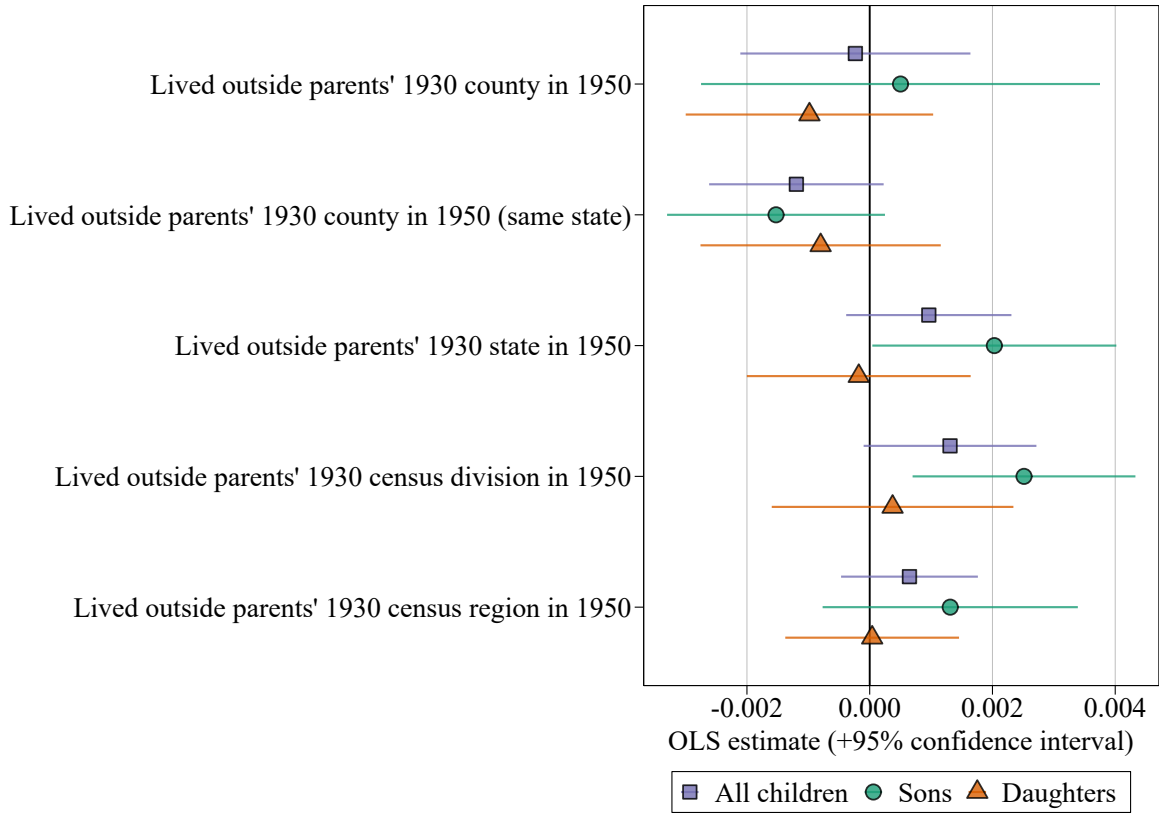

*Notes:* This figure shows the coefficients on father's Social Security exposure (i.e., negative of earliest age of Social Security eligibility) from equation (3), where the sample consists of children in our Numident death-census linked sample and the outcome is the percentile rank of a child's late-life ZIP code in terms of mean AGI, median house values or an index of neighborhood quality. Colors/shapes indicate separate estimates for all children, sons and daughters. All regressions include fixed effects for children's sex-by-birth cohort and fathers' 3-digit occupation-by-birth cohort, Social Security coverage category, race, and birthplace. Observations weighted using inverse propensity score weights that re-weight the sample to match the population of interest in 1930. 95% confidence intervals based on robust standard errors clustered by fathers' 3-digit industry. See Table 4 for detailed regression results.

Figure 8: The impact of father's exposure to Social Security on sons' occupational standing in adulthood



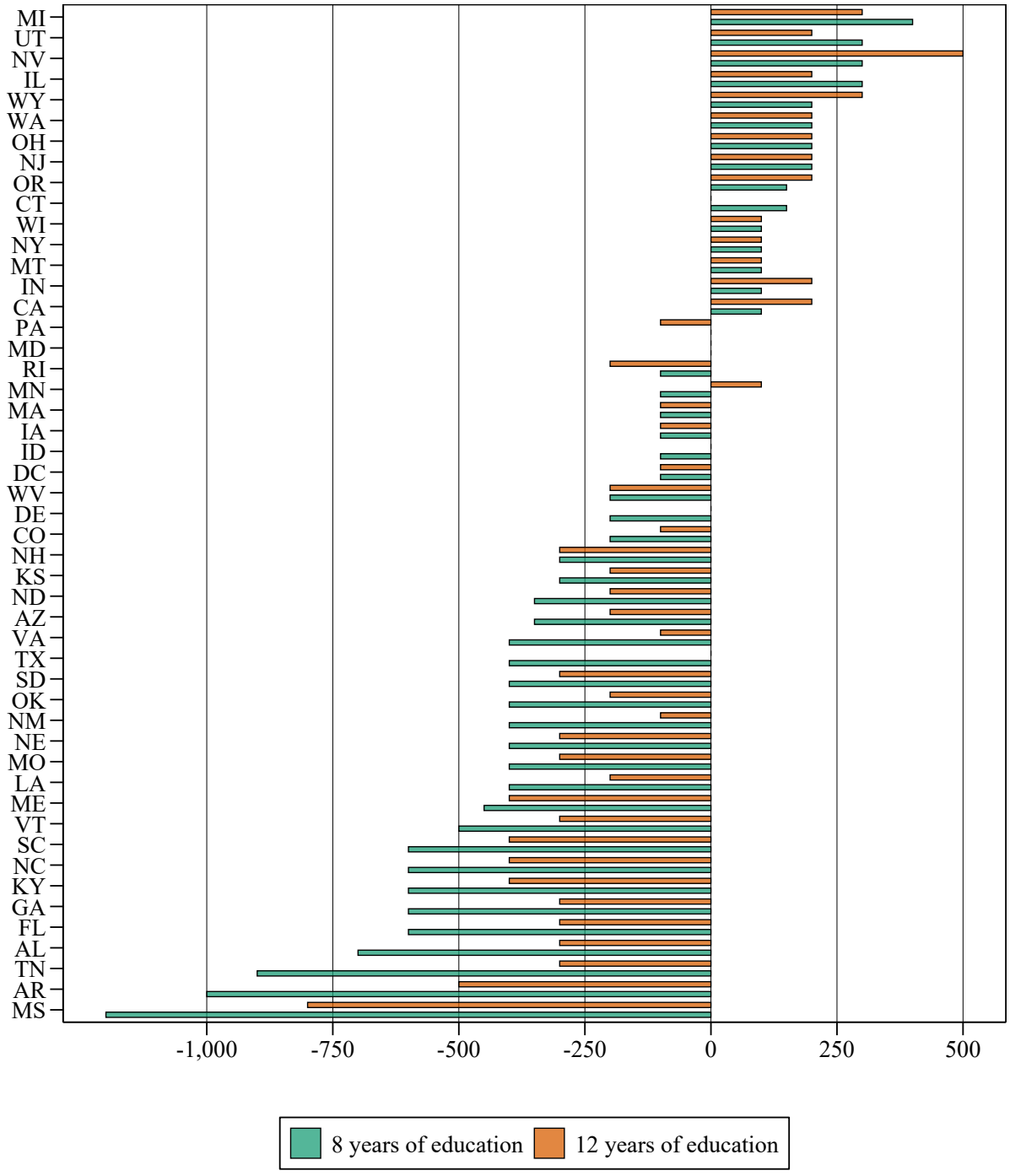

*Notes:* This figure shows the coefficients on father's Social Security exposure (i.e., negative of earliest age of Social Security eligibility) from equation (3), where the sample consists of sons in our 1973 OCG sample and the outcome is the occupational income score associated with sons' first job, job in 1962, or job in 1973 (in logs). To measure occupational standing at these three points in time, we respectively use the 1940, 1960, and 1970-based sex-by-race-by-region-specific occupational income scores from Collins and Wanamaker (2022). Colors/shapes indicate separate estimates for sons born in 1907–1930 and sons born in 1907–1919. All regressions include fixed effects for sons' birth cohort, race and birthplace, and fathers' 3-digit occupation-by-birth cohort and Social Security coverage category. Observations weighted using OCG sampling weights. 95% confidence intervals based on robust standard errors clustered by fathers' 3-digit industry. See Table 5 for detailed regression results.

Figure 9: The impact of father's exposure to Social Security on children's propensity to live outside their parents' 1930 place of residence by 1950



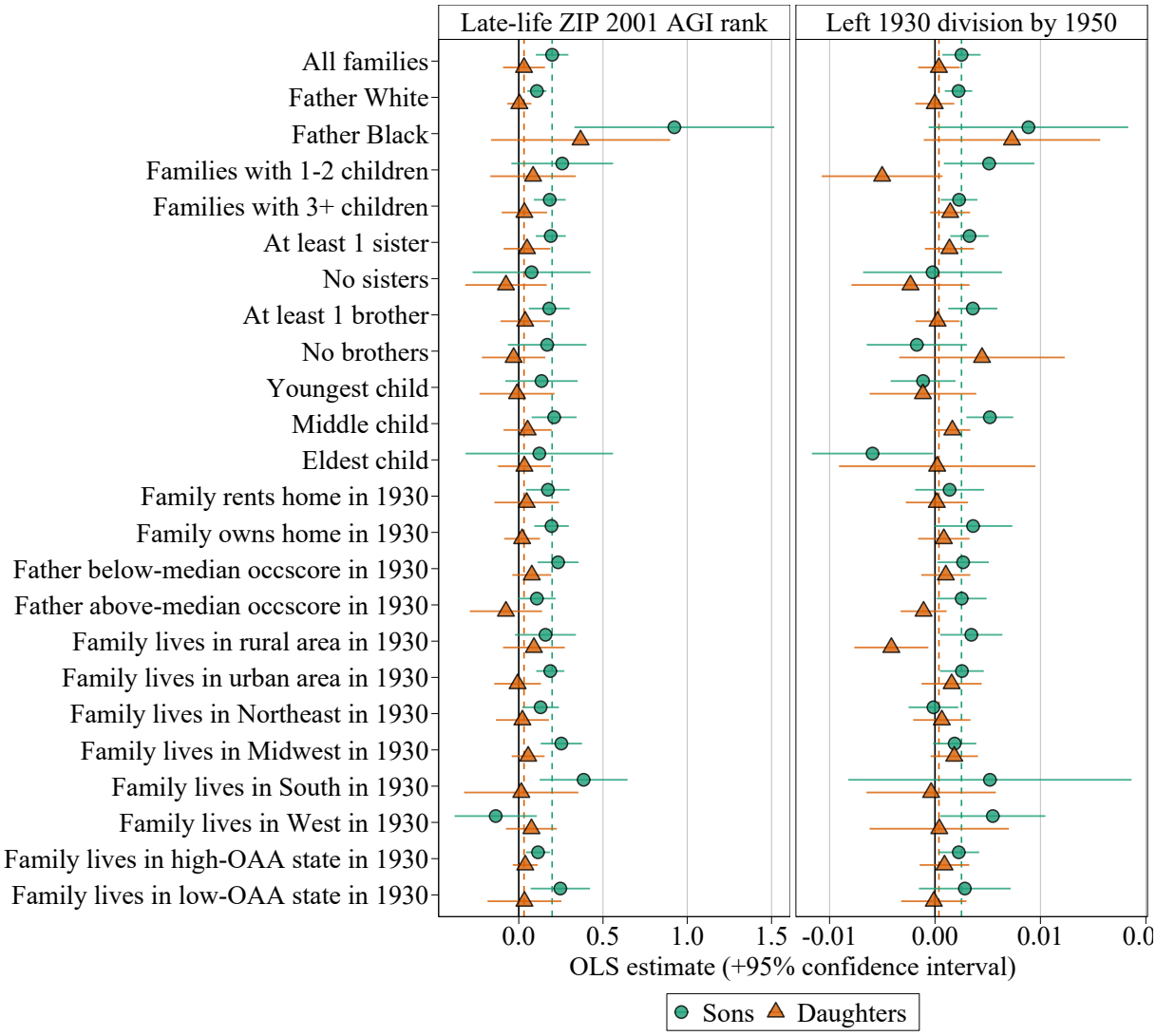

*Notes:* This figure shows the coefficients on father's Social Security exposure (i.e., negative of earliest age of Social Security eligibility) from equation (3), where the sample consists of children in our Numident death-census linked sample who are additionally linked to the 1950 Census and the outcome is an indicator for living outside parents' 1930 county/state/division/region of residence by 1950. Colors/shapes indicate separate estimates for all children, sons and daughters. All regressions include fixed effects for children's sex-by-birth cohort and fathers' 3-digit occupation-by-birth cohort, Social Security coverage category, race, and birthplace. Observations weighted using inverse propensity score weights that re-weight the sample to match the population of interest in 1930. 95% confidence intervals based on robust standard errors clustered by fathers' 3-digit industry. See Table 6 for detailed regression results.

Figure 10: Median earnings by state in 1950 relative to national median, men aged 25–54 with 8 years of education vs. 12 years of education




*Notes:* This figure plots deviations in median earned income by state in 1950 relative to the national median, separately for men aged 25–54 with 8 years of education vs. 12 years of education. Earned income is defined as the sum of wage income and business or farm income. All amounts expressed in 1949 dollars.

Figure 11: The impact of father's exposure to Social Security on children's late-life ZIP rank in terms of 2001 AGI and propensity to live outside their parents' 1930 census division of residence by 1950: Heterogeneity by select characteristics of families and children




*Notes:* This figure shows the coefficients on father's Social Security exposure (i.e., negative of earliest age of Social Security eligibility) from equation (3), where the sample consists of children in our Numident death-census linked sample and the outcome is either the percentile rank of a child's late-life ZIP code in terms of mean AGI (left panel) or an indicator for living outside parents' 1930 census division of residence by 1950 (right panel). Black sample based on 90 percent precision links. Sample restricted to families with at least two children (born in different years) for sibling sex composition (birth order) heterogeneity cuts. High and low-OAA states defined as states with above or below-median OAA payments per person-65+ in December 1939 (Fetter and Lockwood, 2018). Colors/shapes indicate separate estimates for sons and daughters. All regressions include fixed effects for children's sex-by-birth cohort and fathers' 3-digit occupation-by-birth cohort, Social Security coverage category, race, and birthplace. Regressions for sibling sex composition heterogeneity cuts additionally include fixed effects for the number of children in the family. Observations weighted using inverse propensity score weights that re-weight the sample to match the population of interest in 1930. 95% confidence intervals based on robust standard errors clustered by fathers' 3-digit industry.

Table 1: The impact of exposure to Social Security on subsequent claiming behavior

| | Dependent variable: | |
|---|---|---|
| | Linked to claims data | Age at claim |
| | (1) | (2) |
| Social Security exposure | 0.0034*** | -0.1143*** |
| | (0.0005) | (0.0162) |
| Mean of dep. variable | 0.093 | 67.5 |
| *N* | 3,920,008 | 367,732 |

*Notes:* This table shows the coefficients on Social Security exposure (i.e., negative of earliest age of Social Security eligibility) from equation (2), where the sample consists of all men in our population of interest in the 1930 Census in column 1 and the subset of those men appearing in our Numident claims-census linked sample in column 2. In column 1, the outcome is an indicator for appearing in the Numident claims-census linked sample. In column 2, the outcome is age at the time of claiming. All regressions include fixed effects for men's 3-digit occupation-by-birth cohort, Social Security coverage category, race, and birthplace. In column 2, observations are weighted using inverse propensity score weights that re-weight the sample to match the population of interest in 1930. Robust standard errors in parentheses, clustered by men's 3-digit industry. *** 1%, ** 5%, * 10% significance.

Table 2: Selection of 1930 households and children into Numident death-census 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) |
| Social Security exposure | -0.0001 | 0.0003 | 0.0001 | 0.0004 | -0.0002 | -0.0001 |
| | (0.0003) | (0.0007) | (0.0003) | (0.0005) | (0.0002) | (0.0003) |
| Mean of dep. variable | 0.219 | 0.403 | 0.152 | 0.215 | 0.135 | 0.189 |
| *N* | 3,920,008 | 3,920,008 | 3,920,008 | 3,920,008 | 3,920,008 | 3,920,008 |

*Notes:* This table shows the coefficients on Social Security exposure (i.e., negative of earliest age of Social Security eligibility) from equation (2), where the sample consists of all men in our population of interest in the 1930 Census and the outcome is either an indicator for any children/sons/daughters appearing in our Numident death-census linked sample or a discrete variable for the number of children/sons/daughters appearing in our linked sample. All regressions include fixed effects for men's 3-digit occupation-by-birth cohort, Social Security coverage category, race, and birthplace. Robust standard errors in parentheses, clustered by men's 3-digit industry. *** 1%, ** 5%, * 10% significance.

Table 3: Placebo checks: 1930 Census characteristics

| | Dependent variable: Household characteristic in 1930 | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Homeowner | House value (log) | 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* | | | | | | | | | |
| Social Security exposure | -0.0009 | 0.0021 | -0.0002 | 0.0002 | 0.0002 | 0.0020 | -0.0004 | -0.0002 | -0.0001 |
| | (0.0006) | (0.0042) | (0.0009) | (0.0006) | (0.0002) | (0.0023) | (0.0003) | (0.0003) | (0.0003) |
| Mean of dep. variable | 0.49 | 2,750 | 0.47 | 0.74 | 0.05 | 2.2 | 0.94 | 0.93 | 0.11 |
| *N* | 3,912,603 | 1,862,037 | 3,920,008 | 3,920,008 | 3,920,008 | 3,920,008 | 3,920,008 | 3,707,183 | 3,707,185 |
| *Panel (b): Families in Numident death-census linked sample* | | | | | | | | | |
| Social Security exposure | -0.0004 | 0.0040 | 0.0006 | 0.0003 | 0.0004 | 0.0032 | -0.0002 | -0.0002 | -0.0001 |
| | (0.0008) | (0.0032) | (0.0013) | (0.0008) | (0.0004) | (0.0030) | (0.0004) | (0.0004) | (0.0008) |
| Mean of dep. variable | 0.49 | 2,760 | 0.48 | 0.74 | 0.05 | 3.8 | 0.94 | 0.94 | 0.09 |
| *N* | 857,934 | 452,309 | 858,612 | 858,612 | 858,612 | 858,612 | 858,612 | 839,750 | 839,750 |

*Notes:* This table shows the coefficients on Social Security exposure (i.e., negative of earliest age of Social Security eligibility) from equation (2), where the outcome is a specified household characteristic (indicated in the column titles). In panel (a), the sample consists of all men in our population of interest in the 1930 Census while in panel (b) the sample only includes the subset of men who appear in our Numident death-census linked sample. All regressions include fixed effects for men's 3-digit occupation-by-birth cohort, Social Security coverage category, race, and birthplace. In column 2, mean 1930 house values in contemporary dollars are shown. In panel (b), observations are weighted using inverse propensity score weights that re-weight the sample to match the population of interest in 1930. Robust standard errors in parentheses, clustered by men's 3-digit industry. *** 1%, ** 5%, * 10% significance.

Table 4: The impact of father's exposure to Social Security on children's late-life ZIP quality

| | Dependent variable: Death ZIP characteristics | | | | | |
|---|---|---|---|---|---|---|
| | Mean AGI (2001) | | Median home values (2000) | | ZIP quality index | |
| | Rank | Log | Rank | Log | Rank | SD |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| *Panel (a): All children* | | | | | | |
| Social Security exposure | 0.102* | 0.0009 | 0.119** | 0.0022* | 0.105* | 0.0036 |
| | (0.053) | (0.0007) | (0.050) | (0.0011) | (0.058) | (0.0026) |
| Mean of dep. variable | 64.0 | 47,937 | 68.8 | 144,113 | 54.4 | 0.16 |
| *N* | 1,303,454 | 1,303,422 | 1,270,763 | 1,269,551 | 1,267,789 | 1,267,789 |
| *Panel (b): Sons* | | | | | | |
| Social Security exposure | 0.198*** | 0.0024*** | 0.199*** | 0.0043*** | 0.194*** | 0.0054** |
| | (0.049) | (0.0008) | (0.044) | (0.0012) | (0.046) | (0.0024) |
| Mean of dep. variable | 63.8 | 47,639 | 68.7 | 143,128 | 54.3 | 0.15 |
| *N* | 655,886 | 655,862 | 638,754 | 638,195 | 637,235 | 637,235 |
| *Panel (c): Daughters* | | | | | | |
| Social Security exposure | 0.031 | 0.0001 | 0.051 | 0.0005 | 0.062 | 0.0034 |
| | (0.064) | (0.0009) | (0.063) | (0.0015) | (0.073) | (0.0028) |
| Mean of dep. variable | 63.6 | 47,509 | 68.4 | 142,359 | 53.7 | 0.13 |
| *N* | 647,568 | 647,560 | 632,009 | 631,356 | 630,554 | 630,554 |

*Notes:* This table shows the coefficients on father's Social Security exposure (i.e., negative of earliest age of Social Security eligibility) from equation (3), where the sample consists of children in our Numident death-census linked sample and the outcome is the quality a child's late-life ZIP code in terms of mean AGI, median house values or the neighborhood quality index (either in percentile ranks or in logs). Panels show separate estimates for all children, sons and daughters. All regressions include fixed effects for children's sex-by-birth cohort and fathers' 3-digit occupation-by-birth cohort, Social Security coverage category, race, and birthplace. In columns 2 and 4, means of ZIP-level 2001 AGI and 2000 house values in contemporary dollars are shown. Observations weighted using inverse propensity score weights that re-weight the sample to match the population of interest in 1930. Robust standard errors in parentheses, clustered by fathers' 3-digit industry. *** 1%, ** 5%, * 10% significance.

Table 5: The impact of father's exposure to Social Security on sons' occupational standing in adulthood

| | Dependent variable: Occupational income score | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | First occupation | | | 1962 occupation | | | 1973 occupation | | |
| | Log | Level | Missing | Log | Level | Missing | Log | Level | Missing |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| *Panel (a): Sons born in 1909–1930* | | | | | | | | | |
| Social Security exposure | 0.001 | -18 | -0.023 | 0.038** | 363*** | -0.005 | 0.031* | 343* | -0.009 |
| | (0.027) | (95) | (0.018) | (0.016) | (134) | (0.024) | (0.017) | (185) | (0.014) |
| Mean of dep. variable | 8.11 | 3,691 | 0.07 | 8.95 | 8,299 | 0.10 | 9.20 | 10,685 | 0.09 |
| *N* | 2,611 | 2,611 | 2,819 | 2,546 | 2,546 | 2,819 | 2,566 | 2,566 | 2,819 |
| *Panel (b): Sons born in 1909–1919* | | | | | | | | | |
| Social Security exposure | 0.013 | 37 | -0.026 | 0.036* | 363* | -0.021 | 0.034* | 343* | -0.018 |
| | (0.03) | (96) | (0.025) | (0.021) | (186) | (0.028) | (0.019) | (196) | (0.022) |
| Mean of dep. variable | 8.08 | 3,609 | 0.07 | 8.95 | 8,329 | 0.10 | 9.19 | 10,592 | 0.13 |
| *N* | 1,507 | 1,507 | 1,626 | 1,457 | 1,457 | 1,626 | 1,421 | 1,421 | 1,626 |

*Notes:* This table shows the coefficients on father's Social Security exposure (i.e., negative of earliest age of Social Security eligibility) from equation (3), where the sample consists of sons in our 1973 OCG sample and the outcome is the occupational income score associated with sons' first job, job in 1962, or job in 1973 (in logs or in 1972 dollars), or an indicator for having missing occupational information. To measure occupational standing at these three points in time, we respectively use the 1940, 1960, and 1970-based sex-by-race-by-region-specific occupational income scores from Collins and Wanamaker (2022). Panels show separate estimates for sons born in 1907–1930 and sons born in 1907–1919. All regressions include fixed effects for sons' birth cohort, race and birthplace, and fathers' 3-digit occupation-by-birth cohort and Social Security coverage category. Observations weighted using OCG sampling weights. Robust standard errors in parentheses, clustered by fathers' 3-digit industry. *** 1%, ** 5%, * 10% significance.

Table 6: The impact of father's exposure to Social Security on children's propensity to live outside their parents' 1930 place of residence by 1950

| | Dependent variable: Children's 1950 location vs. parents' 1930 location | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Different location | | | | | Distance between counties | | |
| | County | County (same state) | State | Census division | Census region | $>0$ and $\leqslant 100$ miles | $>100$ and $\leqslant 500$ miles | $>500$ miles |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| *Panel (a): All children* | | | | | | | | |
| Social Security exposure | -0.0002 | -0.0012* | 0.0010 | 0.0013* | 0.0006 | 0.0000 | -0.0007 | 0.0005 |
| | (0.0010) | (0.0007) | (0.0007) | (0.0007) | (0.0006) | (0.0007) | (0.0009) | (0.0005) |
| Mean of dep. variable | 0.43 | 0.20 | 0.23 | 0.16 | 0.12 | 0.19 | 0.14 | 0.10 |
| *N* | 601,276 | 601,276 | 601,321 | 601,321 | 601,321 | 601,276 | 601,276 | 601,276 |
| *Panel (b): Sons* | | | | | | | | |
| Social Security exposure | 0.0005 | -0.0015* | 0.0020** | 0.0025*** | 0.0013 | 0.0000 | -0.0013 | 0.0018* |
| | (0.0017) | (0.0009) | (0.0010) | (0.0009) | (0.0011) | (0.0008) | (0.0011) | (0.0010) |
| Mean of dep. variable | 0.42 | 0.19 | 0.23 | 0.16 | 0.13 | 0.18 | 0.13 | 0.11 |
| *N* | 309,657 | 309,657 | 309,682 | 309,682 | 309,682 | 309,657 | 309,657 | 309,657 |
| *Panel (c): Daughters* | | | | | | | | |
| Social Security exposure | -0.0010 | -0.0008 | -0.0002 | 0.0004 | 0.0000 | -0.0003 | -0.0001 | -0.0006 |
| | (0.0010) | (0.0010) | (0.0009) | (0.0010) | (0.0007) | (0.0011) | (0.0011) | (0.0006) |
| Mean of dep. variable | 0.43 | 0.21 | 0.22 | 0.15 | 0.11 | 0.19 | 0.14 | 0.10 |
| *N* | 291,619 | 291,619 | 291,639 | 291,639 | 291,639 | 291,619 | 291,619 | 291,619 |

*Notes:* This table shows the coefficients on father's Social Security exposure (i.e., negative of earliest age of Social Security eligibility) from equation (3), where the sample consists of children in our Numident death-census linked sample who are additionally linked to the 1950 Census and the outcome is an indicator for living outside parents' 1930 county/state/division/region of residence or a specified number of miles away from their 1930 county of residence by 1950. Panels show separate estimates for all children, sons and daughters. All regressions include fixed effects for children's sex-by-birth cohort and fathers' 3-digit occupation-by-birth cohort, Social Security coverage category, race, and birthplace. Observations weighted using inverse propensity score weights that re-weight the sample to match the population of interest in 1930. Robust standard errors in parentheses, clustered by fathers' 3-digit industry. *** 1%, ** 5%, * 10% significance.

Table 7: The impact of father's exposure to Social Security on children's direction of migration

| | Dependent variable: Earnings in 1950 in children's 1950 state vs. parents' 1930 state | | | | | |
|---|---|---|---|---|---|---|
| | Education-specific earnings | | | Overall earnings | | |
| | Median | 25th percentile | 75th percentile | Median | 25th percentile | 75th percentile |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| *Panel (a): All children with information on completed years of schooling from 1940 or 1950* | | | | | | |
| Social Security exposure | 0.38 | -0.52 | 0.60 | -0.74 | -0.74 | -0.30 |
| | (0.67) | (0.42) | (0.66) | (0.73) | (0.46) | (0.67) |
| Mean of dep. variable | 16.2 | 0.8 | 21.2 | 17.7 | 2.2 | 22.3 |
| *N* | 195,533 | 195,533 | 195,533 | 195,533 | 195,533 | 195,533 |
| *Panel (b): Sons with information on completed years of schooling from 1940 or 1950* | | | | | | |
| Social Security exposure | 2.14*** | 0.12 | 2.24*** | 0.10 | -0.55 | 0.01 |
| | (0.74) | (0.47) | (0.59) | (1.46) | (0.54) | (1.07) |
| Mean of dep. variable | 15.6 | -0.3 | 21.9 | 15.9 | 0.7 | 21.5 |
| *N* | 94,046 | 94,046 | 94,046 | 94,046 | 94,046 | 94,046 |
| *Panel (c): Daughters with information on completed years of schooling from 1940 or 1950* | | | | | | |
| Social Security exposure | -1.21 | -1.06* | -0.85 | -1.36 | -0.88 | -0.43 |
| | (0.89) | (0.57) | (0.82) | (1.13) | (0.59) | (0.90) |
| Mean of dep. variable | 16.4 | 1.7 | 20.4 | 19.1 | 3.3 | 22.9 |
| *N* | 101,487 | 101,487 | 101,487 | 101,487 | 101,487 | 101,487 |

*Notes:* This table shows the coefficients on father's Social Security exposure (i.e., negative of earliest age of Social Security eligibility) from equation (3), where the sample consists of children in our Numident death-census linked sample who are additionally linked to the 1950 Census and for which we can observe completed years of schooling either via additional links to the 1940 Census (respondents aged 25 or older and not attending school) or in the 1950 Census (sample-line respondents aged 25 or older and not attending school). In columns 1 to 3, the outcomes are differences in moments of the 1950 education-specific earnings distribution in children's 1950 state of residence vs. their parents' 1930 state of residence (median, 25th percentile, and 75th percentile). In columns 4 to 6, the outcomes are analogous but for moments of the overall earnings distribution by state in 1950. Moments of the earnings distribution by state (and education) are calculated using (sample-line) men aged 25–54 in the 1950 Census. All amounts are expressed in 1949 dollars. Panels show separate estimates for all children, sons and daughters. All regressions include fixed effects for children's sex-by-birth cohort and fathers' 3-digit occupation-by-birth cohort, Social Security coverage category, race, and birthplace. Observations weighted using inverse propensity score weights that re-weight the sample to match the population of interest in 1930. Robust standard errors in parentheses, clustered by fathers' 3-digit industry. *** 1%, ** 5%, * 10% significance.

# Online Appendix for

## Long-Run Intergenerational Effects of Social Security

*By* Daniel K. Fetter, Lee M. Lockwood, and Paul Mohnen

# A Data appendix

## A.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 (including mothers' maiden names).[1] The sibling linkages are then used to create a crosswalk between SSNs and family identifiers. In practice, we use the crosswalk from Bailey et al. (2026). These links 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 linking (since people can have multiple siblings). As a result, trainers were allowed to select multiple links, and their decisions were modeled using a random forest model treating each pair of records as a separate linking decision. In the final crosswalk, 35 percent of individuals can be linked to at least one other sibling in the SS-5 data (see Bailey et al. (2026) for more details).

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 information on parents' year of birth 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 fathers' age in the census. Note that families can consist of 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 fathers' last name initial and the union of children's state/continent of birth.[2] Blocking on the union of children's place of birth implies that the only requirement we impose is that at least one of the children's place of birth in candidate families matches one of the children's place of birth in the family under consideration. Within each block, we select the top 20 candidate households based on a Jaro-Winkler string similarity score between parent names.

Next, we generate training data for a random sample of 500 families using the process described in Section 3. In particular, we display the following information to trainers: fathers' full name, mothers' 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 children that would have been alive at the time of the census, and mask children in the census if there is little chance that they match any of the children in the family under consideration (based on age and birthplace information). We also only display the eight youngest children to avoid overcrowding the computer screen during the training process.

We are able to manually link 61.9 percent of families in the training data to a household in the 1930 Census. To train the two-stage model from Bailey et al. (2023) to mimic the hand-linking process, we first split the training data into a "training set" containing 70 percent of the training data and a "test set" containing the remaining 30 percent. The training set is used to iteratively train the algorithm by varying model features until we are satisfied with its performance, while the test set is used to assess the out-of-sample performance of the model once it is finalized. The final set of features used to train the model are shown in Appendix Table A.12. Following standard practice, probabilities are estimated using 10-fold cross-validation to avoid overfitting. Panel (a) in Appendix Figure A.12 shows the resulting precision-recall curve. The model is able to achieve

[1] Place of birth, year of birth, and race information is also used to influence linking decisions, as siblings are more likely to be born in the same state/country within a few years of each other and have matching race information.

[2] Blocking on a certain characteristics is common practice in historical record linking since it reduces the number of pairs for which string similarity scores must be computed.

a recall rate of 85.6 percent at 97 percent precision in the training set (match rate of 53 percent). The corresponding numbers in the test set are 78.5 percent recall at 96.1 percent precision (match rate of 50.7 percent).[3]

Using the trained model and cross-validated probability threshold, we scale-up linking to the full sample of SS-5 families. This yields 7 million links, for a match rate of 33.9 percent.[4] Among those links, there are cases in which the same census household is linked to two (or more) families in the SS-5 data. While some of these cases are true conflicts (reflecting our model's 3 percent error rate, for example), a significant share of them are not. The reason is that parents can naturally appear on multiple SS-5 records for their different children and we are not able to make all possible links between these records (we can manually link only a subset of all true links due to missing or imperfectly reported names, and the model can only replicate a subset of those links at 97 percent precision). Although we might not be able to link two SS-5 records together solely based on parent name information, we might still 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 200 cases, 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 labeled 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.[5] 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, for a final match rate of 33.6 percent. 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 (match rate of 30.1 percent). This number of links to 1920 is lower because fewer children are born by 1920 in the SS-5 data.

### A.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 process used to link SS-5 families to households in the 1920 and 1930 Censuses. 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 fathers' age in 1920, mothers' age in 1920, and mothers' place of birth during the training process. We are able to manually link 52.6 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). We used a similar set of features as in Appendix Table A.12 to train the model, which is able to

[3] Note that in this instance the training set only contains 150 cases. With more training data the discrepancy between the in-sample and out-of-sample performance statistics would be smaller.

[4] This match rate is lower than the match rates in our training data because we dropped the requirement that SS-5 families have at least one child aged under 25 at the time of the census when scaling up. This added some extra links, but artificially reduced the match rate by expanding the denominator to include families with a much lower probability of being linked.

[5] 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.

achieve a recall rate of 90 percent at 97 percent precision in the training set, for a match rate of 47.3 percent (see panel (b) in Appendix Figure A.12 for the precision-recall curve). The corresponding numbers in the test set are 90.6 percent recall at 100 percent precision (match rate of 51.3 percent).

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 households in the censuses, these are true conflicts since parents can only appear in one household in any given census. To resolve those conflicts, we used model-based match probabilities to prioritize higher-quality links and dropped all remaining conflicts. The final number of households headed by a man born in 1860–1905 in the 1930 Census linked to a household in the 1920 Census at 97 percent precision is 7.6 million (match rate of 47.3 percent).

### A.3 Linking men and women in SS-5 records to the 1950 Census

To link male and female children in SS-5 records to themselves in the 1950 Census, we followed the process described in Bailey et al. (2026) to link SS-5 records to the 1940 Census (including recycling their training data and model). While we are only interested in a subset of children for our analysis, we simply linked all children who would have been alive in 1950 and imposed our sample restrictions after combining all links. Below we provide an overview of how we adapted the process from Bailey et al. (2026) for our purposes, but more details can be found in their paper.

For male 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 person's age in 1950. Taking the union of blocks based on first or last name initials implies that potentials candidates only need to satisfy one of those two conditions, which is a weaker requirement. We then compute Jaro-Winkler name similarity scores and retain 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: first, middle and last names, age in 1940, country of birth (if born outside the US), and race.[6] The manual match rate in this training data was 42.6 percent, and the recall rate achieved by the model in the training set was 68.6 percent at 97 percent precision (match rate of 29.2 percent). The corresponding numbers in the test set were 69.9 percent recall at 95.3 percent precision (match rate of 28.5 percent). We used this model and the cross-validated probability threshold to generate 97-percent precision links to the 1950 Census, ultimately linking 6.5 million men in SS-5 records for a final match rate of 33 percent.[7]

The process of linking female SS-5 records is analogous, with a few key differences. First, multiple versions of each female record are generated, one with women's birth name as their last name and others using all possible married names as their last name (most women have at most one married name, but some have multiple married names reflecting re-marriages).[8] Second, we additionally block on marital status when generating potential candidates. Women linked using

[6] Race and country of birth were displayed as optional pieces of information to take into account since they are not always accurately reported.

[7] Similar to linking households across censuses, it is possible for the same census individual to be linked to two different SSNs, which are true conflicts since a person can only be associated with one SSN. To resolve those cases, we again prioritized higher-quality links and discarded the remaining conflicts.

[8] In practice, we use every unique last name variant across all SS-5 entries for the same SSN, which can include name misspellings that were later corrected. This generally leads to duplicate links that can be dropped at the end.

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 (married, divorced, or widowed). The original training data for linking women to the 1940 Census was based on a random sample of 500 SSNs (corresponding to around 1,500 name variants). Unlike other links, it is possible for female SSNs to be linked to different people in the census (if their birth and married names happen to be common names, for example). For women that had multiple links in the training data, those cases were manually resolved by either choosing one link and discarding the others or discarding all links if there was too much ambiguity. The SSN-level match rate in this training data was 52 percent and the SSN-level recall rate achieved by the model in the training set was 70 percent at 96.6 percent precision (match rate of 36.4 percent). The corresponding numbers in the test set were 68.3 percent recall at 95.3 percent precision (match rate of 34.4 percent). Again, we used this trained model and the cross-validated probability threshold to generate 97-percent precision links to the 1950 Census for women (here we simply discarded all links when there was a within-SSN conflict). The final number of links for women was 5.7 million, for a final match rate of 33.4 percent.

### A.4 Linking individuals in claims records to 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 retirement benefits, we focus on the subset of life claims and drop death claims.[9] We separately link claims records with non-missing birthplace information (40 percent of records) and claims records with missing birthplace information (60 percent of records). Claims records with non-missing birthplace information are linked using name, year of birth, and birthplace 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 name similarity scores 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). We also displayed the state associated with the first 3 digits of a person's SSN and the state of residence of potential candidates in the 1930 Census. Until 1972, the first 3 digits of an SSN was tied to the field office in which the application was filed. 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 52.8 percent of claims records with non-missing birthplace information to the 1930 Census, and our model achieves a recall rate of 66.5 percent at 97 percent precision in the training set for a match rate of 35.4 percent (see panel (c) in Appendix Figure A.12 for the precision-recall curve). The corresponding numbers in the test set are 67.5 percent recall at 97.3 percent precision (match rate of 37.7 percent).

We use an analogous procedure to link claims records with missing birthplace 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 more ambiguity in the set of potential candidates. Consequently, we are only able to manually link 33.9 percent of claims records with missing birthplace information to the 1930 Census. The model recall rate at 97 percent precision is 67.1 percent, for a match rate of 22.7 percent (see panel (d) in

[9] Life claims capture retirement and disability benefits. Death claims capture survivor and death benefits. Around 77 percent of records in the claims data are life claims.

Appendix Figure A.12 for the precision-recall curve). The corresponding numbers in the test set are 64.5 percent recall at 98.4 percent precision (match rate of 20.3 percent).

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 1.3 million links, for a final match rate of 26.2 percent. To augment our death-census linked sample, we also link men born between 1895 and 1920 in claims records to the 1920 Census using the same process (here we keep all types of claims since we are not interested in the claims records themselves). This yields 1.2 million links, for a match rate of 13.5 percent. This match rate is lower because the share of claims records with non-missing birthplace information in this sample is lower (16 percent).

### A.5 Details on HRS sample construction

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.[10] 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 and late-career earnings based on linkages to administrative 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 in the HRS files (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 are likely a better measure for women in the HRS cohorts, 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.[11] We can also generate estimates of late-career earnings by subtracting lifetime earnings through age 50 from lifetime earnings through age 65 (i.e., cumulative earnings between ages 50 and 65). 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 income or other characteristics. The final sample contains around 22,200 HRS respondents aged 65+ at the time of their last interview, around 55 percent of whom are women.

### A.6 Measuring family size, sibling sex composition, and birth order

Since the 1920 and 1930 Censuses and Numident records only contain a subset of all children ever born in a given family, and because there are no identifiers that we can use to identify unique

[10]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.

[11]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 algorithm. For more details, see Fang (2021).

children (i.e., the same child can appear across multiple records), characteristics like family size, sibling sex composition, and birth order must be approximated using what we can observe in the data.[12] Here we describe our approach.

To obtain estimates of family size (i.e., the total number of children in a family), we add together: (1) the number of sons and daughters that appear in the household in 1920, (2) the number of sons and daughters who were born after 1920 that appear in the household in 1930, and (3) the number of sons and daughters who were born after 1930 that appear in our linked sample (i.e., in Numident records). In some families, the unconditional total number of sons and daughters in our linked sample (i.e., regardless of when they are born) can strictly exceed the corresponding total number of sons and daughters obtained by adding (1), (2) and (3). This can happen for a variety of reasons, including measurement error in self-reported sex and year of birth information, the fact that some children born before 1930 can appear in Numident records but not in the censuses (e.g., older children who left the household prior to 1920), but also linking errors. In those cases, we use the higher number as our final estimate of the total number of sons and daughters. Those numbers are combined to obtain an estimate of the total number of children, and we also use them to determine the sex composition of siblings. For example, a daughter has at least 1 brother if the number of sons in the family is greater than 0, and she has at least 1 sister if the number of daughters is greater than 1. In the analysis, we focus on families with at least two children and control for family size to try to avoid confounding differences in sex composition with differences in family size.

Rather than trying to determine the exact birth order of a given child in a family, we only try to determine whether that child is the youngest child, the eldest child, or a middle child. To that end, we compute the minimum and maximum year of birth among all children we observe across the 1920 Census, 1930 Census, and Numident records. We then define a child to be a youngest child if his or her year of birth in the Numident data is equal to the maximum year of birth, an eldest child if it is equal to the minimum, and a middle child if it is strictly between the minimum and maximum. Note that eldest and youngest children are not necessarily unique due to multiple births (e.g., twins). Also, it is possible for the minimum and maximum year of birth to coincide (e.g., only-children). In the analysis, we focus on families with at least two children born in different years to avoid having the same children simultaneously being counted as eldest and youngest children.

# B Classification of employment into OASI coverage categories

## B.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

[12] For the most part, children can be uniquely identified within Numident records and within each decennial census (apart from rare cases where the same person was erroneously assigned two SSNs or accidentally listed twice during the census enumeration process). The challenge lies in identifying unique children across records.

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.[13] 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

[13]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.

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., 2025b), 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 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.[14] 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

[14]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.

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.

### B.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 percent 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 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.

## C 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).[15] 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).

## D Reconciling late-career earnings and late-life neighborhood estimates

Our estimates imply that a ten-year earlier predicted Social Security eligibility for fathers increased sons' late-career occupational income scores by over 30 percent while increasing ZIP-level average income and median home values near the end of sons' lives by only about 2 to 4 percent. These magnitudes are not directly comparable: The former is an individual-level measure of earnings potential, whereas the latter are neighborhood-level averages. In this section, we show that the two

[15] 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.

sets of estimates are quantitatively consistent once one calibrates from individual late-career earnings to late-life ZIP-level outcomes using the cross-sectional relationship observed in the HRS. We interpret this exercise as a calibration and consistency check rather than as a formal decomposition.

We estimate the cross-sectional relationship between individual late-career earnings and late-life ZIP income using the HRS sample described in Section 5.1 and Appendix A.5. Specifically, among male HRS respondents we estimate

$$\log(\text{ZIP income}_i) = \alpha + \beta \log(\text{earnings}_i) + \varepsilon_i, \tag{4}$$

where ZIP income$_i$ is the mean adjusted gross income in 2001 in individual $i$'s last known ZIP code of residence, and earnings$_i$ is a measure of individual $i$'s late-career earnings from SSA administrative records, defined as cumulative earnings between ages 50 and 65. We estimate $\hat{\beta} = 0.075$ (SE: 0.003), implying that a 10-percent increase in individual earnings is associated with about a 0.75-percent increase in ZIP-level income (Appendix Table A.13, column 1).[16,17]

We use $\hat{\beta}$ as a calibration to assess whether our occupational-income and ZIP-based estimates are quantitatively consistent. This calculation requires two conditions. First, sons whose late-career earnings and late-life neighborhoods were affected by their fathers' Social Security eligibility must sort across neighborhoods conditional on earnings in roughly the same way as the broader population in which $\hat{\beta}$ is estimated. Second, the conditional expectation of log ZIP income given log earnings must be approximately linear over the relevant range. Under these conditions, $\hat{\beta}$ provides the relevant mapping from individual earnings gains to expected ZIP gains.[18]

Using this calibration, our estimated increase of roughly 35 percent in late-career occupational income scores implies a ZIP-income effect of $0.35 \times 0.075 \approx 2.6$ percent, close to our direct estimate of 2.4 percent. The analogous calculation for ZIP-level home values, using the corresponding gradient estimate of $\hat{\beta}_{\text{hp}} = 0.076$ (SE: 0.005; Appendix Table A.13, column 2), yields a predicted effect of $0.35 \times 0.076 \approx 2.7$ percent, somewhat below our direct estimate of 4.3 percent but within its 95-percent confidence interval $[1.9\%, 6.7\%]$.

We interpret this comparison as a calibration rather than a formal mediation exercise. The predicted effect of 2.6 percent is close to our estimate of 2.4 percent, but the prediction itself is imprecise because it is based on the imprecise occupational income score estimate. Ignoring sampling variation in $\hat{\beta}$, the 95-percent confidence interval for the predicted ZIP-income effect is roughly $[0.1\%, 5.2\%]$.[19] Thus, while the match is consistent with the earnings channel accounting

[16]The modest magnitude of this gradient reflects substantial heterogeneity in individual resources within ZIP codes. A ZIP code with mean income of, say, \$50,000 contains residents spanning a wide range of individual incomes; consequently, even large differences in individual earnings are associated with only modest differences in the average income of one's neighborhood.

[17]One might wonder about doing the projection the other way around, by regressing log individual earnings on log ZIP income. This "reverse regression" is not the right tool for the reconciliation: OLS projections are asymmetric, and the reverse gradient $\beta_{\text{R}}$ is not the inverse of the "forward" gradient $\beta$ except under strong assumptions likely to be violated in practice. In our data, $\hat{\beta} \times \hat{\beta}_{\text{R}} = R^2 \approx 0.075 \times 0.79 \approx 0.059$, so ZIP income explains only about 6 percent of the cross-sectional variation in individual late-career earnings. Accordingly, multiplying the ZIP effect by $\hat{\beta}_{\text{R}}$ would severely understate the implied individual-level earnings change and is not a valid predictor thereof.

[18]Throughout this section, we treat the occupational income score gain as a proxy for the individual earnings gain. This score is actually the sex-by-race-by-region-specific average income in the individual's occupation from Collins and Wanamaker (2022), not individual earnings. To the extent that the occupational income score itself is an imprecise proxy for individual earnings, the 35-percent figure carries additional uncertainty beyond its standard error.

[19]This is $\hat{\beta}\times$ (95% CI for the occscore effect) $= 0.075 \times [0.01, 0.69]$. The true confidence interval, including the

for most or all of the ZIP gains, it does not rule out meaningful contributions from other channels. At the same time, the comparison does show that the order-of-magnitude gap between the occupational-income and ZIP-based estimates is what one would expect when a treatment-induced change in individual resources is observed through neighborhood-level averages, and that there is no *need* to invoke a large additional channel to reconcile the magnitudes.

In summary, our estimates of sons' gains in terms of late-career occupational income scores and late-life neighborhood characteristics are quantitatively consistent. The calibration relies on common sorting and approximate linearity, and should therefore be viewed as a consistency check rather than a formal decomposition.

**Inferring late-career earnings gains from ZIP-at-death results**

Under an additional condition—that the treatment affects late-life ZIP outcomes only through individual earnings (an "exclusion restriction")—we can invert the forward gradient to infer the individual-level late-career earnings gain implied by the ZIP estimates:

$$\text{Implied } \Delta \log(\text{earnings}) = \frac{\Delta \log(\text{ZIP income})}{\hat{\beta}} = \frac{0.024}{0.075} \approx 32\%.$$

Using the delta method (treating the ZIP effect and HRS gradient as independent, since they come from different samples), this estimate has a standard error of about 11 percentage points and a 95-percent confidence interval of roughly $[11\%, 53\%]$.[20] This is consistent with the direct OCG-based estimate of roughly 35 percent (SE $\approx$ 17 percent, 95% CI: $[2\%, 68\%]$) but substantially more precise, because it draws on the large linked Numident sample rather than the small OCG survey.

The analogous calculation for ZIP-level home values yields a larger implied gain of $0.043/0.076 \approx 57$ percent (SE $\approx$ 16 percent). This is consistent with the ZIP-income-based estimate given the statistical precision. We treat the ZIP-income-based estimate as primary but note that the ZIP-home-value-based estimate provides a useful cross-check and, if anything, suggests that the ZIP-income-based estimate may be conservative.

The exclusion restriction—that the treatment affects late-life ZIP outcomes only through individual earnings—is the strongest requirement for the structural inversion. The forward-calculation match provides some evidence for it, but that evidence is limited by the imprecision of the occscore estimate, as discussed above. The most natural concern about the common-sorting condition—that compliers had unusually back-loaded earnings profiles and therefore sorted into worse neighborhoods than others with similar late-career earnings—would cause the HRS gradient to overstate the complier-specific gradient, biasing the structural inversion conservatively (understating the implied earnings gain).

---

contribution of sampling variation in $\hat{\beta}$, is wider. We ignore sampling variation in $\hat{\beta}$ for this calculation since for our purpose here of assessing whether one estimate is quantitatively consistent with another, a shorter confidence interval is more conservative in that it makes more likely the conclusion that the estimates are *not* quantitatively consistent.

[20] $\text{SE}(g) = |g|\sqrt{(\text{SE}_\delta/\hat{\delta})^2 + (\text{SE}_\beta/\hat{\beta})^2} = 0.32 \times \sqrt{(0.008/0.024)^2 + (0.003/0.075)^2} \approx 0.107.$

## E Details of the implications calculations (Section 6)

Section 6 reports several quantities derived from the estimated late-career earnings gains: the full-career average earnings gain, the expected present value (EPV) of lifetime earnings gains, the ratio of children's gains to Social Security benefits, and fiscal effects. Here we describe these calculations, including how sampling uncertainty in the key estimates propagates to the "downstream" results.

The calculations from the late-career earnings gain $g$ to downstream quantities proceed as follows:

1. Late-career to full-career earnings gains (percent and EPV): As described in Section 6.1, we linearly interpolate between the estimated occupational income score effects at different career stages: no gain at first job and in 1950, and the full late-career gain in 1962 and 1973, for sons of fathers born around 1877. The baseline earnings path is median annual earnings for male four-quarter workers from the Social Security Bulletin Annual Statistical Supplements (Social Security Administration, 1965, 1975). We assume sons were born in 1910, and we discount the implied earnings and income gains to reflect time discounting (3% annual real interest rate, discounted to 1950) and mortality (1910 cohort life table for males from the SSA). This implies a ratio of late-career to full-career average earnings gains of about 2.4 (35 percent / 14.8 percent) and an EPV gain (in 2020 dollars, discounted to 1950) of about $160,000 per son.

2. Collective gains: Per-son gains are multiplied by the average number of sons per family: 2.3.

3. Ratio to Social Security benefits: Collective gains are divided by the EPV of the difference in Social Security benefits from a ten-year eligibility difference: $\approx$ $38,000.

4. Fiscal effects: Net fiscal cost per dollar of Social Security benefits equals $1 + \rho\tau - r\tau$, where $\rho \approx 1.25$ is the assumed recipient earnings reduction per dollar of benefits, $\tau = 0.25$ is the assumed average marginal tax rate, and $r$ is the ratio from step 4.

We use three complementary approaches to estimate sons' late-career earnings gains $g$. The first uses the occupational income score estimates, yielding a point estimate of roughly 35 percent (SE $\approx$ 17 percent). The second uses the late-life ZIP income estimate together with the cross-sectional relationship between late-life ZIP income and late-career earnings, yielding an implied late-career gain of roughly 32 percent (SE $\approx$ 11 percent). The third uses the late-life ZIP home price estimate together with the cross-sectional relationship between late-life ZIP home prices and late-career earnings, yielding an implied late-career gain of roughly 57 percent (SE $\approx$ 16 percent). The first two estimates are in close agreement. The third is somewhat higher in point estimate but consistent with the others given the precision. The occupational income score approach is most direct but least precise, while the late-life ZIP approaches are more precise but require the additional assumption that the treatment affects late-life neighborhood characteristics only through individual earnings. All three rely on different data and assumptions, so their broad accordance increases confidence that the late-career gain is in the range of roughly 30 to 35 percent, with the ZIP-home-value-based estimate suggesting this may be conservative.

Since all downstream quantities are approximately linear in the late-career gain $g$, standard errors scale proportionally through the chain. Under our primary specification (ZIP-income-based

late-career gain with the linear-interpolation mapping), the key downstream quantities and their 95-percent confidence intervals (reflecting sampling uncertainty only) are:

- Full-career average earnings gain: $\approx 14\%$ (95% CI: $[5\%, 23\%]$)
- EPV of lifetime earnings gains per son: $\approx$ \$146,000 (95% CI: [\$50,000, \$243,000])
- Ratio of children's collective gains to Social Security benefits: $\approx 9$ (95% CI: $[3, 15]$)
- Net fiscal cost per dollar of Social Security benefits: $\approx -\$0.91$ (95% CI: $[-\$2.38, \$0.55]$)

Under the OCG-based late-career gain, the point estimates are somewhat larger and the confidence intervals are much wider (e.g., ratio $\approx 10$, 95% CI: $[0.5, 19]$).

Two important caveats apply. First, the reported confidence intervals capture only sampling uncertainty in the estimated treatment effects and HRS gradient. They do not account for additional uncertainty from the interpolation scheme, the baseline earnings path, the Social Security benefit simulation, or, for the occupational income scores-based estimate, the mapping from occupational income scores to earnings. The true uncertainty is therefore larger than what the confidence intervals suggest. Second, the fiscal calculations additionally assume a 25-percent marginal tax rate and a 25-percent reduction in recipients' labor supply after the eligibility age, both of which are approximate. For the labor supply reversal—the result that these Social Security expansions increased total family earnings—the key threshold is whether children's earnings gains exceed recipients' earnings losses. Under our assumed value of recipients' earnings losses, this holds across the 95-percent confidence interval even under the most conservative specification. But the result that these Social Security expansions paid for themselves, i.e., had negative cost to the government, does not hold under the most conservative bound of the confidence interval in the most conservative specification.

# F Potential efficiency gains from Social Security smoothing the costs of old-age support

Social Security may have facilitated the exploitation of gains from trade in old-age support that were otherwise foregone due to missing markets and transaction costs. The key idea is that Social Security smoothed the financial component of old-age support both across individuals and over time, transforming it from being highly concentrated (often borne mainly by just one or two children during a relatively brief spell) to more dispersed (borne partly in the form of modest tax payments by all workers throughout their working years). This smoothing, in turn, likely reduced the opportunity cost of migrating away from one's parents and thereby facilitated a more efficient allocation of people across places.

Without Social Security, the costs of old-age support were highly concentrated for three reinforcing reasons. First, missing markets—especially the lack of insurance markets for hedging the risk in how much old-age support will be demanded of a particular individual—greatly limited the extent to which old-age support costs could be spread across individuals, especially across families but also within them. Historical financial markets were less developed than today, and instruments for insuring or trading idiosyncratic old-age support risk were essentially nonexistent (on the limited development of financial markets in this period, see, e.g., Bleakley and Ferrie, 2016). Second,

capital market imperfections limited the extent to which individuals could smooth old-age support costs over time, especially by borrowing against future income to finance current support. This constraint was particularly binding given that the gains from migration are frequently back-loaded (see, e.g., Nakamura et al., 2022), meaning that much of sons' gains from migrating would have accrued after their parents died, making it difficult to use those gains to finance old-age support.[21] Third, key features of old-age support technologies further concentrated costs: There are large economies of scale in providing support through shared housing or shared location, which are most efficiently provided by one or few providers, creating strong incentives for at least one child to live near the parents even at the expense of being poorly matched to the local labor market. The resulting concentration of old-age support was considerably greater than it likely would be in the first-best allocation with more risk sharing and a better matching of individuals to labor markets, so the equilibrium without Social Security was likely highly inefficient.

Social Security made the financial component of old-age support much more uniform by having virtually all working-age individuals pay a modest share of their earnings as taxes throughout their working lives, independent of the level and timing of their parents' demand for support. This smoothing reduced the costs borne by would-be primary providers of family old-age support, thereby reducing their gains from living near their parents during old age and so reducing their opportunity cost of migration. In this way, Social Security relaxed for at least some children a "location constraint" requiring them to live near their elderly parents, potentially enabling them to sort to labor markets that better matched their skills. Our evidence suggests this was a major efficiency benefit of Social Security.[22]

That Social Security facilitated a more efficient allocation is consistent with our finding that Social Security significantly increased the total family earnings of recipients and their children. This raises the question: If the gains from migration were so large, why didn't families find ways to capture them even without Social Security, while ensuring elderly members were supported? While answers are necessarily speculative given the lack of direct evidence, we consider several factors beyond missing markets that may have contributed.

First, families may have underestimated the gains from migration. Quasi-experimental and structural studies in many contexts find that migration generates large financial gains for those induced to migrate (e.g., Bryan et al., 2014; Lagakos et al., 2023). While a leading explanation is that migration costs are very large, a complementary explanation is that individuals may underestimate migration's benefits. To the extent that families were migrating less than efficiently without Social Security, Social Security-induced migration would have direct efficiency consequences beyond the fiscal externalities from higher earnings and so higher net tax payments (since it has a first-order effect on the families themselves).

Second, norms or transaction costs within families may have limited families' ability to exploit

[21] Another important missing market was that for insuring the risky returns to migration. Migration is costly, risky, and hard to insure (Munshi and Rosenzweig, 2016; Meghir et al., 2022), so even if it pays off handsomely on average, it may not pay off enough to finance old-age support in certain states of the world.

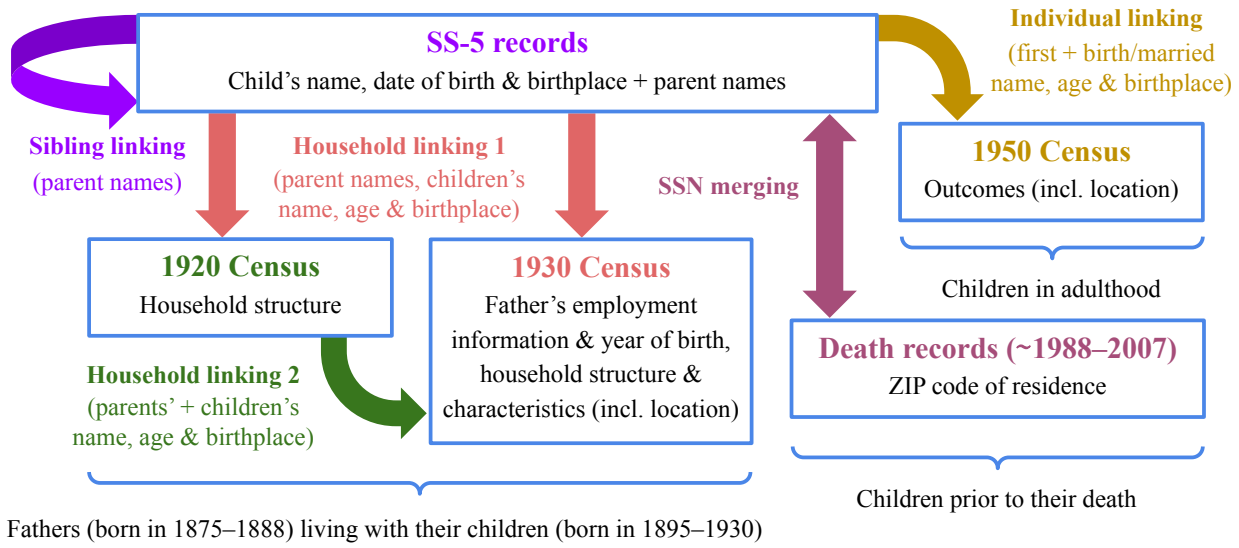
[22] Our evidence of large, migration-driven gains to recipients' children from Social Security is suggestive of a large gap between the equilibrium without Social Security and the first best. That co-residence appears to be an inferior good (Rosenzweig and Wolpin, 1993) reinforces this idea. Note that Social Security did not *force* families to smooth old-age support costs; the average family retained their no-Social Security allocation in their with-Social Security choice set. That families chose a smoother distribution with Social Security than without is suggestive that such a distribution is more efficient—i.e., that Social Security effectively created valuable otherwise-missing markets for trade in old-age support.

gains from trade. For example, norms might have favored local, in-kind old-age support, such as a norm that at least one child stays near the parents. A transaction cost rationale for such a norm is that non-local children will not provide much old-age support, perhaps due to lower connection with their parents. While there is a force toward efficiency-promoting norms, norms that promoted efficiency in one environment may not continue to do so as the economic landscape changes (Bau, 2021).

We emphasize that these considerations are speculative. The extent to which the efficiency gain came from substituting for missing markets versus correcting mistakes does not affect its size but is relevant for understanding its generalizability to contexts in which key markets or beliefs may differ. This is an important question for future research.

## Appendix Figures and Tables

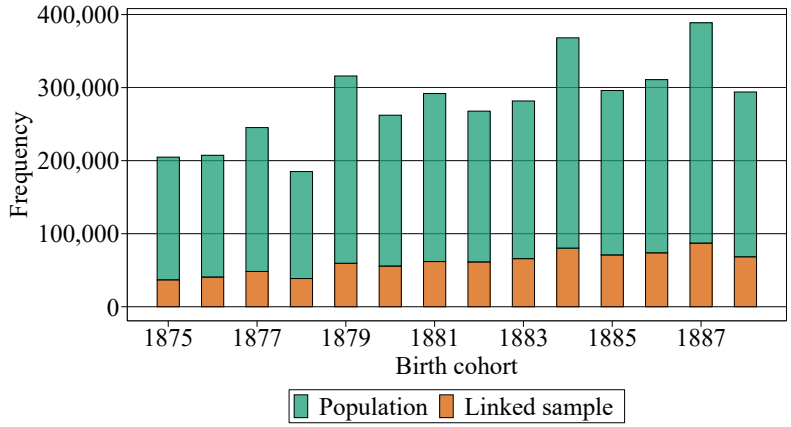
Figure A.1: Overview of Numident death-census linked sample construction



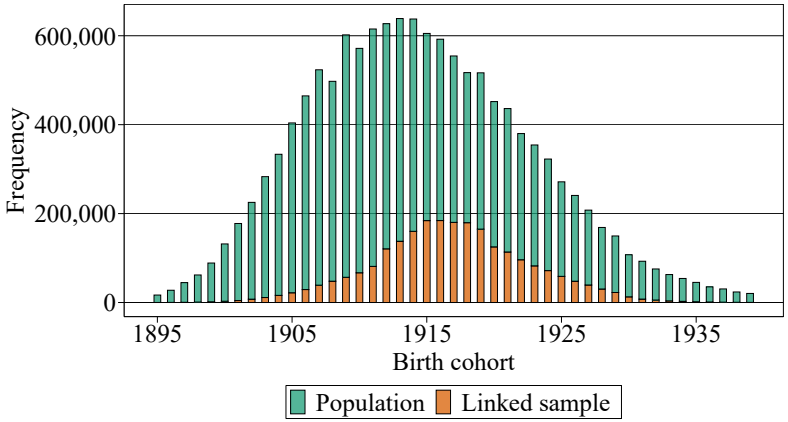

*Notes:* The key variables of interest from each dataset are indicated in the boxes. The variables used for each type of linkage are indicated in parentheses.

Figure A.2: Coverage of men in our population of interest and their children in Numident death-census linked sample

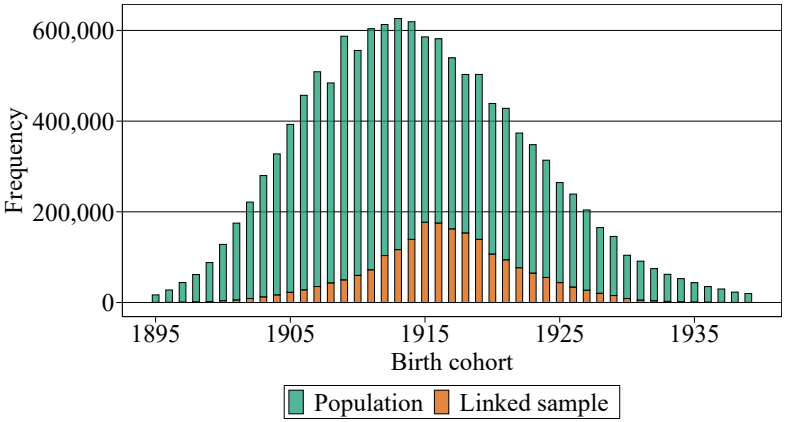
(a) Men born in 1875–1888 covered by Social Security in 1935 or 1950 based on 1930 employment




(b) Sons of men born in 1875–1888

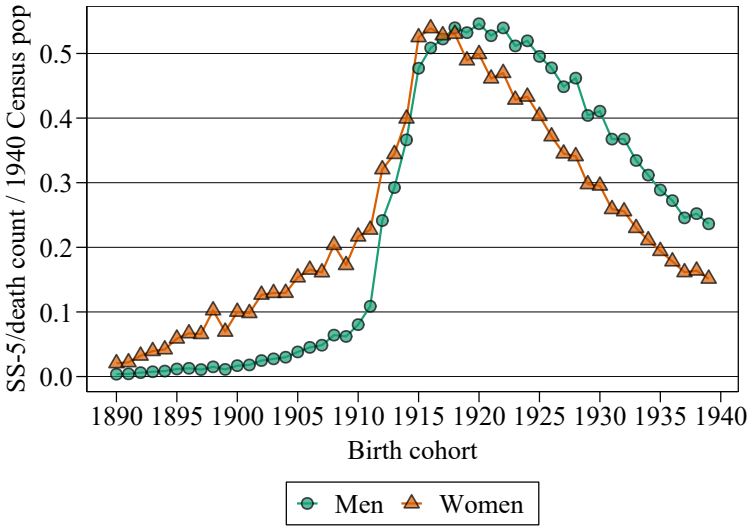



(c) Daughters of men born in 1875–1888



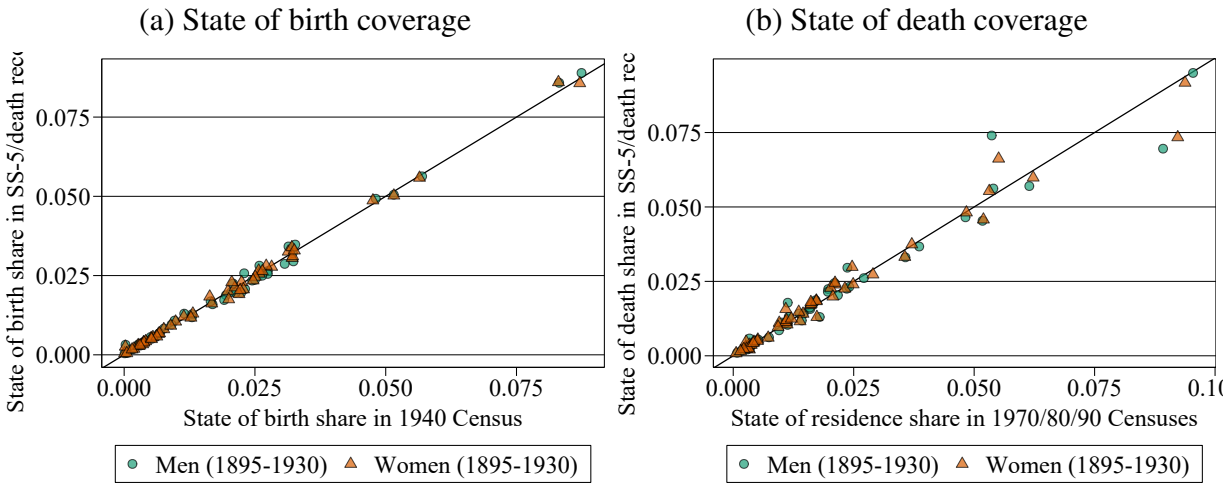

*Notes:* Population in panel (a): men born in 1875–1888 in the 1930 Census who were covered by Social Security in 1935 or 1950 based on their employment information (excluding the self-employed). Linked sample in panel (a): fathers in Numident death-census linked sample satisfying analogous restrictions. Population in panels (b)–(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 samples in panels (b)–(c): sons/daughters born in 1895–1939 in Numident death-census linked sample whose fathers were born in 1875–1888.

Figure A.3: Cohort coverage of Numident records by sex




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

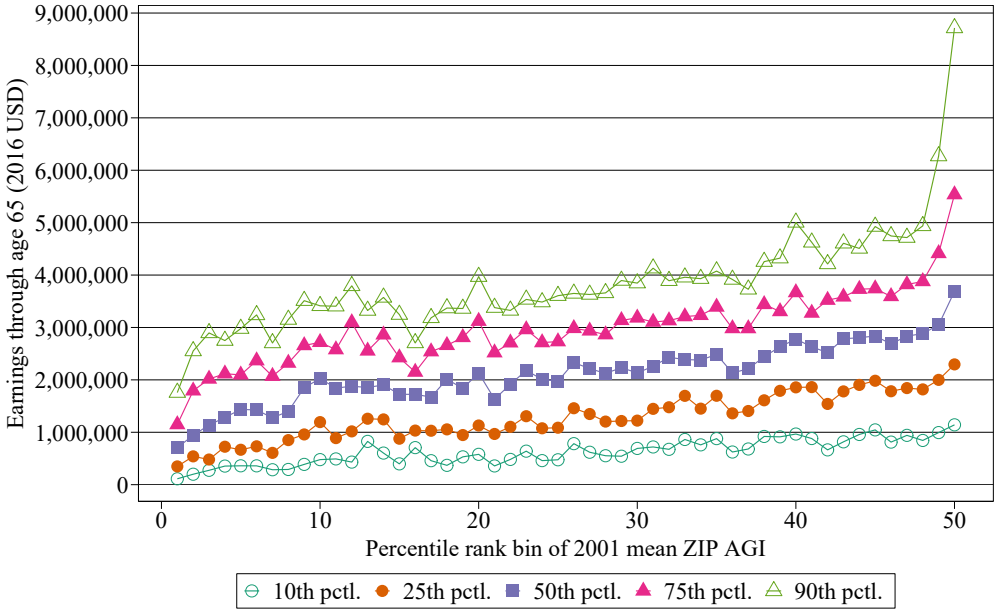
Figure A.4: State of birth and state of death coverage of Numident records by sex



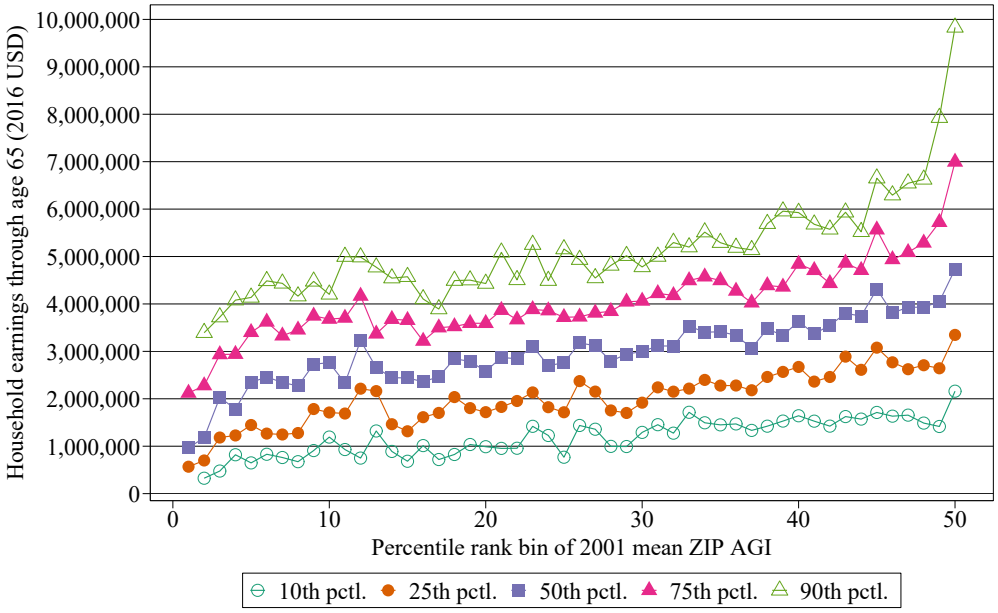

*Notes:* Panel (a) plots state of birth shares among SSNs that can be found in Numident SS-5 and death records against the corresponding shares in the 1940 Census for individuals born between 1895 and 1930, separately by gender (US-born individuals only). For the same set of cohorts and SSNs, panel (b)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

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

(a) Men, own lifetime earnings

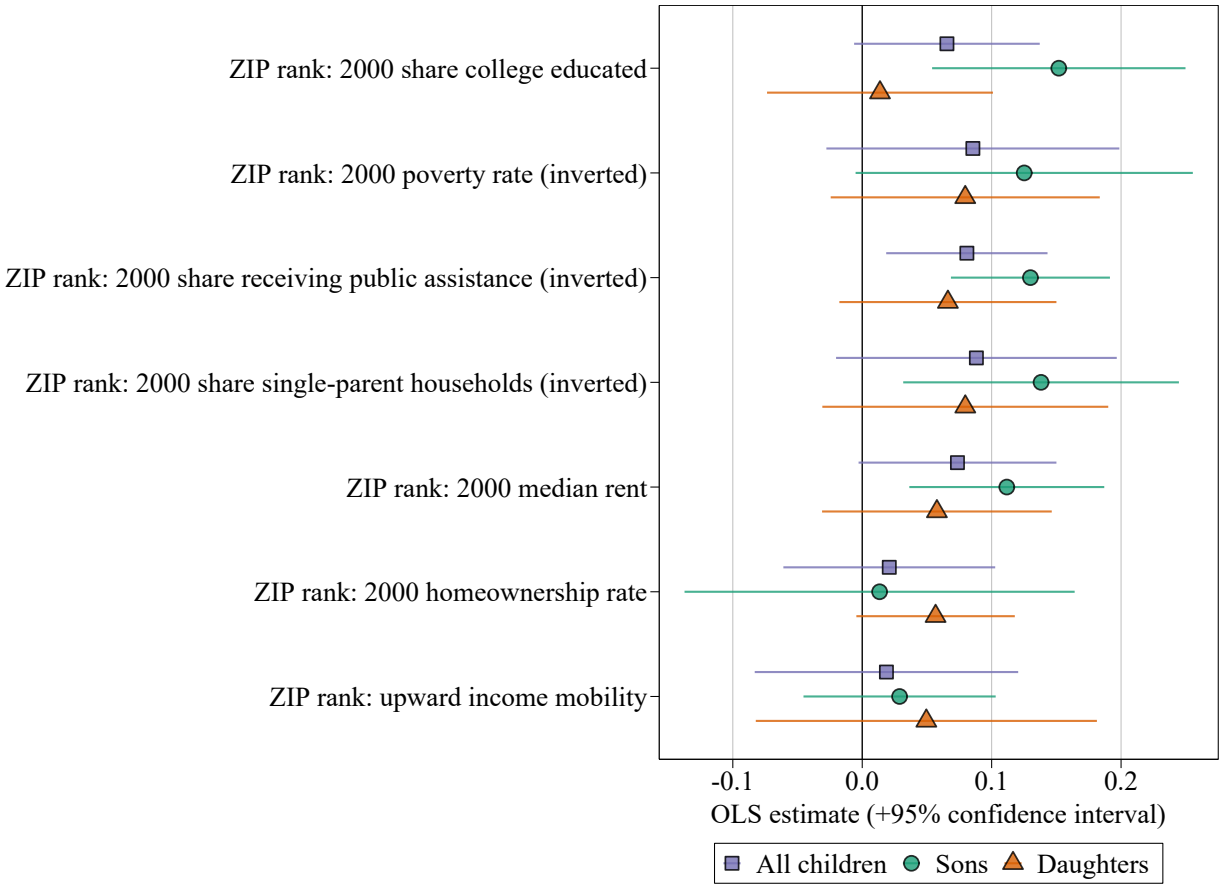



(b) Women, household lifetime earnings



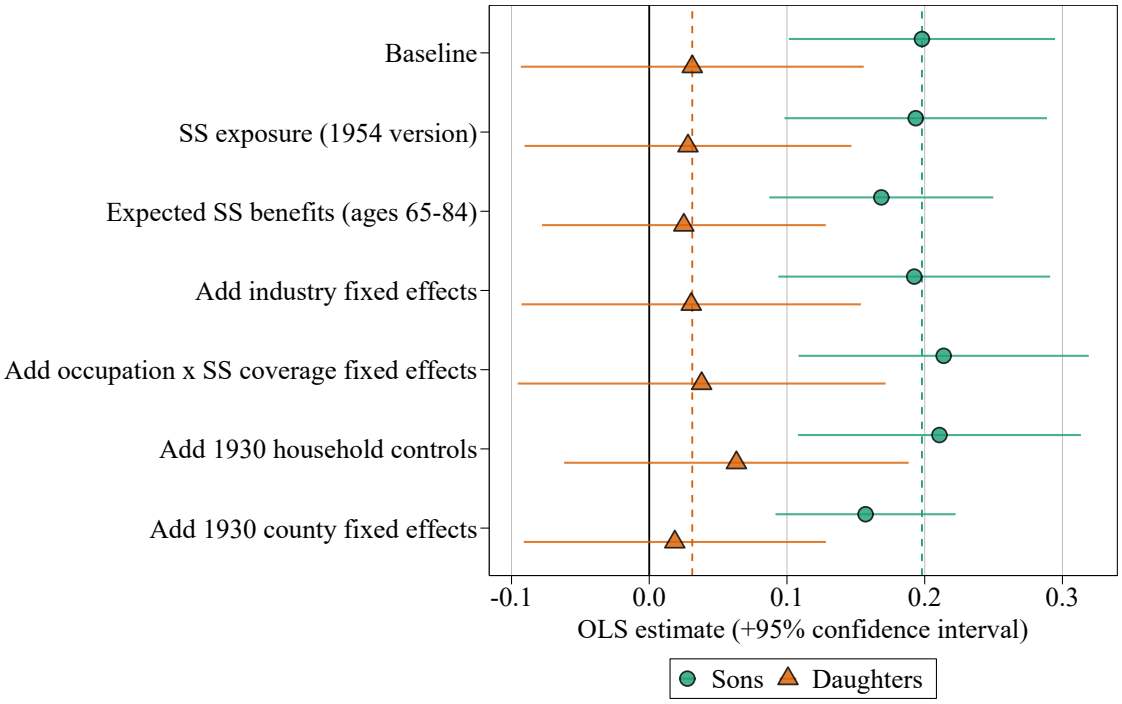

*Notes:* In this figure, the sample consists of HRS respondents aged 65 or older when they were last interviewed and for which we have ZIP of residence and SSA earnings information (see Appendix A.5 for details). Panel (a) plots moments of the distribution of lifetime earnings through age 65 for men by the rank of their last known ZIP code of residence, where ZIP codes are ranked according to mean AGI in 2001 and categorized into 50 bins each containing two adjacent percentile ranks (e.g., the 1st bin includes the 1st and 2nd percentiles, etc.). Panel (b) analogously plots moments of the household lifetime earnings distribution for women with spouses, where household earnings are defined as the sum of respondents' lifetime earnings through age 65 and that of their spouses.

Figure A.6: The impact of father's exposure to Social Security on children's late-life neighborhood quality in terms of various socioeconomic indicators



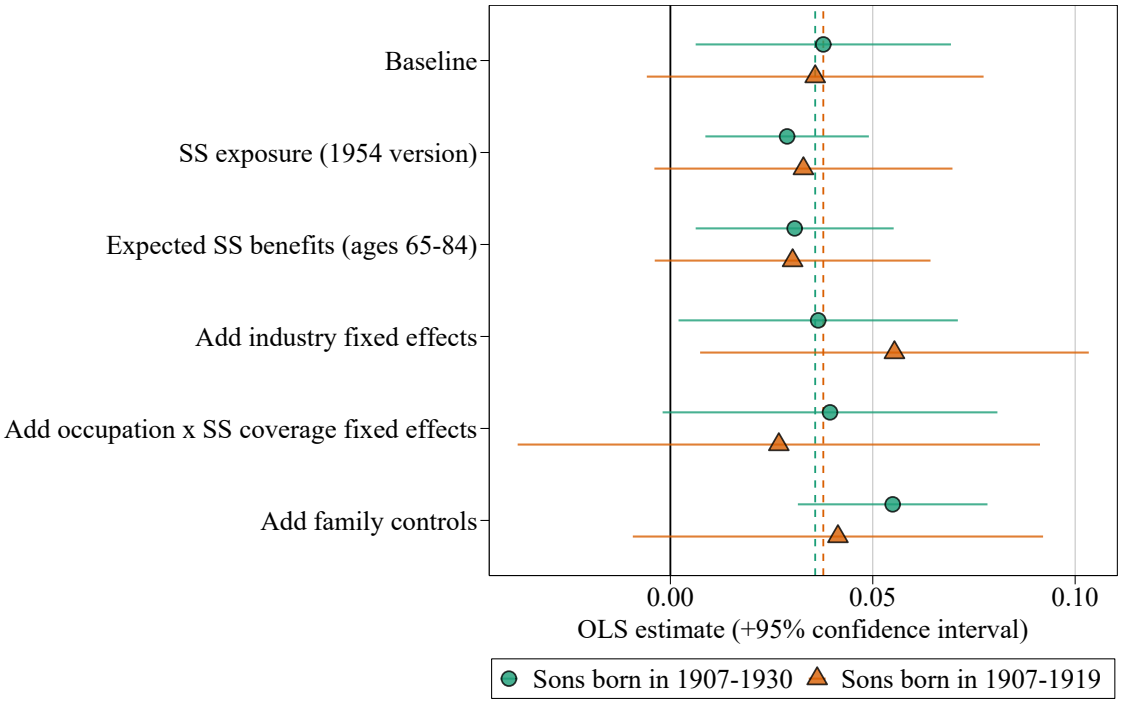

*Notes:* This figure shows the coefficients on father's Social Security exposure (i.e., negative of earliest age of Social Security eligibility) from equation (3), where the sample consists of children in our Numident death-census linked sample and the outcome is the percentile rank of children's late-life ZIP code in terms of various socioeconomic indicators. Signs of the coefficients for the ZIP-level poverty rate, ZIP-level share of households receiving public assistance, and ZIP-level share of single-parent families have been inverted so that a positive coefficient means father's Social Security exposure leads children to live in better neighborhoods late in life. Colors/shapes indicate separate estimates for all children, sons and daughters. All regressions include fixed effects for children's sex-by-birth cohort and fathers' 3-digit occupation-by-birth cohort, Social Security coverage category, race, and birthplace. Observations weighted using inverse propensity score weights that re-weight the sample to match the population of interest in 1930. 95% confidence intervals based on robust standard errors clustered by fathers' 3-digit industry.

Figure A.7: The impact of father's exposure to Social Security on children's late-life ZIP rank in terms of 2001 AGI: Robustness checks



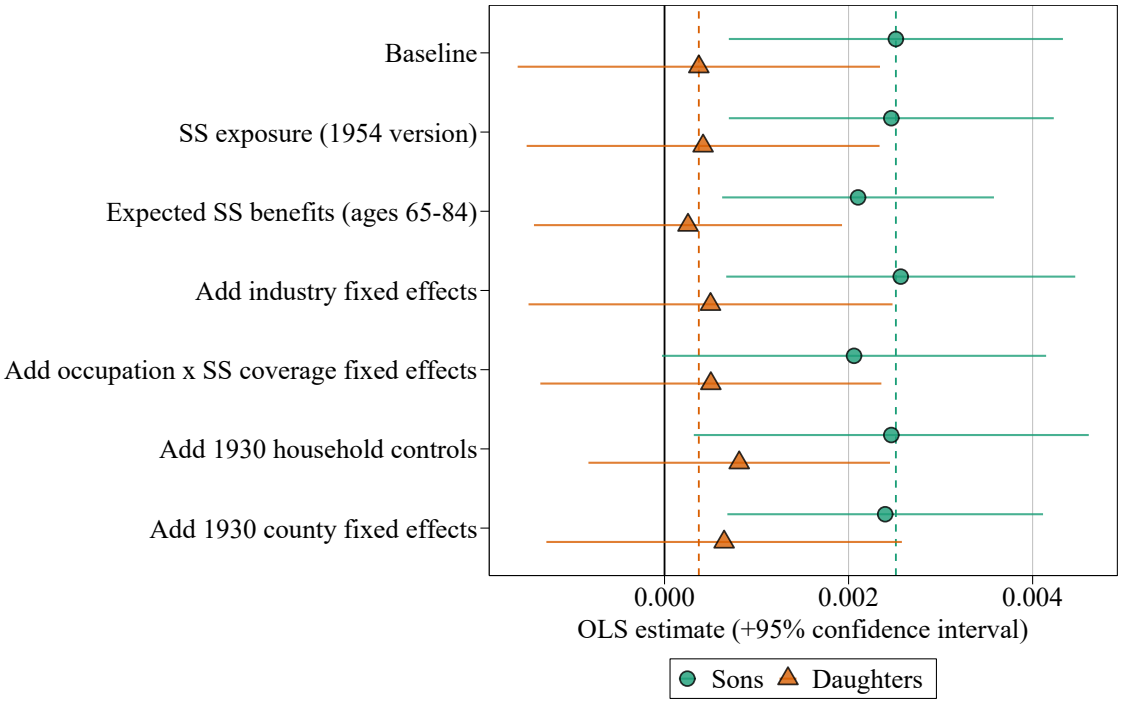

*Notes:* The top row in this figure shows the baseline coefficients from Figure 7 (where the outcome is late-life ZIP 2001 AGI rank). The next two rows show estimates using alternative measures of Social Security exposure (variant of the baseline measure that assumes eligibility starts in 1954 instead of 1950 for 1950-covered workers, or expected average monthly benefits between ages 65 and 84 based on the earnings history of a median earner, in tens of 2000 dollars). The bottom four rows show estimates under variants of the baseline specification that include various additional controls (3-digit industry fixed effects, 3-digit occupation-by-Social Security coverage category fixed effects, fixed effects for all outcomes in Table 3 except house values which is included as a linear control, or fixed effects for parents' 1930 county of residence). Colors/shapes indicate separate estimates for sons and daughters. All regressions include fixed effects for children's sex-by-birth cohort and fathers' 3-digit occupation-by-birth cohort, Social Security coverage category, race, and birthplace. Observations weighted using inverse propensity score weights that re-weight the sample to match the population of interest in 1930. 95% confidence intervals based on robust standard errors clustered by fathers' 3-digit industry.

Figure A.8: The impact of father's exposure to Social Security on sons' occupational income score in 1962 (log): Robustness checks



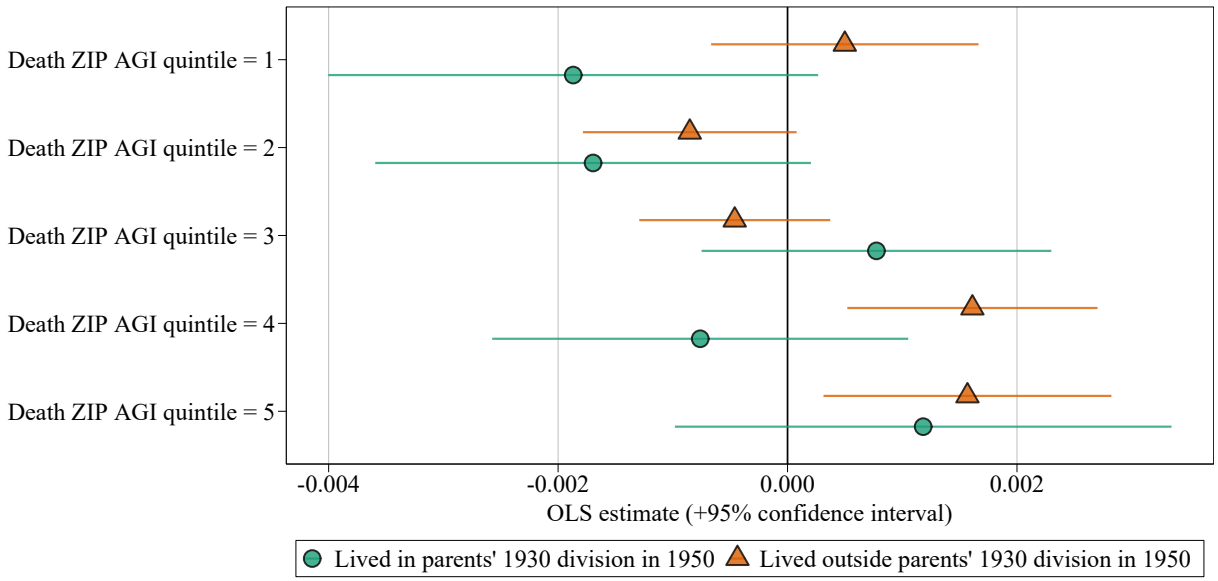

*Notes:* The top row in this figure shows the baseline coefficients from Figure 8 (where the outcome is the log occupational income score associated with sons' jobs in 1962). The next two rows show estimates using alternative measures of Social Security exposure (variant of the baseline measure that assumes eligibility starts in 1954 instead of 1950 for 1950-covered workers or the expected average monthly benefits between ages 65 and 84, in tens of 2000 dollars). The bottom three rows show estimates under variants of the baseline specification that include various additional controls (3-digit industry fixed effects, 3-digit occupation-by-Social Security coverage category fixed effects, or fixed effects for the number of children, both parents' years of education, and bins for family income at age 16). Colors/shapes indicate separate estimates for sons born in 1907–1930 and sons born in 1907–1919. All regressions include fixed effects for sons' birth cohort, race and birthplace, and fathers' 3-digit occupation-by-birth cohort and Social Security coverage category. Observations weighted using OCG sampling weights. 95% confidence intervals based on robust standard errors clustered by fathers' 3-digit industry.

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




*Notes:* The top row in this figure shows the baseline coefficients from Figure 9 (where the outcome is living outside parents' 1930 census division of residence in 1950). The next two rows show estimates using alternative measures of Social Security exposure (variant of the baseline measure that assumes eligibility starts in 1954 instead of 1950 for 1950-covered workers, or expected average monthly benefits between ages 65 and 84 based on the earnings history of a median earner, in tens of 2000 dollars). The bottom four rows show estimates under variants of the baseline specification that include various additional controls (3-digit industry fixed effects, 3-digit occupation-by-Social Security coverage category fixed effects, fixed effects for all outcomes in Table 3 except house values which is included as a linear control, or fixed effects for parents' 1930 county of residence). Colors/shapes indicate separate estimates for sons and daughters. All regressions include fixed effects for children's sex-by-birth cohort and fathers' 3-digit occupation-by-birth cohort, Social Security coverage category, race, and birthplace. Observations weighted using inverse propensity score weights that re-weight the sample to match the population of interest in 1930. 95% confidence intervals based on robust standard errors clustered by fathers' 3-digit industry.

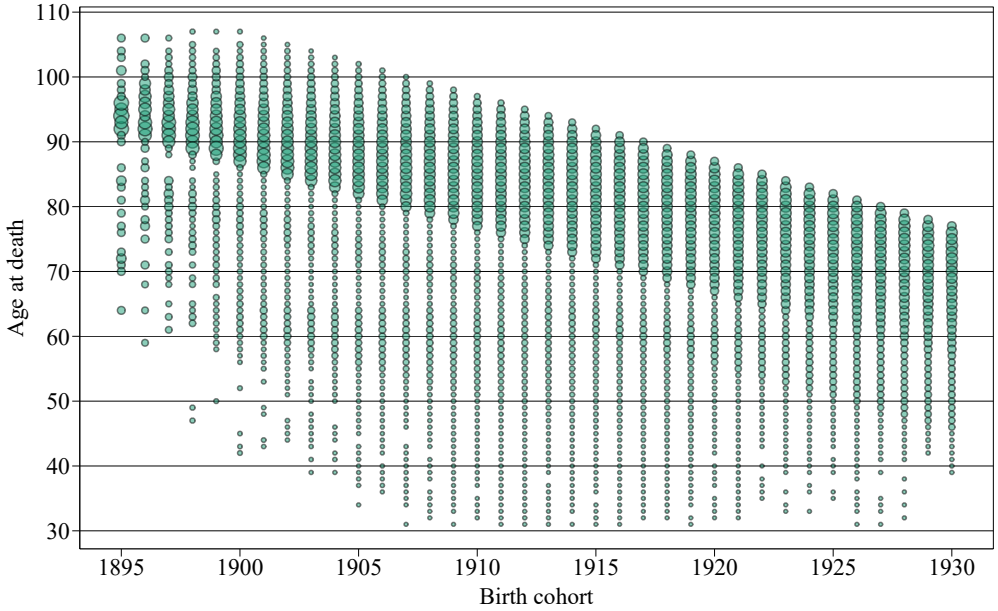
Figure A.10: The impact of father's exposure to Social Security on sons' joint migration status in 1950 and late-life ZIP quintile in terms of 2001 mean AGI



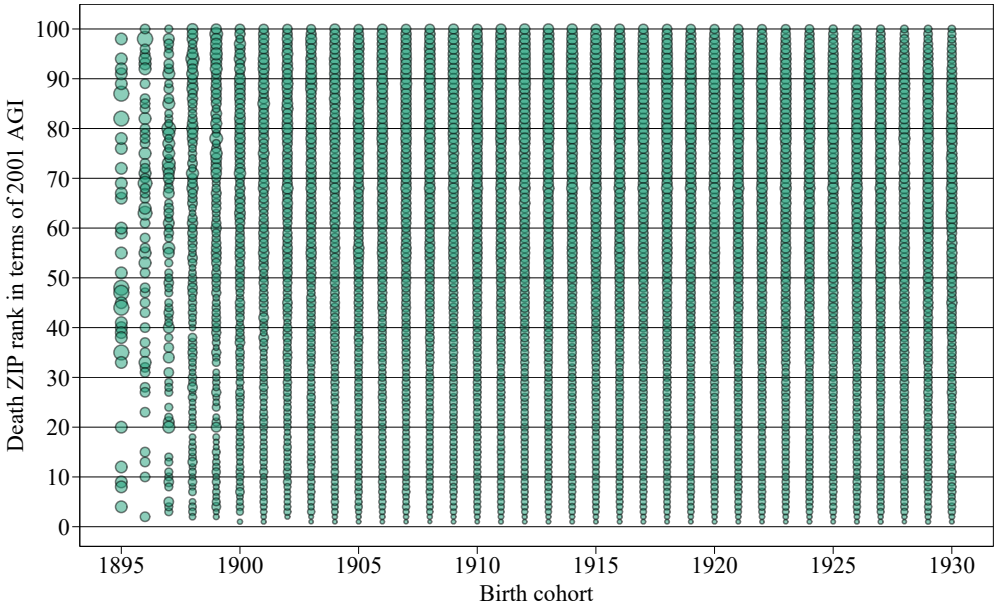

*Notes:* This figure shows the coefficients on father's Social Security exposure (i.e., negative of earliest age of Social Security eligibility) from equation (3), where the sample consists of sons in our Numident death-census linked sample who are additionally linked to the 1950 Census and the dependent variable is an indicator for the joint outcome of living in or outside parents' 1930 census division of residence by 1950 and living in a specified ZIP quintile at the time of death (where ZIP quintiles are based on mean AGI in 2001). All regressions include fixed effects for sons' birth cohort and fathers' 3-digit occupation-by-birth cohort, Social Security coverage category, race, and birthplace. Observations weighted using inverse propensity score weights that re-weight the sample to match the population of interest in 1930. 95% confidence intervals based on robust standard errors clustered by fathers' 3-digit industry.

Figure A.11: Age at death and late-life ZIP AGI rank distribution by birth cohort in Numident death-census linked sample

(a) Age at death distribution



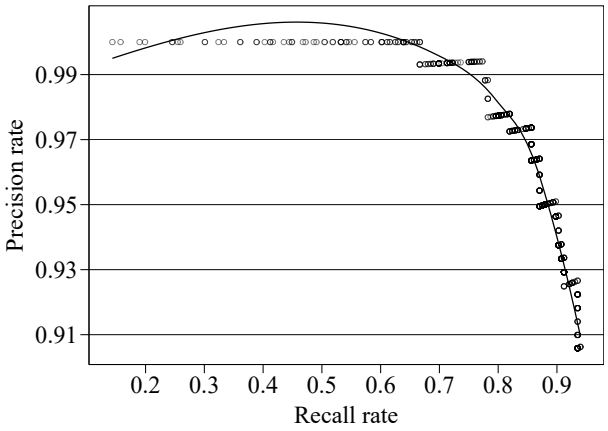

(b) Late-life ZIP AGI rank distribution



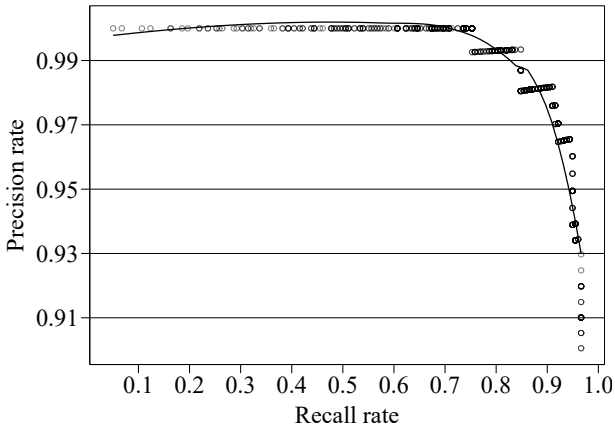

*Notes:* Panel (a) plots the age at death distribution by birth cohort among children in our Numident death-census linked sample (where the size of each dot is proportional to the cohort-specific share). Panel (b) similarly plots the distribution of late-life ZIP AGI percentile ranks by birth cohort among children our Numident death-census linked sample.

Figure A.12: Model performance: Precision-recall curves

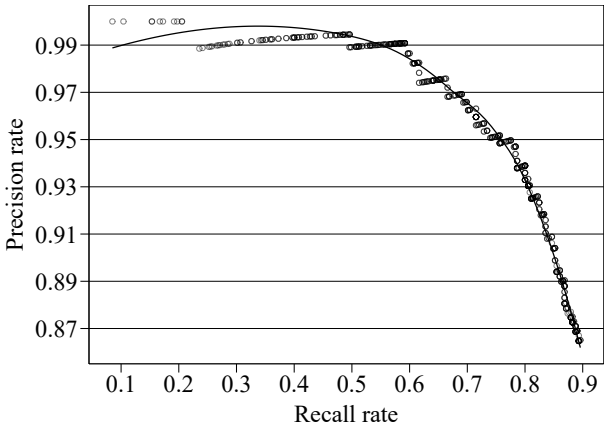
(a) Numident SS-5-1930 Census linking



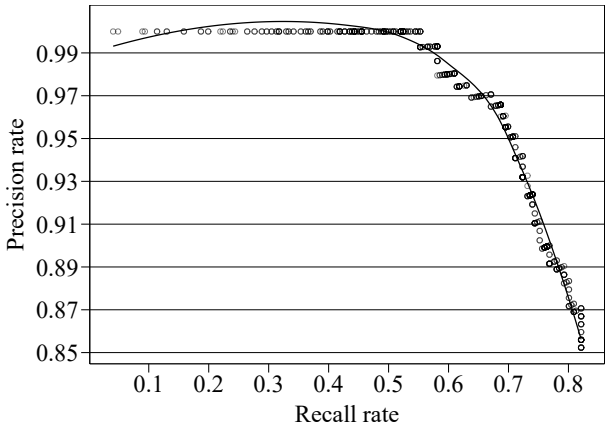

(b) 1930-1920 Census linking




(c) Numident claims-1930 Census linking (records with non-missing birthplace)




(d) Numident claims-1930 Census linking (records with missing birthplace)




*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 | 0.074 | 0.028 | 0.072 | -0.046 | -0.002 |
| | | | | [0] | [0.001] |
| Born in 1875–1879 | 0.296 | 0.264 | 0.297 | -0.032 | 0.001 |
| | | | | [0] | [0.161] |
| Born in 1880–1884 | 0.375 | 0.383 | 0.375 | 0.007 | -0.001 |
| | | | | [0] | [0.461] |
| Born in 1885–1888 | 0.329 | 0.354 | 0.329 | 0.025 | 0 |
| | | | | [0] | [0.499] |
| Born abroad | 0.309 | 0.302 | 0.309 | -0.007 | 0 |
| | | | | [0] | [0.59] |
| Born in Northeast | 0.203 | 0.221 | 0.203 | 0.017 | 0 |
| | | | | [0] | [0.564] |
| Born in Midwest | 0.25 | 0.274 | 0.25 | 0.024 | 0 |
| | | | | [0] | [0.81] |
| Born in South | 0.215 | 0.182 | 0.215 | -0.034 | -0.001 |
| | | | | [0] | [0.198] |
| Born in West | 0.023 | 0.022 | 0.023 | -0.001 | 0 |
| | | | | [0.002] | [0.623] |
| Literate | 0.941 | 0.955 | 0.942 | 0.015 | 0.001 |
| | | | | [0] | [0.001] |
| Covered by Social Security in 1935 | 0.893 | 0.899 | 0.891 | 0.005 | -0.003 |
| | | | | [0] | [0] |
| Covered by Social Security in 1950 | 0.107 | 0.101 | 0.109 | -0.005 | 0.003 |
| | | | | [0] | [0] |
| Farmer | 0.002 | 0.003 | 0.003 | 0 | 0 |
| | | | | [0] | [0.041] |
| Unskilled worker | 0.265 | 0.234 | 0.26 | -0.031 | -0.005 |
| | | | | [0] | [0] |
| Skilled/semi-skilled worker | 0.474 | 0.506 | 0.492 | 0.032 | 0.018 |
| | | | | [0] | [0] |
| White-collar worker | 0.259 | 0.257 | 0.245 | -0.001 | -0.013 |
| | | | | [0.014] | [0] |
| *(table continues on next page)* | | | | | |
| *N* | 3,920,008 | 858,612 | | | |

*Notes:* Columns (3) and (5) weighted using inverse propensity score weights. *p*-values for two-sided *t*-test of equality of means in brackets. Sample restricted to ever-married men who are household heads (or spouses of heads) in 1930.

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 | 0.369 | 0.387 | 0.378 | 0.018 | 0.009 |
| | | | | [0] | [0] |
| Live in Midwest | 0.321 | 0.346 | 0.33 | 0.025 | 0.009 |
| | | | | [0] | [0] |
| Live in South | 0.202 | 0.176 | 0.2 | -0.027 | -0.002 |
| | | | | [0] | [0] |
| Live in West | 0.108 | 0.092 | 0.092 | -0.016 | -0.016 |
| | | | | [0] | [0] |
| Live on farm | 0.051 | 0.06 | 0.051 | 0.009 | 0 |
| | | | | [0] | [0.535] |
| Live in urban area | 0.739 | 0.712 | 0.737 | -0.027 | -0.002 |
| | | | | [0] | [0.018] |
| Own home | 0.488 | 0.552 | 0.492 | 0.064 | 0.004 |
| | | | | [0] | [0] |
| Own a radio | 0.475 | 0.509 | 0.477 | 0.034 | 0.002 |
| | | | | [0] | [0.005] |
| *N* | 3,920,008 | | 858,612 | | |

*Notes:* Columns (3) and (5) weighted using inverse propensity score weights. *p*-values for two-sided *t*-test of equality of means in brackets. Sample restricted to ever-married men who are household heads (or spouses of heads) in 1930.

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 | 0.074 | 0.026 | 0.076 | -0.048 | 0.002 |
| | | | | [0] | [0.01] |
| Born in 1875–1879 | 0.296 | 0.357 | 0.287 | 0.062 | -0.009 |
| | | | | [0] | [0] |
| Born in 1880–1884 | 0.375 | 0.392 | 0.374 | 0.017 | -0.001 |
| | | | | [0] | [0.148] |
| Born in 1885–1888 | 0.329 | 0.213 | 0.339 | -0.116 | 0.01 |
| | | | | [0] | [0] |
| Born abroad | 0.309 | 0.268 | 0.311 | -0.041 | 0.002 |
| | | | | [0] | [0.047] |
| Born in Northeast | 0.203 | 0.24 | 0.202 | 0.037 | -0.001 |
| | | | | [0] | [0.187] |
| Born in Midwest | 0.25 | 0.311 | 0.248 | 0.062 | -0.002 |
| | | | | [0] | [0.02] |
| Born in South | 0.215 | 0.155 | 0.216 | -0.06 | 0.001 |
| | | | | [0] | [0.35] |
| Born in West | 0.023 | 0.025 | 0.023 | 0.002 | 0 |
| | | | | [0] | [0.731] |
| Literate | 0.941 | 0.969 | 0.941 | 0.028 | 0 |
| | | | | [0] | [0.853] |
| Covered by Social Security in 1935 | 0.893 | 0.938 | 0.924 | 0.044 | 0.03 |
| | | | | [0] | [0] |
| Covered by Social Security in 1950 | 0.107 | 0.062 | 0.076 | -0.044 | -0.03 |
| | | | | [0] | [0] |
| Farmer | 0.002 | 0.002 | 0.002 | -0.001 | 0 |
| | | | | [0] | [0.003] |
| Unskilled worker | 0.265 | 0.195 | 0.241 | -0.07 | -0.024 |
| | | | | [0] | [0] |
| Skilled/semi-skilled worker | 0.474 | 0.527 | 0.513 | 0.053 | 0.039 |
| | | | | [0] | [0] |
| White-collar worker | 0.259 | 0.277 | 0.244 | 0.018 | -0.014 |
| | | | | [0] | [0] |
| *(table continues on next page)* | | | | | |
| *N* | 3,920,008 | | 367,732 | | |

*Notes:* Columns (3) and (5) weighted using inverse propensity score weights. *p*-values for two-sided *t*-test of equality of means in brackets. Sample restricted to ever-married men who are household heads (or spouses of heads) in 1930.

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 | 0.369 | 0.374 | 0.364 | 0.005 | -0.005 |
| | | | | [0] | [0] |
| Live in Midwest | 0.321 | 0.37 | 0.338 | 0.049 | 0.017 |
| | | | | [0] | [0] |
| Live in South | 0.202 | 0.142 | 0.189 | -0.06 | -0.013 |
| | | | | [0] | [0] |
| Live in West | 0.108 | 0.114 | 0.109 | 0.006 | 0.001 |
| | | | | [0] | [0.065] |
| Live on farm | 0.051 | 0.038 | 0.051 | -0.013 | 0 |
| | | | | [0] | [0.647] |
| Live in urban area | 0.739 | 0.767 | 0.74 | 0.028 | 0 |
| | | | | [0] | [0.636] |
| Own home | 0.488 | 0.556 | 0.488 | 0.067 | 0 |
| | | | | [0] | [0.763] |
| Own a radio | 0.475 | 0.562 | 0.473 | 0.087 | -0.001 |
| | | | | [0] | [0.184] |
| *N* | 3,920,008 | | 367,732 | | |

*Notes:* Columns (3) and (5) weighted using inverse propensity score weights. *p*-values for two-sided *t*-test of equality of means in brackets. Sample restricted to ever-married men who are household heads (or spouses of heads) in 1930.

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 emp. categories (excl. public emergency), male workers aged 25–64* | | |
| Covered in 1935 based on 1940 employment | 9.3 | 90.7 |
| Covered in 1950 or later based on 1940 employment | 73 | 27 |
| *Panel (b): Emp. categories in analysis (excl. public emergency), male workers aged 51–64* | | |
| Covered in 1935 based on 1940 employment | 12.8 | 87.2 |
| Covered in 1950 based on 1940 employment | 75.5 | 24.5 |

*Notes:* Panel (a) shows the share of male workers aged 25–64 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. 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. We exclude public emergency workers from this analysis as their coverage status is hard to determine.

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

| | Men born in 1875–1888 in 1930 Census | | | | | | |
|---|---|---|---|---|---|---|---|
| | Population | | | Death-census linked sample | | Claims-census linked sample | |
| Occupation | Number (1) | Share covered in 1935 (%) (2) | Share (%) (3) | Share covered in 1935 (%) (4) | Share (%) (5) | Share covered in 1935 (%) (6) | Share (%) (7) |
| Laborers (n.e.c.) | 586,498 | 90.6 | 14.96 | 90.8 | 14.95 | 93.4 | 14.28 |
| Managers, officials, and proprietors (n.e.c.) | 289,471 | 98.3 | 7.38 | 98 | 7.21 | 98.7 | 6.97 |
| Carpenters | 204,143 | 98.5 | 5.21 | 98.5 | 5.51 | 98.7 | 5.58 |
| Foremen (n.e.c.) | 119,435 | 98.4 | 3.05 | 98.2 | 3.28 | 99.2 | 3.38 |
| Clerical and kindred workers (n.e.c.) | 73,422 | 92.3 | 1.87 | 92 | 1.77 | 95.2 | 1.87 |
| Painters, construction and maintenance | 69,527 | 97.2 | 1.77 | 97.4 | 1.66 | 97.4 | 1.69 |
| Truck and tractor drivers | 68,466 | 95.4 | 1.75 | 95.4 | 1.86 | 96.5 | 1.82 |
| Stationary engineers | 63,644 | 92.7 | 1.62 | 92.6 | 1.64 | 95.4 | 1.65 |
| Janitors and sextons | 57,344 | 56.9 | 1.46 | 51.9 | 1.5 | 67.3 | 1.43 |
| Plumbers and pipe fitters | 38,794 | 96.8 | 0.99 | 96.6 | 1 | 97.5 | 1 |
| Electricians | 34,909 | 95.3 | 0.89 | 94.7 | 0.88 | 97.2 | 0.93 |
| Mechanics and repairmen (n.e.c.) | 30,130 | 97.1 | 0.77 | 97.1 | 0.77 | 98.1 | 0.84 |
| Teachers (n.e.c.) | 28,506 | 4.3 | 0.73 | 3.4 | 0.83 | 5.8 | 0.44 |
| Stationary firemen | 28,068 | 93.6 | 0.72 | 93.5 | 0.79 | 95.6 | 0.77 |
| Guards, watchmen, and doorkeepers | 27,065 | 88.1 | 0.69 | 88.2 | 0.72 | 91.8 | 0.63 |
| Accountants and auditors | 23,524 | 94.8 | 0.6 | 94.8 | 0.55 | 95.6 | 0.61 |
| Bookkeepers | 23,382 | 96.2 | 0.6 | 96.1 | 0.58 | 97.1 | 0.65 |
| Taxicab drivers and chauffers | 23,062 | 43.7 | 0.59 | 43.8 | 0.55 | 41.3 | 0.56 |
| Cooks, except private household | 22,531 | 95.9 | 0.57 | 94.6 | 0.36 | 95.8 | 0.52 |
| Gardeners, except farm and groundskeepers | 22,236 | 58 | 0.57 | 57.8 | 0.55 | 60.1 | 0.46 |

*Notes:* This table shows summary statistics for the top 20 occupations with some variation in Social Security coverage year in the 1930 Census (defined as occupations with a share of workers covered in 1935 between 1 percent and 99 percent). Population consists of men born in 1875–1888 whose current employment was covered in 1935 or 1950 (excluding the self-employed). Columns 4 and 5 restrict the sample to men who appear in our Numident death-census linked sample. Columns 6 and 7 restrict the sample to men who appear in our Numident claims-census linked sample. Observations in columns 4–7 weighted using inverse propensity score weights that re-weight the sample to match the population of interest in 1930.

Table A.5: Selection of 1930 households and children into our Numident death-census linked sample, by child 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* | | | | | | |
| Social Security exposure | -0.0003 | -0.0004 | -0.0001 | 0.0001 | 0.0004 | 0.0006 |
| | (0.0003) | (0.0003) | (0.0004) | (0.0005) | (0.0003) | (0.0005) |
| Mean of dep. variable | 0.043 | 0.049 | 0.178 | 0.262 | 0.071 | 0.093 |
| *Panel (b): Sons* | | | | | | |
| Social Security exposure | -0.0000 | -0.0000 | 0.0000 | 0.0002 | 0.0002 | 0.0002 |
| | (0.0002) | (0.0002) | (0.0003) | (0.0004) | (0.0002) | (0.0003) |
| Mean of dep. variable | 0.023 | 0.025 | 0.112 | 0.138 | 0.045 | 0.052 |
| *Panel (c): Daughters* | | | | | | |
| Social Security exposure | -0.0003** | -0.0004*** | -0.0001 | -0.0001 | 0.0003 | 0.0004** |
| | (0.0001) | (0.0001) | (0.0002) | (0.0002) | (0.0002) | (0.0002) |
| Mean of dep. variable | 0.022 | 0.023 | 0.102 | 0.124 | 0.036 | 0.041 |
| *N* | 3,920,008 | 3,920,008 | 3,920,008 | 3,920,008 | 3,920,008 | 3,920,008 |

*Notes:* This table shows the coefficients on Social Security exposure (i.e., negative of earliest age of Social Security eligibility) from equation (2), where the sample consists of all men in our population of interest in the 1930 Census and the outcome is either an indicator for any children/sons/daughters born in a specified cohort range appearing in our Numident death-census linked sample or a discrete variable for the number of such children/sons/daughters appearing in our linked sample. All regressions include fixed effects for men's 3-digit occupation-by-birth cohort, Social Security coverage category, race, and birthplace. Robust standard errors in parentheses, clustered by men's 3-digit industry. *** 1%, ** 5%, * 10% significance.

Table A.6: The impact of father's exposure to Social Security on parents' educational attainment and fathers' wage income and employment status in 1940

| | Dependent variable: Parents' outcomes in 1940 | | | |
|---|---|---|---|---|
| | Years of education | Wage income (log) | In labor force | Employed |
| | (1) | (2) | (3) | (4) |
| *Panel (a): Men in 1930–1940 CLP linked sample (ABE conservative algorithm)* | | | | |
| Social Security exposure | 0.0052 | -0.0001 | -0.0012 | -0.0015 |
| | (0.0085) | (0.0035) | (0.0009) | (0.0010) |
| Mean of dep. variable | 7.6 | 1,037 | 0.89 | 0.81 |
| *N* | 725,039 | 547,692 | 740,257 | 740,257 |
| *Panel (b): Fathers in Numident death-census sample who are additionally linked to 1940* | | | | |
| Social Security exposure | 0.0113** | -0.0018 | 0.0001 | -0.0013 |
| | (0.0046) | (0.0039) | (0.0009) | (0.0010) |
| Mean of dep. variable | 7.4 | 1,039 | 0.90 | 0.81 |
| *N* | 513,978 | 398,824 | 525,524 | 525,524 |
| *Panel (c): Mothers in Numident death-census linked sample who are additionally linked to 1940* | | | | |
| Social Security exposure | 0.0010 | — | — | — |
| | (0.0074) | | | |
| Mean of dep. variable | 7.5 | | | |
| *N* | 520,681 | | | |

*Notes:* This table shows the coefficients on father's Social Security exposure (i.e., negative of earliest age of Social Security eligibility) from equation (2), where the outcome is indicated in the column title. In panel (a), the sample consists of men in our population of interest in the 1930 Census who are linked to the 1940 Census via linkages from the Census Linking Project using the ABE conservative algorithm (Abramitzky et al., 2025). In panels (b) and (c), the sample consists of fathers and mothers in our Numident death-census linked sample who are additionally linked to the 1940 Census via linkages from Bailey et al. (2026). All regressions include fixed effects for fathers' 3-digit occupation-by-birth cohort, Social Security coverage category, race, and birthplace. In column 2, means of wage income in 1939 dollars are shown. Observations weighted using inverse propensity score weights that re-weight the sample to match the population of interest in 1930. Robust standard errors in parentheses, clustered by fathers' 3-digit industry. *** 1%, ** 5%, * 10% significance.

Table A.7: The impact of father's exposure to Social Security on children's post-1950 migration patterns

| | Dependent variable: Children's location at death vs. location in 1950 | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Different location | | | | | Distance between counties | | |
| | County | County (same state) | State | Census division | Census region | $>0$ and $\leqslant 100$ miles | $>100$ and $\leqslant 500$ miles | $>500$ miles |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| *Panel (a): All children* | | | | | | | | |
| Social Security exposure | -0.0011 | 0.0005 | -0.0016 | -0.0018 | -0.0020* | 0.0006 | 0.0000 | -0.0016* |
| | (0.0013) | (0.0007) | (0.0012) | (0.0013) | (0.0011) | (0.0006) | (0.0010) | (0.0009) |
| Mean of dep. variable | 0.51 | 0.23 | 0.28 | 0.21 | 0.18 | 0.21 | 0.13 | 0.17 |
| *N* | 507,232 | 507,232 | 507,936 | 507,936 | 507,936 | 507,232 | 507,232 | 507,232 |
| *Panel (b): Sons* | | | | | | | | |
| Social Security exposure | -0.0022 | 0.0000 | -0.0022* | -0.0026* | -0.0028*** | -0.0006 | -0.0007 | -0.0010 |
| | (0.0016) | (0.0014) | (0.0013) | (0.0014) | (0.0009) | (0.0013) | (0.0009) | (0.0008) |
| Mean of dep. variable | 0.51 | 0.22 | 0.29 | 0.22 | 0.18 | 0.20 | 0.13 | 0.18 |
| *N* | 253,174 | 253,174 | 253,564 | 253,564 | 253,564 | 253,174 | 253,174 | 253,174 |
| *Panel (c): Daughters* | | | | | | | | |
| Social Security exposure | -0.0002 | 0.0013 | -0.0015 | -0.0012 | -0.0017 | 0.0017 | 0.0006 | -0.0026** |
| | (0.0015) | (0.0011) | (0.0013) | (0.0013) | (0.0014) | (0.0012) | (0.0013) | (0.0011) |
| Mean of dep. variable | 0.50 | 0.23 | 0.27 | 0.20 | 0.17 | 0.21 | 0.13 | 0.16 |
| *N* | 254,058 | 254,058 | 254,372 | 254,372 | 254,372 | 254,058 | 254,058 | 254,058 |

*Notes:* This table shows the coefficients on father's Social Security exposure (i.e., negative of earliest age of Social Security eligibility) from equation (3), where the sample consists of children in our Numident death-census linked sample who are additionally linked to the 1950 Census and the outcome is an indicator for migrating to a different county/state/division/region or migrating a specified number of miles away between 1950 and the time of death. Panels show separate estimates for all children, sons and daughters. All regressions include fixed effects for children's sex-by-birth cohort and fathers' 3-digit occupation-by-birth cohort, Social Security coverage category, race, and birthplace. Observations weighted using inverse propensity score weights that re-weight the sample to match the population of interest in 1930. Robust standard errors in parentheses, clustered by fathers' 3-digit industry. *** 1%, ** 5%, * 10% significance.

Table A.8: The impact of father's exposure to Social Security on children's long-run mortality

| | Dependent variable: | |
|---|---|---|
| | Died between ages 65 and 77 | Age at death (conditional on dying between ages 65 and 77) |
| | (1) | (2) |
| *Panel (a): All children (born in 1923–1930)* | | |
| Social Security exposure | -0.001 | 0.035*** |
| | (0.001) | (0.010) |
| Mean of dep. variable | 0.58 | 71.6 |
| *N* | 225,731 | 131,277 |
| *Panel (b): Sons (born in 1923–1930)* | | |
| Social Security exposure | -0.001 | 0.047** |
| | (0.002) | (0.018) |
| Mean of dep. variable | 0.58 | 71.5 |
| *N* | 128,313 | 74,225 |
| *Panel (c): Daughters (born in 1923–1930)* | | |
| Social Security exposure | 0.001 | 0.020 |
| | (0.002) | (0.016) |
| Mean of dep. variable | 0.58 | 71.8 |
| *N* | 97,418 | 57,052 |

*Notes:* This table shows the coefficients on father's Social Security exposure (i.e., negative of earliest age of Social Security eligibility) from equation (3), where the sample consists of children born in 1923–1930 in our Numident death-census linked sample. In column 1, the outcome is an indicator for dying between ages 65 and 77. In column 2, the outcome is age at death, for the subset of children who died between ages 65 and 77. Panels show separate estimates for all children, sons and daughters. All regressions include fixed effects for children's sex-by-birth cohort and fathers' 3-digit occupation-by-birth cohort, Social Security coverage category, race, and birthplace. Observations weighted using inverse propensity score weights that re-weight the sample to match the population of interest in 1930. Robust standard errors in parentheses, clustered by fathers' 3-digit industry. *** 1%, ** 5%, * 10% significance.

Table A.9: The impact of father's exposure to Social Security on children's living arrangements, labor market outcomes, marital status, and fertility in 1950

| | Dependent variable: Outcome in 1950 | | | | | | |
|---|---|---|---|---|---|---|---|
| | Living with parent | Living with father | Living with mother | In labor force | Occscore (log) | Married | Any children |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| *Panel (a): All children* | | | | | | | |
| Social Security exposure | -0.0012** | -0.0010* | -0.0010* | 0.0010 | -0.0005 | -0.0001 | 0.0010 |
| | (0.0005) | (0.0005) | (0.0006) | (0.0008) | (0.0010) | (0.0007) | (0.0009) |
| Mean of dep. variable | 0.19 | 0.12 | 0.16 | 0.63 | 2,706 | 0.80 | 0.65 |
| *N* | 601,321 | 601,321 | 601,321 | 601,321 | 357,190 | 601,321 | 601,321 |
| *Panel (b): Sons* | | | | | | | |
| Social Security exposure | -0.0014 | -0.0012 | -0.0010 | 0.0006 | 0.0001 | 0.0009 | 0.0011 |
| | (0.0009) | (0.0014) | (0.0009) | (0.0008) | (0.0010) | (0.0014) | (0.0017) |
| Mean of dep. variable | 0.18 | 0.12 | 0.16 | 0.94 | 2,855 | 0.80 | 0.63 |
| *N* | 309,682 | 309,682 | 309,682 | 309,682 | 270,215 | 309,682 | 309,682 |
| *Panel (c): Daughters* | | | | | | | |
| Social Security exposure | -0.0009 | -0.0009 | -0.0008 | 0.0008 | 0.0001 | -0.0009 | 0.0003 |
| | (0.0013) | (0.0010) | (0.0012) | (0.0015) | (0.0020) | (0.0013) | (0.0012) |
| Mean of dep. variable | 0.20 | 0.12 | 0.17 | 0.32 | 2,254 | 0.80 | 0.67 |
| *N* | 291,639 | 291,639 | 291,639 | 291,639 | 86,975 | 291,639 | 291,639 |

*Notes:* This table shows the coefficients on father's Social Security exposure (i.e., negative of earliest age of Social Security eligibility) from equation (3), where the sample consists of children in our Numident death-census linked sample who are additionally linked to the 1950 Census and the outcome is indicated in the column title. Panels show separate estimates for all children, sons and daughters. All regressions include fixed effects for children's sex-by-birth cohort and fathers' 3-digit occupation-by-birth cohort, Social Security coverage category, race, and birthplace. In column 2, means of 1950-based IPUMS occupational income scores in 1949 dollars are shown. Observations weighted using inverse propensity score weights that re-weight the sample to match the population of interest in 1930. Robust standard errors in parentheses, clustered by fathers' 3-digit industry. *** 1%, ** 5%, * 10% significance.

Table A.10: Median earnings by state in 1950 relative to national median, men aged 25–54

| | Median earnings by state in 1950 relative to US median, selected education levels | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | All years | 4 years | 6 years | 8 years | 10 years | 12 years | 14 years | 16 years |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| United States | 2,450 | 1,550 | 1,950 | 2,450 | 2,650 | 2,850 | 3,050 | 3,050 |
| Alabama | -800 | -350 | -600 | -700 | -600 | -300 | -400 | -200 |
| Arizona | -300 | 100 | -100 | -350 | -200 | -200 | -450 | -100 |
| Arkansas | -900 | -500 | -700 | -1,000 | -700 | -500 | -500 | -400 |
| California | 300 | 400 | 100 | 100 | 200 | 200 | 0 | 0 |
| Colorado | 0 | 100 | -100 | -200 | -200 | -100 | -300 | -100 |
| Connecticut | 200 | 700 | 400 | 150 | 0 | 0 | 0 | 0 |
| Delaware | -200 | 0 | -300 | -200 | -100 | 0 | -50 | 500 |
| District of Columbia | 0 | 400 | 100 | -100 | -100 | -100 | -100 | 0 |
| Florida | -600 | -200 | -400 | -600 | -500 | -300 | -500 | -500 |
| Georgia | -900 | -300 | -500 | -600 | -600 | -300 | -400 | -100 |
| Idaho | 100 | 400 | 100 | -100 | -50 | 0 | -50 | 0 |
| Illinois | 400 | 800 | 500 | 300 | 400 | 200 | 0 | 0 |
| Indiana | 200 | 500 | 200 | 100 | 200 | 200 | 0 | 50 |
| Iowa | 100 | 400 | 100 | -100 | 0 | -100 | 0 | 0 |
| Kansas | 0 | 325 | 100 | -300 | -200 | -200 | -200 | 0 |
| Kentucky | -700 | -400 | -550 | -600 | -300 | -400 | -500 | -450 |
| Louisiana | -700 | -150 | -400 | -400 | -300 | -200 | -250 | 0 |
| Maine | -400 | 100 | -100 | -450 | -400 | -400 | -600 | -200 |
| Maryland | 0 | 150 | 100 | 0 | 0 | 0 | 0 | 100 |
| Massachusetts | 100 | 500 | 200 | -100 | -100 | -100 | -200 | 0 |
| Michigan | 500 | 1,000 | 700 | 400 | 400 | 300 | 0 | 200 |
| Minnesota | 100 | 400 | 0 | -100 | 200 | 100 | 0 | 100 |
| Mississippi | -1,200 | -600 | -900 | -1,200 | -925 | -800 | -500 | 0 |
| Missouri | -400 | -275 | -200 | -400 | -200 | -300 | -300 | -150 |
| Montana | 150 | 400 | 100 | 100 | 200 | 100 | -200 | 250 |
| Nebraska | -100 | 300 | -50 | -400 | -200 | -300 | -150 | 0 |
| Nevada | 600 | 800 | 100 | 300 | 400 | 500 | 500 | 800 |
| New Hampshire | -100 | 300 | 100 | -300 | -300 | -300 | -300 | 0 |
| New Jersey | 300 | 500 | 500 | 200 | 200 | 200 | 0 | 400 |

*(table continues on next page)*

*Notes:* This table shows deviations in median earned income by state in 1950 relative to the national median among men aged 25–54, separately for selected education levels. Earned income is defined as the sum of wage income and business or farm income. All amounts expressed in 1949 dollars.

Table A.10 (cont.): Median earnings by state in 1950 relative to national median, men aged 25–54

| | Median earnings by state in 1950 relative to US median, selected education levels | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | All years | 4 years | 6 years | 8 years | 10 years | 12 years | 14 years | 16 years |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| New Mexico | -400 | 0 | -400 | -400 | -350 | -100 | 0 | 200 |
| New York | 150 | 500 | 200 | 100 | 100 | 100 | 0 | 0 |
| North Carolina | -700 | -200 | -400 | -600 | -600 | -400 | -500 | -250 |
| North Dakota | -200 | 100 | -100 | -350 | -200 | -200 | -150 | 300 |
| Ohio | 300 | 600 | 500 | 200 | 300 | 200 | 50 | 100 |
| Oklahoma | -400 | -300 | -400 | -400 | -300 | -200 | -200 | 0 |
| Oregon | 350 | 500 | 400 | 150 | 200 | 200 | 0 | 50 |
| Pennsylvania | 100 | 500 | 300 | 0 | -100 | -100 | -100 | 0 |
| Rhode Island | -100 | 500 | 100 | -100 | -200 | -200 | -200 | 300 |
| South Carolina | -900 | -400 | -400 | -600 | -600 | -400 | -500 | -200 |
| South Dakota | -300 | 0 | -400 | -400 | -300 | -300 | -200 | 25 |
| Tennessee | -800 | -400 | -550 | -900 | -600 | -300 | -400 | -300 |
| Texas | -400 | -50 | -200 | -400 | -200 | 0 | 0 | 0 |
| Utah | 500 | 600 | 500 | 300 | 300 | 200 | 0 | 400 |
| Vermont | -300 | -100 | -200 | -500 | -400 | -300 | -400 | 0 |
| Virginia | -400 | -100 | -200 | -400 | -200 | -100 | -100 | 300 |
| Washington | 450 | 500 | 300 | 200 | 300 | 200 | 100 | 200 |
| West Virginia | -200 | 400 | 100 | -200 | -200 | -200 | -300 | 0 |
| Wisconsin | 200 | 500 | 400 | 100 | 200 | 100 | 0 | 0 |
| Wyoming | 500 | 600 | 250 | 200 | 400 | 300 | 200 | 600 |

*Notes:* This table shows deviations in median earned income by state in 1950 relative to the national median among men aged 25–54, separately for selected education levels. Earned income is defined as the sum of wage income and business or farm income. All amounts expressed in 1949 dollars.

Table A.11: The impact of father's exposure to Social Security on sons' educational attainment, veteran status and union status

| | Dependent variable: | | | | | |
|---|---|---|---|---|---|---|
| | Years of education | | | Veteran status | | Union status |
| | 1940 | 1950 | 1973 | 1950 | 1973 | 1973 |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| *Panel (a): Sons born in 1907–1930 in 1973 OCG sample* | | | | | | |
| Social Security exposure | — | — | -0.023 | — | 0.007 | -0.030 |
| | | | (0.153) | | (0.028) | (0.023) |
| Mean of dep. variable | | | 11.3 | | 0.62 | 0.33 |
| *N* | | | 2,714 | | 2,819 | 2,790 |
| *Panel (b): Sons born in 1907–1919 in 1973 OCG sample* | | | | | | |
| Social Security exposure | — | — | -0.008 | — | -0.003 | -0.039 |
| | | | (0.186) | | (0.044) | (0.033) |
| Mean of dep. variable | | | 10.9 | | 0.48 | 0.33 |
| *N* | | | 1,565 | | 1,626 | 1,604 |
| *Panel (c): Sons in Numident death-census linked sample who are additionally linked to 1940 or 1950* | | | | | | |
| Social Security exposure | -0.018 | -0.039** | — | 0.000 | — | — |
| | (0.017) | (0.020) | | (0.004) | | |
| Mean of dep. variable | 11.0 | 10.1 | | 0.61 | | |
| *N* | 83,507 | 50,739 | | 51,149 | | |

*Notes:* This table shows the coefficients on father's Social Security exposure (i.e., negative of earliest age of Social Security eligibility) from equation (3), where the outcome is indicated in the column title. In panels (a) and (b), the samples are sons born in 1907–1930 or 1907–1919 in our 1973 OCG sample. In panel (c), the sample consists of sons in our Numident death-census linked sample who are additionally linked to the 1950 Census (columns 2 and 4) or to the 1940 Census via linkages from Bailey et al. (2026) (column 1). In column 1, the sample is restricted to sons aged 25 or older and not attending school in 1940. In column 2, the sample is restricted to sample-line sons aged 25 or older and not attending school in 1950. All regressions include fixed effects for sons' birth cohort and fathers' 3-digit occupation-by-birth cohort, Social Security coverage category, race, and birthplace. Observations weighted using inverse propensity score weights that re-weight the sample to match the population of interest in 1930. Robust standard errors in parentheses, clustered by fathers' 3-digit industry. *** 1%, ** 5%, * 10% significance.

Table A.12: Linking families in SS-5 records to households in the 1930 Census: Model features

| Stage 1 model features (at the set level) |
|---|
| Description |
| Gap between highest and second-highest JW score |
| Max JW score |
| Mean JW score |
| Absolute difference in number of matching children (method 1) between highest and second-highest JW score potential |
| Gap between highest and second-highest father last JW score |
| Max father last JW score |
| Mean father last JW score |
| Absolute difference in number of matching children (method 1) between highest and second-highest father last JW score potential |
| Unique potential with 1+ matching children (method 1) |
| Unique potential with 2+ matching children (method 1) |
| Unique potential with max JW score or max father last JW score & 1+ matching children (method 1) |
| Unique potential with max JW score or max father last JW score & 2+ matching children (method 1) |
| Unique potential with max JW score or max father last JW score & father or mother middle initial match & 1+ matching children (method 1) |
| Unique potential with max JW score or max father last JW score & father or mother middle initial match & 2+ matching children (method 1) |
| Unique potential with 1+ matching children (method 1) & 1 matching child (method 2) |
| Unique potential with 1+ matching children (method 1) & 2+ matching children (method 2) |
| Unique potential with max JW score or max father last JW score & 1+ matching children (method 1) & 1 matching child (method 2) |
| Unique potential with max JW score or max father last JW score & 1+ matching children (method 1) & 2+ matching children (method 2) |
| Unique potential with max JW score or max father last JW score & father or mother middle initial match & 1+ matching children (method 1) & 1 matching child (method 2) |
| Unique potential with max JW score or max father last JW score & father or mother middle initial match & 1+ matching children (method 1) & 2+ matching children (method 2) |
| Unique potential with max JW score or max father last JW score & 1+ matching children (methods 1 or 2) & father or mother middle initial match & 1+ matching children (method 3) |
| Unique potential with max JW score or max father last JW score & 1+ matching children (methods 1 or 2) & father or mother middle initial match & 1+ matching children (method 3) |
| Unique potential with max JW score or max father last JW score & 1+ matching children (methods 1 or 2) & father or mother middle initial match & 1+ matching children (method 4) |
| Unique potential with max JW score or max father last JW score & 1+ matching children (methods 1 or 2) & father or mother middle initial match & 1+ matching children (method 4) |

| Stage 2 model features (at the pair level) | |
|---|---|
| Description | Description |
| JW score | Modal race match |
| Father last JW score | Father middle initial match |
| Number of matching children (method 1) | Mother middle initial match |
| Number of matching children (method 2) | Max JW score or max father last JW score & 1+ matching children (method 1) |
| Number of matching children (method 3) | Max JW score or max father last JW score & 1+ matching children (method 2) |
| Number of matching children (method 4) | Max JW score or max father last JW score & 1+ matching children (method 3) |
| Rank | Max JW score or max father last JW score & 1+ matching children (method 4) |
| Rank = 1 | |

*Notes:* "JW score" is the average of three Jaro-Winkler scores comparing the following strings: (1) father first names, (2) father last names, and (3) mother first names. Two children match according to method 1 if three conditions hold: (1) the Jaro-Winkler score between their first names is above 0.8 or 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.13: The relationship between late-life ZIP quality and late-career earnings

| | Dependent variable: Late-life ZIP quality | |
|---|---|---|
| | Mean AGI in 2001 (log) | Median home values in 2000 (log) |
| | (1) | (2) |
| Late-career earnings (log) | 0.075*** | 0.076*** |
| | (0.003) | (0.005) |
| $R^2$ | 0.059 | 0.033 |
| $N$ | 8,100 | 7,900 |

*Notes:* Each entry in this table corresponds to the estimated slope coefficient from a linear regression of a measure of late-life ZIP quality (log of ZIP-level mean AGI in 2001 or ZIP-level median home values in 2000) on a constant and the log of late-career earnings between ages 50 and 65, where the sample consists of male HRS respondents aged 65 or older when they were last interviewed and for which we have ZIP of residence and SSA earnings information (see Appendix A.5 for details). Robust standard errors in parentheses. Sample sizes have been rounded to the nearest hundredth following HRS disclosure guidelines. *** 1%, ** 5%, * 10% significance.