Living the American Dream: How Norway Became a High-Mobility Country*

Eirik Berger Abel
Norwegian School of Economics

Job Market Paper
(Click here for the most recent version)

November 21, 2023

Abstract

I estimate long-run trends in intergenerational mobility in income in Norway during a period that includes World War II and the creation of the welfare state. To overcome the challenge of data availability, I digitize and link a novel dataset with more than 16 million observations from tax records between 1924 and 1964 to data on family ties from historical censuses and modern administrative data. I show that persistence between fathers and sons in the income percentile rank was high in the early 20th century but decreased substantially for cohorts born between the 1920s and 1940s. The convergence of rural and urban areas explains about half of the total fall in persistence. First, I link this result to changes in education by using plausibly exogenous variation in the intensity of schooling from a primary school reform, which reduced the gap between cities and rural areas, and find that it significantly decreased persistence. Second, I show that the return to education fell considerably over time, particularly at the beginning of World War II. Comparing persistence for a set of father-son pairs but using income for the father measured just before and after this shock, I find about 13 percent lower persistence in income percentile ranks. These results suggest that equal access to education and low income inequality are two key drivers behind Norway's transition to high mobility.

JEL Classification: N34, N44, I14, I24
Keywords: Education, Inequality, Mobility

*I am thankful for advice and comments from Ran Abramitzky, Kjell Gunnar Salvanes, Ola Grytten, René Karadakic, Andreas Haller, Samuel David Hirshman, Katrine Løken, Manudeep Bhuller, Mikko Silliman and participants at seminars at NHH. The activities at NHH are funded by the Research Council of Norway through project No 275800 and through its Centres of Excellence Scheme, FAIR project No 262675.

1 Introduction

Inequality has risen throughout Western countries since at least the 1980s, leading to concerns that children growing up today do not have the same opportunity for upward mobility as their parents (Corak, 2013). In cross-country comparisons, the Scandinavian welfare states outperform most other countries on measures of relative intergenerational mobility in income - meaning that relative incomes in Norway, Denmark and Sweden are less tied to that of their parents than in almost any other country. It is often suggested that this results from a compressed distribution of incomes, transfers to the poor and welfare states providing equal access to education, health and social insurance (Berger et al., 2023; Pekkarinen, Salvanes and Sarvimäki, 2017). But despite the strong interest in this topic from policymakers and researchers, we have little evidence on how the high rates of mobility in Scandinavia came about.

Existing evidence shows that trends in intergenerational income mobility in the Scandinavian countries have been flat for male cohorts born between 1951 and 1979 (Ahrsjö, Karadakic and Rasmussen, 2023). If the welfare state increased mobility, it must have happened for earlier cohorts. This is indicated by Pekkarinen, Salvanes and Sarvimäki (2017), who push the Norwegian administrative data to its limit by estimating intergenerational mobility for cohorts back to the 1930s with incomes from the late 1960s and 1970s. But if we expect that income inequality directly impacts intergenerational mobility, we would want to measure incomes before the massive compression in incomes that happened in Norway and many other countries in the 1940s and 1950s (Abel, Abramitzky and Salvanes, 2023). However, the type of individual-level data with multiple observations on income linked with family ties needed to produce mobility measures in this period does not exist (Mogstad and Torsvik, 2023; Black and Devereux, 2011).[1]

In this paper, I provide the first estimates of intergenerational mobility in income in Norway from the early 1900s – when inequality in income was high, transfers to the poor were limited, and the welfare state was still only in its early infancy – and show how it has evolved until today. After documenting that intergenerational mobility in income used to be very low, I use a series of shocks to investigate the causal drivers of income mobility and, more specifically, how Norway became a high-mobility society. I do this by extending the Norwegian tax registers almost 50 years back to the early 1920s by constructing a novel individual-level dataset of more than 16 million observations from Norwegian tax records from 1924-1964 (Abel and Salvanes, 2023). I link individuals in the historical tax data to modern administrative data and historical population censuses to get links between family members. This allows me to estimate intergenerational mobility in income for cohorts born between 1910 and 1980 and to average multiple observations on income for each person to estimate lifetime income (Black and Devereux, 2011; Mogstad and Torsvik, 2023).

[1]This has spurred an extensive literature in economic history using measures from sociology and other alternative measures of intergenerational mobility by using data on occupations from linked census data (Björklund, Jäntti et al., 1999; Long and Ferrie, 2013) and most recently survey data (Jácome, Kuziemko and Naidu, 2021).

I first show that persistence in income rank across generations for cohorts born in the 1910s and 1920s was more than twice as high as today's – with a rank-rank persistence of about 0.55 compared to 0.20 today. The decrease in persistence is isolated to cohorts born in the early 1920s to 1940s and the level has remained relatively constant ever since. This shows that Norway has not always been the mobile society it is today and that this mobility must have been caused by something happening primarily to cohorts from the first part of the 1900s. To better understand these country-wide patterns, I follow Jácome, Kuziemko and Naidu (2021) and decompose the rank-rank mobility measure into components within and between rural and urban areas. Cohorts born in cities before 1935 earn better than those growing up in rural areas, even if their fathers have the same income. Using data from the entire sample, I find that the gap between urban and rural areas was dramatically reduced during and after World War II and virtually disappeared by the late 1980s. Given a set of assumptions, this reduction has contributed to a 16 percentage point reduction in rank-rank persistence – or almost half the overall decline – found in this period.

In the second part of the paper, I investigate the possible drivers of this convergence. Changes in mobility can result from pre-market, market or post-market factors. I focus on the first two by looking at education (pre-market) and the returns to education (labor market), and close down the direct effect of post-market factors by looking at income measures pre-tax and pre-transfers. First, I show that the gap in educational attainment between rural and urban areas was large and relatively stable until cohorts born in the late 1920s when the share with more than primary schooling started to converge between rural and urban areas. Gaps in the attainment of primary, secondary and higher education were closed sequentially, and the gap in completed years of education was closed by cohorts born in the mid-1960s. Education is a key determinant of income, so we might expect a reduction in inequalities in educational attainments across areas to decrease persistence in income across generations. To provide evidence of this mechanism, I use variation from the implementation of the 1936 rural primary school reform in Norway, which differentially increased weeks of schooling during primary school for different municipalities and cohorts, as a basis for causal inference. I find that the reform decreased persistence in income rank by more than five percentage points, suggesting that it had the strongest effect on kids from poorer households.

Second, I investigate educational premiums over time, which would impact mobility by changing the pay for already educated individuals or the incentives for the young to pursue additional education. Educational premiums estimated using a Mincer equation were around 15 percent in the 1930s before dropping sharply at the start of World War II to about 8 percent. A second reduction in educational premiums followed in the 1970s, before trending upwards, as in many Western countries (Goldin and Katz, 2009; Autor, 2019). Premiums were low in the 1920s, but this could be the result of the severe economic crisis of the 1920s in Norway rather than the 'normal' levels. Taken together, these results point to the occupation of Norway during World War II as the most important driver of educational premiums in this period. To understand the direct impact of this change in the income structure, I compare persistence across generations for a set of father-son pairs, but using income for the father measured just before and after the occupation of Norway during World War II. I find about four percentage points – or 13 percent – lower persistence in income percentile ranks after the shock.

In this paper, I make three main contributions to our understanding of intergenerational mobility mobility. First, I contribute to a large and growing literature estimating historical rates of intergenerational mobility (Ferrie, 2005; Long and Ferrie, 2013; Modalsli, 2017; Berger et al., 2023; Ward, 2021; Jácome, Kuziemko and Naidu, 2021). I separate myself from this literature using frequently measured individual income data from historical tax registers and constructing consistent measures of persistence using data spanning 100 years. Existing evidence on historical rates of intergenerational mobility in the Scandinavian countries follows Long and Ferrie (2013) in linking together historical decennial population censuses and looking at changes in occupations between fathers and sons. To get closer to a measure of income mobility, they often impute incomes based on occupation and other variables. A study from Norway using census data on occupations finds a gradual increase in mobility until today. However, it lacks available census years for large parts of the 20th century (1910-1960) (Modalsli, 2017). My approach of directly utilizing individual-level income data for estimating mobility offers a more accurate representation of economic opportunity than methods reliant on imputed income from occupations, which overlook the income disparities within the same occupational categories and typically are restricted to a low number of occupational categories. Such imputations also typically cannot pick up year-to-year changes in income levels, which is important for studying the interaction between income inequality and intergenerational mobility. Finally, these papers typically also use just one observation for each person, which will tend to bias persistence toward zero because of noise (Ward, 2021).

None of these papers estimate intergenerational mobility with individual incomes for parents and children. The exception is Feigenbaum (2018), which links the 1915 Iowa census with the 1940 US census to estimate intergenerational mobility in income for a sample in the US. In a new paper, Jácome, Kuziemko and Naidu (2021) uses historical survey data with information on family income and fathers' occupation to estimate relative intergenerational mobility in income for cohorts between 1910 and 1970 in the US. They find a U- or L-shaped pattern in intergenerational persistence, where persistence decreased substantially for cohorts between the 1910s and 1940s. This contrasts previous findings for the US using occupations, which finds a gradual decrease in mobility since the 19th century. The paper more closely matches new findings from Ward (2021), which looks at intergenerational mobility using imputed occupational incomes for multiple census years for each person.

Second, I contribute to the empirical literature on schooling and intergenerational mobility by providing causal evidence of the importance of education in explaining Norway's rise to a high-mobility country. In Scandinavia, researchers have argued about the equalizing impact of early education (Karlson and Landersø, 2021; Pekkarinen, Salvanes and Sarvimäki, 2017; Pekkarinen, Uusitalo and Kerr, 2009). However, these papers cannot speak strongly to the causal effect of the expansion of education in the early part of the 1900s, when we know that much of the improvements in mobility happened. In the context of the early 20th century US, it has been shown that the Rosenwald school construction program had the highest gains in the most disadvantaged counties, suggesting that it impacted intergenerational mobility (Aaronson and Mazumder, 2011). Similarly, Card, Domnisoru and Taylor (2022) use data from the 1940 census and show that school quality increased upward mobility in relative levels of education. Still, they do not link their results to broader changes in the trends in income mobility rates over time.

Finally, I contribute to an extensive literature on the changes and determinants of historical rates of educational premiums. Studies from the US find a drop in educational premiums in the early 20th century and attribute this to the race between technology and human capital, where significant increases in the educated workforce pressed down high-skilled wages (Goldin and Margo, 1992; Katz and Murphy, 1992; Goldin and Katz, 2009). I estimate educational premiums for Norway and show that the pattern matches qualitatively with the US. However, using yearly data, I find that the fall in premiums is almost entirely isolated to the first years of World War II, making it highly unlikely to be driven by increases in the educated workforce. Instead, it suggests that World War II played an important role in equalizing incomes and reducing educational premiums, as argued for the Norwegian context by Abel, Abramitzky and Salvanes (2023).

This paper is structured as follows. First, I provide an overview of Norway during the 20th century and a review of changes in inequalities. Second, I describe how I construct my full linked dataset. Third, I present evidence on levels and trends in intergenerational mobility in income, focusing on the rural-urban gap. Fourth, I present causal evidence on the effect of education and educational premiums on intergenerational mobility. Finally, I discuss my results and conclude.

2 Historical Background

This paper investigates intergenerational mobility in income for cohorts born during the century when Norway transformed itself from a rural country dominated by farming and fishery to one of the richest in the world (Grytten, 2020). To better understand the history and institutions affecting rates of intergenerational mobility, I will first provide an overview of Norway's economic transformation in this period before detailing how the welfare state was developed to create equal chances for success in life for all its inhabitants. Finally, I show how inequality in income, wealth, educational attainment and health have developed in this period. I pay particular attention to forces impacting cohorts born in the 1940s and earlier, as this is the period where my findings indicate that mobility is increasing.

2.1 Norway During the 1900s

In the 1880s, Norway was a predominantly rural society, with most people working in farming, fishery, and forestry. While it has been portrayed as poor and underdeveloped, Norway was likely quite average in terms of GDP per capita among the developed countries in Western Europe (Myhre, 2022; Grytten, 2020). It also had among the lowest infant mortality rates in the world, the highest life expectancy in northern Europe and ranked among the highest in reading and writing skills (Regidor et al., 2011). It seems fair to say that Norway was a reasonably developed country long before the development of the welfare state.

Compared to other countries in Western Europa, industrialization in Norway happened relatively late, with 11.9 percent of the working population employed in the industry at the turn of the century mainly concentrated in larger cities (Leknes and Modalsli, 2018; Venneslan, 2009). It wasn't until 1905, when Norway got its independence from Sweden and "Norsk Hydro" started up its electricity-intensive industry, that industrialisation started to take off. In the following 30 years, more than 140 hydroelectric power plants were constructed, mostly in rural areas, providing electricity to local energy-intensive industries. This development brought a great structural transformation of the Norwegian economy, increased resources and improved long-term health (Leknes and Modalsli, 2018; Karadakic, 2023). Norway reached peak industrialization after World War II, with 35 percent of its workforce working in the industry sector and significant deindustrialization beginning in the 1970s (Grytten, 2020).

Employment in agriculture, fishery and forestry had remained relatively constant until World War II, when employment in agriculture, in particular, started to decline. These sectors had struggled with overproduction and reductions in real incomes in the inter-war period and underemployment was a massive problem. The German occupation in 1940 led to a short-term stillstand in the economy, but only months after the capitulation the economy was booming and unemployment quickly disappeared (Abel, Abramitzky and Salvanes, 2023; Ingulstad, Hatlehol and Frøland, 2017). Disruptions to trade, the presence of foreign troops and workers and massive investments in industry and infrastructure led to massive increases in demand for food and raw materials, resulting in increases in the wages in the previously struggling primary sectors. Technological change reached the agriculture sector, which adopted mechanical milking machines after World War II, inducing young women to find alternative work in the cities (Ager, Goñi and Salvanes, 2023).

The years following World War II saw unprecedented growth in GDP and the years 1950-1973 have been known as "the golden years" in Norwegian history for low levels of unemployment and inflation (Grytten, 2008). Oil was found in the North Sea in 1969 and production began in the 1970s. This supported a growing public sector and new and large transfer programs initiated by the Norwegian Labor Party. The stable economic situation stands in contrast to that of the inter-war years – with high peaks and low troughs. The two highest peaks came in 1916 and 1930 with GDP about eight percent above trend - with the 1930 peak being the highest economic upturn of the 20th century (Eika, 2008). The largest troughs came in 1919 and 1921, with GDP about 11 and 8 percent below trend, respectively.

2.2 The Development of the Welfare State

Historically, the responsibility for social problems was predominantly borne by families, churches, and individual parishes. However, in the 18th and 19th centuries, the Norwegian authorities began implementing public measures to combat poverty and social distress. Throughout the 19th century, public benefits remained minimal, and their use was associated with considerable social stigma. The early seeds of the welfare state came in the first part of the 20th century, with some municipalities taking on the role of pioneers and introducing welfare programs themselves, such as unemployment and retirement benefits.

While the ideas for the Norwegian welfare state came decades before World War II, the real breakthrough happened after World War II. The Norwegian Labor Party had been in power since 1935, but was in exile in London during the war. It received a strong mandate for change during the election 1945, and remained in power more or less until 1965. It was during this period that historian Jens Arup Seip coined "the one-party state", and the labor party was able to pass reforms such as universal unemployment insurance and public pensions (Sejersted, 2021; Seip, 1994). The Norwegian Labor Party was focused on decreasing inequalities further and increasing the welfare of the middle and lower classes.

Tax policies were another arena for reducing inequalities. This area remains heavily understudied in Norway, but it is clear that taxes increased from the already high levels put in place during the war, and they were made more progressive. To exemplify, in 1949, the wealth tax was increased fivefold to balance the budget Lie (1995). This was all a part of their ambition to replace private savings with public savings and to distribute financing to firms as loans through public institutions and direct investments (Sejersted, 2021). The Norwegian historian Francis Sejersted later stated, "[...] true to Social Democracy's hegemonic nature, no one really stood up for the rich, the real capitalists". While the measures that the Labor Party passed were drastic by today's standard, they represented a moderation of the previous anti-marked stands of the Party.

2.3 Inequality in Income, Wealth, Education and Health

Inequalities of economic and non-economic factors might directly impact the next generation and thus strengthen intergenerational persistence (Corak, 2013). During the 1900s, Norway implemented a range of policies aimed at decreasing inequality through increasing educational attainment (Abel, Buetikofer and Salvanes, 2023; Acemoglu et al., 2021; Black, Devereux and Salvanes,

2005) and improving health among those from low SES families (Bütikofer, Mølland and Salvanes, 2018; Bütikofer, Løken and Salvanes, 2019; Bütikofer and Salvanes, 2020). I will, therefore, briefly review changes in inequality occurring in Norway during the 1900s, which could have a direct impact on intergenerational mobility in income and social mobility more broadly.

While it has been suggested that Norway had low levels of inequality for a long time, research on inequality in income and wealth suggests that it was significantly higher at the turn of the 20th century than today – with large drops in inequality occurring during World War II and the post-war period. I plot estimates of the pre-tax Gini coefficient for taxpayers from 1925 to today in Figure A.1. The income Gini coefficient seems to have been high – in the interval 50-60 percent – before the second world war, before dropping by 20 percentage points and stabilizing after 1955 (Abel and Vagle, 2017; Aaberge, Atkinson and Modalsli, 2020; Abel and Vagle, 2017). Top income shares closely follow this development (Aaberge and Atkinson, 2010; Abel and Vagle, 2017). Wealth follows a similar trend, but this is less studied and surrounded by more uncertainty (Aaberge, Modalsli and Solbakken, 2018).

Inequalities in health are difficult to measure consistently across time, and to my knowledge, no research has been able to look at this in Norway in the very long run. Bütikofer, Karadakic and Salvanes (2021) study inequalities in mortality over the income distribution and find that income gradient in infant mortality across municipalities was flat by the late 1960. The gradient for older ages across municipalities and the individual-level income gradient in infant mortality lasted into the 21st century. To look at the very long run, I plot the Gini coefficient for the age of death for each death cohort between 1900 and 2014 in Figure A.2. It starts noisily for cohorts who die at the beginning of the century but drops sharply following the end of World War I and the Spanish flu. There was a temporary increase during World War II. Inequalities continued to fall after World War II, but improvements were slower, and the trend was virtually flat after the 1990s.

Finally, I look at inequalities in educational attainment in Figure A.3. Inequalities in education is a broader topic I will get back to in Section 5. Here, I plot the Gini coefficient of years of education across the 20th century and the fraction with less than ten years of education. The Gini coefficient follows an inverse U-shape that increases until cohorts are born around the 1930s and decreases afterwards. This might not seem intuitive, but the low inequality in years of education early in this period comes from the fact that the majority had the same number of years of education (7). As this fraction decreases, the difference between those with less than ten years of schooling and those with more becomes increasingly important. The shaded grey area 1924-1931 are cohorts affected by the 1936 rural primary school reform, and the shaded area 1947-1958 are cohorts affected by the roll-out of ten years of compulsory schooling.

3 Data and Measurement

This paper combines newly digitized data from Norwegian tax registers from 1925 to 1965 and modern tax registers from 1967 to 2018 with information on parents to estimate intergenerational mobility for birth cohorts between 1910 and 1980. The resulting dataset contains multiple observations on taxpayers' income, wealth and occupation from the yearly municipal and state tax assessments (Abel and Salvanes, 2023). I use all available full-count historical censuses and modern population registers to identify parents whenever possible and link them to data from tax registers using fuzzy string matching (Abramitzky et al., 2021). To determine the transmission of a broader set of characteristics, I also include data on educational attainment from the population censuses and the Norwegian Educational Register.

3.1 Individual-level Historical Income Data

I use newly digitized individual-level data with information from Norwegian tax authorities, including information on name, occupation, place of residence, income, and wealth (Abel and Salvanes, 2023). These records have been open to the public since the modern income and wealth tax was created at the turn of the 1880s and include data on all who paid taxes (Gerdrup, 1998). The dataset is constructed by digitizing a previously overlooked series of county-level Norwegian address books published since 1900, resulting in a dataset with more than 16 million individual observations. The address books were published only infrequently for the first decades but started being published regularly for a more extensive set of counties every other year in 1924.

Whenever yearly estimates are presented in this article, missing data for a county is replaced by the inflation-adjusted data from the previous years. The exception is from 1936 to 1947 when I use aggregated municipality-level income data to interpolate incomes. Precise adjustments for these years are particularly important because there were large changes in incomes between municipalities, and the absence of such adjustments would make estimates choppy.[2] Furthermore, it is critical to be able to create consistent samples over time. For this reason, I operate with two samples: Counties that I can follow from 1925 and those I can follow from 1935. Whenever data from modern tax registers are used in the same calculation or figures, the sample is selected to reflect this. I also drop individuals without positive recorded incomes and who do not pay taxes, as this reflects the sample of taxpayers that would be included in the historical tax register.

The definition of income and wealth in the dataset follows that of the Norwegian tax system, which lists income and wealth before most deductions and taxes (Gerdrup, 1998; Abel and Vagle, 2017). Income includes wages, financial income and net income from self-employment. The income measure includes work-related cash transfers, such as unemployment benefits and short-term sickness benefits at least since 1967 (Bhuller, Mogstad and Salvanes, 2017). Although the tax

[2] See Abel, Abramitzky and Salvanes (2023) for details on this interpolation.

system changed in certain respects during this period, the basic income definition used is the same throughout the whole period and is the same used in the modern tax register from 1967. Income and wealth were taxed at the individual level, but wives and children below the age of 15 were taxed jointly with the husband.

The series of address books typically lists the name, occupation, place of residence, income and wealth of all taxpayers as seen in Figure A.4. Such pages were digitised with a primarily automated routine, using computer vision (CV), optical character recognition (OCR) and machine learning (ML) learning methods from the data science literature Abel and Salvanes (2023). The pipeline had three pain parts: First, pre-processing the images and identifying columns with relevant text and second, reading the columns using OCR. Third, the data is structured using a combination of *regular expression*, named entity recognition and text classification. A fourth step performs a long series of post-processing steps to remove noise and add more data. See Abel and Salvanes (2023) for the full documentation of this dataset, the digitization process and quality checks.

In addition to these sources, for this paper, I use a version of the algorithm developed in Abel and Salvanes (2023) to digitize all address books for the cities of Oslo and Trondheim, as these are missing from the dataset created by Abel and Salvanes (2023). These books do not include information from the city tax offices but are based on the municipal population registers that have occupational titles. I can impute incomes based on gender, occupation and age using the data from Abel and Salvanes (2023). The data includes all heads of households in these cities, and the books were published yearly in this period. I also digitize books on top taxpayers from Oslo and the soundings starting in 1936. This allows me to take a detailed look at intergenerational mobility for some of the wealthiest people in Norway.[3]. See Appendix B for more information and descriptive statistics on these data sources.

3.2 Modern Administrative Data and Family Ties

The link with modern administrative data allows me to include data on younger generations, income after 1967, and educational attainment. Notably, the income data in the modern registers comes from the same source as the historical data I use, ensuring continuity in how income is measured. Educational attainment is obtained from the 1960, 1970 and 1980 population censuses and the educational register maintained by Statistics Norway. Unfortunately, we don't have any nationwide dataset on educational attainment further back in time. To verify my estimates, I collect data from several surveys conducted in the 1950s and later that contain data on the respondent's educational level and that of their father and replicate measures of intergenerational mobility in years of education from Mjølsnes (2019).

[3]In 1930, Oslo and the surrounding area was the region with the highest number of affluent taxpayers, followed by Bergen and Trondheim.

To construct a dataset of child-parent links across 100 years, I combine data from all available full-count population censuses with family ties from the modern population register. Statistics Norway maintains the modern population register, which includes the personal identifier of the parents of most individuals born after the 1960s. For older cohorts, I use the 1900, 1910 and 1920 population censuses to extract family ties and use these for all cohorts born before 1920. I can provide reliable family ties for cohorts born between 1945 and 1960 by using the 1960 census. Cohorts born between 1920 and 1945 are challenging, but the Norwegian population data contain links for these years as well, although the fraction with data on parents decreases going back in time.[4]

3.3 Fuzzy String Matching and the Intergenerational Sample

This paper links the historical tax register to Norwegian register data and historical censuses using fuzzy string matching. Linking historical sources in economics typically follows a variant of the Abramitzky, Boustan and Eriksson matching algorithm and uses information on name, birth year and place of birth (Abramitzky et al., 2021). Linking the historical tax register data to this dataset is more problematic because it doesn't include any information on the place or year of birth. In the following, I detail how I perform these linings and give descriptive statistics on matching rates and the balance of the linked sample.

I require the first name and surname's first letter to be perfect matches across the two matched sources. This is justified by the observation that the first letter is unlikely to be digitized incorrectly as it is capitalized. It also massively reduced the number of possible matches we need to investigate. I do not have information on the date of birth for both sources, but I restrict possible matches in the administrative and census data to those between the ages of 15 and 75 based on the year of the tax data I matching with. Depending on the version of the linking approach, I might require the matches to have the same municipality or county of residence as well. To compare observations, I estimate the Jaro Winkler string distance for all potential matches – a score equal to 0 if there is no overlap and 1 if they are identical. We typically require matches to be better than 0.8 or 0.9 to characterize it as a match. Contrary to the census linking literature, I do not require matches to have a string distance with some arbitrary distance to the next best match. The reason is that the data are digitized from printed records and I thus expect a very low level of transcription errors.

Different levels of 'blocking' – municipality, county or no additional blocking – have different upsides and downsides. A detailed blocking level ensures high-quality matches for those with constant blocking characteristics but ignores movers. On the other hand, matching rates without any additional blocking typically result in significantly lower matching rates. In a less conventional

[4] To increase the number of matches in this period, I will get access to data from the historical birth registers currently being digitized through the Historical Population Register project (Norsk Regnesentral, 2023). This will allow me to have access to the 1930 and 1950 censuses and have close to full coverage of family links for the 20th century.

method, I use an iterative approach to get the best of both approaches (Abel and Salvanes, 2023). It entails starting with the strictest level of blocking, but observations that are not matched will be attempted to be matched using the next strictest method, and so on. The final sample will be weakly larger than all the other matching specifications and typically be of higher quality.[5]

Matching rates with the administrative data are shown in Figure A.5. It fluctuates between 28 percent in 1924 and 33 percent in 1932 (excluding the early 1920s, for which I have very few counties) and gradually increases from the late 1930s to the 1960s. I present descriptive statistics for the various stages of data processing in Table A.1. The initial sample contains about 24 million observations, of which about 8 million are observations created from interpolation. Linking this dataset with the Norwegian administrative data leaves 7.3 million observations or 5.1 million when a perfect match is required. The size of the dataset is reduced significantly to about 15.000 observations once linking together sons with fathers and averaging incomes between the ages of 30 and 35 for sons and 55 to 60 for fathers. However, note that each family contains observations from, on average, three observations for the father and three for the son, as can be seen in Figure A.6.[6]. The linked sample has a somewhat higher income but lower wealth than the unlinked sample. The family sample is also significantly less urban, with a share of only 15 percent, compared to 23 percent in the full sample. The share of males in the full sample is about 83 percent, but all females are filtered out in the family sample as it is currently too small to construct meaningful estimates.

4 Relative Intergenerational Mobility in Income

This section presents my main descriptive results. I start by estimating conventional rank-rank mobility estimates for incomes, combining modern administrative data with novel historical data described in Section 3. The data sources and linking practises introduce sample selections, and I implement a series of different weighting schemes to account for this. Then, I perform a series of sample splits to understand better where income persistence across generations comes from and focus mainly on the rural-urban divide. Finally, I decompose mobility rates by rural and urban areas and estimate the contribution of the rural-urban divide to the rank-rank persistence in income across generations.

[5]The current linking setup does not use information on municipality or county of residence primarily because I lack high-quality data on people's places of residence in different years. I am waiting for this data and expect it to increase the quantity and quality of my matches significantly (Norsk Regnesentral, 2023).

[6]These numbers are impacted by the lack of data with family ties for cohorts born between 1920 and 1940. I expect the sample to increase significantly once I get this data through the Norwegian Historical Population Register (Norsk Regnesentral, 2023).

4.1 Levels and Trends in Rank-Rank Persistence in Income

I estimate intergenerational persistence in income using income ranks calculated within income year and five-year birth cohort bins. All estimates are currently restricted to fathers and sons due to data limitations. Income ranks are average across all available observations between the ages of 30 and 35 for sons and 55 to 60 for fathers. Each estimate is from a separate regression estimated in five-year bins for the historical data and yearly for the modern data. The estimating regression is as follows:

$$\text{Child Rank}_i = \alpha + \beta \times \text{Parent Rank}_i + \epsilon_i \tag{1}$$

where Child Rank_i and Parent Rank_i are the income rank for for child i and the parent of child i, respectively. β is the parameter of interest and should be interpreted as the increase in the expected income rank of the child by moving up the income rank by one percentage point. α is the expected income rank for someone growing up with a date and the lowest income rank. An alternative setup would be to estimate the relationship between log incomes. However, the rank-rank estimate has the property that it is independent of changes in inequality across generations. Importantly, these estimates should not be interpreted causally but as correlations, as I cannot control for omitted variables. Estimates of intergenerational rank-rank persistence in income for sons in cohorts born between 1910 and 1980 are shown in Figure A.7. Bars show 95 percent confidence intervals. Rank-rank persistence using historical data starts at around 0.4 for birth cohorts from 1910 to 1920 and then declines until the last cohorts in the historical dataset born in the early 1930s. Estimates of rank-rank persistence for cohorts born in the late 1930s are done using modern administrative data. It continues the falling trend from previous cohorts but is relatively stable for cohorts following the 1950s.

Estimates from cohorts born in the 1930s are relatively similar when using historical and modern datasets. However, while the matching errors are 0 percent in the modern sample, they are not in the historical sample. Assuming that the matching errors are random, this will tend to bias rates of persistence toward zero; however, how much is challenging to say without a clear idea of the rates of false matches in the historical data. Any correction for this would make the fall in persistence – and rise in mobility – for cohorts born between the 1920s and 1950s even more dramatic. To avoid much bias in my results, I require the match to be perfect (the name strings have to be identical) for the estimate in Figure A.7, but I still expect some false matches. A survey of papers using fuzzy matching techniques suggests that the rate of false matches can be from about 16 percent to more than half of the sample when matching across population censuses (Bailey et al., 2020). Assuming a rate of false matches in this interval, I should adjust the estimated intergenerational persistence rates by between 19 and 100 percent.[7]

[7] Adjustment factor = 1/(1-error rate).

The linked sample might not correctly reflect the composition of the full population, either because of biased matching or missing data from particular regions or cities. To account for this, I perform a series of re-weighting exercises in Table A.2 for both by five-year bins of sons' year of birth and the full historical sample. Weights for occupation, municipality and county are based on the composition of the full un-linked historical sample. In contrast, the urban status weights are based on aggregate statistics from Statistics Norway from 1938 because the historical sample does not cover all areas (SSB, 1940). The results for the five-year bins are noisy, but for the full sample, I find that weighting by municipality, county and rural status increases persistence rates across generations. Weighting by occupation leads to a slight decrease in measured persistence rates. However, this might be caused by the fact that not all occupations are present in the linked sample. Therefore, I am unable to account for differences in occupational composition completely. Re-weighting is not necessary with modern data, as matching rates are 100 percent and fully represent the composition of the full population. The adjustment upward of 3-5 pp. using the geographical weights suggests that the downward sloping trend in persistence rates across the 20th century if anything, is even more significant than shown in Figure A.7.

The rates of rank-rank persistence in the early 1900s are substantially higher than in Norway today, with just above 0.2 for cohorts born in 1980 – representing a reduction in rank-rank persistence of about 50 percent. A comparison with other countries in the early 1900s is difficult because few estimates of persistence in income across generations exist. In a study linking the 1915 Iowa census with the 1940 US census, Feigenbaum (2018) estimates a rank-rank persistence of 0.26 for cohorts born in this period, suggesting that Norway was significantly less mobile than the US. However, newer work by Jácome, Kuziemko and Naidu (2021) using survey data with income for sons and imputed income for fathers finds rates of rank-rank persistence of 0.37. This is still lower than my estimate for Norway, but within the 95 percent confidence interval[8].

I explore heterogeneity in Table A.3 by re-estimating rates of intergenerational persistence with sample splits based on rural status, education and wealth. While persistence rates for the five-year bins are noisy, some interesting patterns emerge for the full sample. Rural areas are somewhat less mobile than urban areas, with rank-rank measures of 0.32 and 0.30, respectively. This is consistent with evidence from the US (Feigenbaum, 2018). Rates of persistence are somewhat higher for those with only primary schooling than those with more, but the difference is not statistically significant. The population with some non-zero wealth have the highest persistence rates, with 0.35 compared with 0.32 for those without wealth.

[8]The estimate by Feigenbaum (2018) seems to match evidence on occupational mobility (Ferrie, 2005; Long and Ferrie, 2013) that finds high levels of mobility compared to European countries, which gradually decline over time. However, new evidence from Ward (2021) corrects for measurement errors in occupational titles and finds a pattern similar to that of Jácome, Kuziemko and Naidu (2021) – low levels of mobility in the early 1900s and then a substantial increase.

4.2 The Rural-Urban Divide

The fact that the rank-rank coefficient for the full sample is larger than that of both rural and urban areas separately means that level differences in income ranks between the areas drive part of the persistence. To investigate this further, I plot the average income rank for children by their father's location in the income distribution in Figure A.8 separately for fathers living in rural and urban areas. The slope of these lines will be equal to the group-specific rank-rank persistence between fathers and sons. To compare changes over time, I show separate panels for the historical sample (cohorts 1910-1935) and the modern sample (1937-1964). I find significant level differences between rural and urban fathers in the historical sample for the same absolute income level. Sons growing up in cities do better in terms of income than somebody growing up in a rural area, even if their fathers have the same income. The grey dashed line is the joint regression line whose slope equals the intergenerational mobility for the combined sample. Interestingly, this line is steeper than the line for rural and urban areas. For cohorts born 1937-1964, the level difference is almost completely gone, and the combined line has a slope similar to the urban and rural lines.

The explanation is that dads from urban areas have significantly higher average income ranks. This point is similar to that of Jácome, Kuziemko and Naidu (2021), who looks at intergenerational mobility in income in the US and the income gap between the white and black part of the population. One can easily imagine this point holds more generally for many groups as income inequality is reduced, and those who previously earned very different incomes see their earnings become more similar. To understand the drivers of the reduction in persistence, I decompose the persistence into drivers related to the rural or urban status of the father's municipality of residence. I follow Jácome, Kuziemko and Naidu (2021) who arrive at the following equation for decomposing into groups $g \in G$:

$$\gamma^{\mathrm{RR}} = \underbrace{12 \times \sum_{g} p_g \,\mathrm{Var}\left(\mathrm{Rank}^p \mid g\right) \gamma_g^{\mathrm{RR}}}_{\text{weighted slopes}} + \underbrace{12 \times \sum_{g} p_g \mathrm{E}\left[\mathrm{Rank}^p \mid g\right] \mathrm{E}[\mathrm{Rank} \mid g] - 3}_{\text{level effect}} \tag{2}$$

where γ^{RR} is the rank-rank persistence for the full population, and $Rank^p$ and $Rank^p$ is the income rank of the parent and child, respectively. γ_g^{RR} is the persistence within subgroup g and p_g is the groups fraction of the total sample. The two terms translate into a weighted average of the group-specific slopes and a between-group component. This decomposition gives a clear intuition as to why the income gap between urban and rural areas shown in Figure A.8 leads to higher persistence across generations.

We know that income inequality was dramatically reduced in Norway between the 1930s and 1950s, and incomes in urban and rural areas converged. I will use the decomposition presented in Equation 2 to understand the impact this had on intergenerational mobility. To reduce noise and provide a clearer picture, I will estimate it based on the whole historical tax register and not just the linked sample. This has the added benefit of allowing me to go further back in time. I simplify the interpretation by asserting that ranks for fathers and sons are measured in the same year and that the expected ranks of fathers and sons in rural and urban areas are the same. Estimates are presented biannually in Figure A.9, starting at almost 16 percent and showing a massive decline in the between-group component of rank-rank persistence. The level differences between rural and urban areas were virtually insignificant by the 1970s – suggesting that the convergence between rural and urban areas can account for a 15 percentage point decrease in rank-rank persistence in this period. I do not currently estimate the contribution of the group-specific slopes because the urban estimates are too noisy.

5 Mechanisms

I have established that rates of relative intergenerational mobility in income in Norway were much higher in the early 1900s, and that a large increase in mobility happened for cohorts born between 1920 and 1940. Using a decomposition framework, I showed that a convergence between rural and urban areas caused a drop in the rank-rank persistence of up to 15 percent throughout this period. In this section, I investigate the possible drivers of this convergence. We can consider possible mechanisms as either pre-market, labor market or post-market. I focus on the two first by looking at education (pre-market) and the returns to education (labor market). I do not look at the direct effect of post-market factors as my measures of income are pre-tax and include few transfers.

5.1 Rural and Urban Education

This section investigates the development of rural and urban education in Norway over the 1900s and its impact on intergenerational mobility in income. I provide a brief overview of educational differences between rural and urban areas in Norway during the 20th century. Then, I estimate the causal effect of an increase in the intensity of primary schooling in this period on intergenerational mobility.

5.1.1 Development of the Rural-Urban Education Gap

Schooling in rural and urban areas was highly segregated in the first part of the 1900s, with one law governing the rural primary schools and one governing urban schools (Acemoglu et al., 2021). While hours of schooling per week were higher in rural schools, weeks of schooling in cities were substantially higher, giving pupils in cities almost twice as many hours in school in the late 1800s. This was justified by the lower need for schooling in rural areas and the need for children to help at home. However, it made it difficult for pupils from rural schools to pursue higher education.

The two schooling systems started converging following the 1936 rural primary school reform, which was the first step towards a unified primary schooling system (Acemoglu et al., 2021; Abel, Buetikofer and Salvanes, 2023). The reform increased central financing for schools, decreased maximum class sizes and, most importantly, increased the minimum number of weeks of schooling during a year by around four weeks (about 30 percent). I show minimum hours of schooling in rural and urban school districts Figure A.10 using data from (Acemoglu et al., 2021). Minimum hours of schooling in rural and urban primary schools had been virtually constant since the last extensive primary school reform in 1889, with rural schooling providing around 55 percent of that in urban schools. The 1936 rural primary school reform increased rural schooling to about 75 percent of urban schools. The gap was not closed before 1959 when the two schooling systems became governed by the same law, both providing a minimum of 5,814 hours of schooling during primary school.

The period following the cohorts born in the 1920s also saw large convergences among several dimensions of educational attainment. I plot differences in the share with different educational attainment in Figure A.11a. It shows that gaps in educational attainment started large for cohorts born in the 1910s and were gone by the late 1960s. It began with rural pupils increasingly moving on to middle schools and were overrepresented in this group after the early 1930s. The final wave of convergence occurred between the early 1930s and late 1960s when pupils born in rural areas moved into higher education (here defined as 13 or more years of completed education). Measured as years of schooling, the difference in educational attainment between rural and urban areas increased up until the 1920s, and it didn't decrease significantly until the 1940s. The gap in completed years of education was gone in the late 1970s.

For the changes in educational attainment in Norway to have impacted intergenerational mobility in Norway, we would expect that it happened through an increase in persistence in educational attainment. I estimate the persistence in education between fathers and sons as follows:

$$\text{Child Education}_i = \alpha + \beta \times \text{Parent Education}_i + \epsilon_i \tag{3}$$

where Child Education$_i$ and Parent Education$_i$ are the years of completed education for child i and the parent of child i, respectively. β is the parameter of interest and should be interpreted as the persistence in years of education across generations. The results are estimated in 5-year birth cohort bins and are shown in Figure A.12. The persistence was above 0.5 for men and just below 0.5 for women until the late 1930s, which coincides with the convergence in years of schooling between rural and urban areas. It is unclear if this is the effect of an increase in education at the bottom or a more general increase in mobility. We would usually construct rank-rank measures to remove the effect of changes in inequality, but this is not possible as large fractions of the population have the same years of completed schooling, making it highly sensitive to how you deal with ranking ties. As a robustness check, I re-estimate intergenerational mobility using a sample of surveys similar to Mjølsnes (2019). I present the results in Appendix C and find that it is qualitatively very similar to the estimates based on register data. Still, the survey-based estimates have much higher persistence in the early 1900s.

5.1.2 The 1936 Primary School Reform

To get at the causal effect of an increase in schooling, I estimate the causal effect of education on intergenerational mobility in income using the 1936 Norwegian primary school reform as exogenous variation (Acemoglu et al., 2021; Abel, Buetikofer and Salvanes, 2023). The reform increased the minimum allowed weeks of schooling for rural municipalities in Norway and increased funding for schools, teachers and teaching materials. I follow Acemoglu et al. (2021) and Abel, Buetikofer and Salvanes (2023) and use an intensity of treatment design to estimate the causal effect of the reform. The reform had a different 'bite' depending on how many weeks of schooling the municipality had to increase teaching to reach the new minimum. The minimum weeks of schooling were 12 (1-3 grade) and 14 (4-7 grade) weeks before the reform and 16 (1-3 grade) and 18 (4-7 grade) weeks after the reform. Let $b_j^{\text{små}}$ be the increase necessary in weeks of schooling for 1-3 grade and b_j^{stor} the necessary in weeks of schooling for 4-7 grade. The intensity of treatment for municipality j is then given by P_j:

$$P_j = \frac{3 \times b_j^{\text{små}} + 4 \times b_j^{stor}}{28} \tag{4}$$

where P_j can be interpreted as the share of the full reform experienced by pupils starting primary school in municipality j after the reform is implemented. In addition to the geographical variation in exposure, exposure to the reform also varies by how many years of your time in primary school is during the reformed system. The measure Z_{jt} is the sum of increases in weeks of schooling induced by the reform for cohort t in municipality j. Similar to Acemoglu et al. (2021) and Abel, Buetikofer and Salvanes (2023), I calculate it as:

$$Z_{jt} = \frac{\sum_{a=5}^{7} b_j^{\text{små}} \mathbb{1}(treat_t = a) + \sum_{a=1}^{4} b_j^{stor} \mathbb{1}(treat_t = a)}{28} \tag{5}$$

where $\mathbb{1}(treat_t = a)$ indicate whether individual i experienced a years of schooling in the reformed primary school system. My regression setup assumes that the impact of the reform is linear in treatment intensity. In my first regression setup, I estimate the impact of the reform on individual incomes specified as follows:

$$Y_{ijt} = \alpha + \beta Z_{jt} + \sum_l \gamma_l (d_l \times \mathbf{X_j}) + \theta_j + \eta_t + \varepsilon_{ijt} \tag{6}$$

where Z_{jt} represents the proportion of the full reform that an individual born in year t in municipality j undergoes. The term $d_l \times \mathbf{X_j}$ encapsulates the interaction between year-of-birth fixed effects and certain municipality-level control variables. The coefficient β is of primary interest as it quantifies the effect of transitioning from the previous legal minimum requirement $Z_{jt} = 1$ to the updated requirements $Z_{jt} = 0$. Based on the linearity assumption, the influence of progressing from 1 to 2 years of education under the reformed system is equivalent to that of advancing from 6 to 7 years.

When estimating the impact of the reform on intergenerational mobility in income using a rank-rank setup, I interact the reform exposure with the father's income rank. This setup is similar to that of Bütikofer, Dalla-Zuanna and Salvanes (2022). I use cohorts born between 1910 and 1950. Income ranks are averaged at ages 30 to 35 for sons and 55 to 60 for fathers, similar to in Equation 1. My regression looks like the following:

$$Y_{ijt}^{son} = \alpha + Y_{ijt}^{father} + Y_{ijt}^{son} \times \beta_1 Z_{jt} + \beta_2 Z_{jt} + \sum_l \gamma_l (d_l \times \mathbf{X_j}) + \theta_j + \eta_t + \varepsilon_{ijt} \tag{7}$$

β_1 is the reform's effect on intergenerational mobility measured as rank-rank persistence. If the reform disproportionately increases incomes at the lower part of the income distribution, we would expect this term to be negative. β_2 is the intergenerational mobility when controlling for the exposure of the reform.

5.1.3 Effects on Intergenerational Mobility

Table A.4 presents results from the simple causal setup from Equation 6, as well as for the regression investigating the effect on intergenerational mobility in Equation 7. Starting with years of education in column 1, I find that the reform increased years of education by 0.57 years of schooling. This increase is somewhat larger than the one found in previous papers and less precise because of the smaller sample linked to their fathers (Abel, Buetikofer and Salvanes, 2023). In column 2, I estimate the reform's effect on intergenerational mobility in education but find no statistically significant effect.

For income, I find in column 3 that the reform increased the income rank of cohorts affected by 0.058, a smaller effect than in previous research using average income over a more extended period (Abel, Buetikofer and Salvanes, 2023). However, in column 4, I find a negative impact of the primary school reform on intergenerational mobility of 5.3 percent at the 10 percent significance level. This suggests that reform increased incomes for sons with low-income fathers more than others. These estimates are less precise than wanted because of the limited sample size but indicate that the primary school reform had a sizable impact on intergenerational mobility.

5.2 The Role of Educational Premiums and Income Inequality

In a model of intergenerational mobility where income is determined at least partly by educational attainment, a change in the returns to education can impact rates of relative intergenerational mobility in income (Becker and Tomes, 1979, 1986). An increase in the returns to education would, for example, result in non-borrowing-constrained parents increasing investments in their child's human capital (Corak, 2013). Alternatively, higher returns to education would tend to increase rates of intergenerational persistence if ability is strongly inherited across generations. This makes estimating the returns to education central to understanding changes in intergenerational mobility.

Despite the implications of the returns to education for investments in human capital and intergenerational mobility more broadly, comprehensive estimates of the returns to education over the long run are hard to find, with the partial exception of the US (Goldin and Margo, 1992; Goldin and Katz, 2009). Minde (1998) study the returns to education in Norway between 1885 and 1955 and find a long-term decline in the profitability of schooling. However, he looks only at a sample of government employees and only has data for every decade.

5.2.1 Estimating Returns to Years of Education

I estimate the yearly returns to education in Norway between 1925 and 2014 and use the historical tax register developed in Abel and Salvanes (2023) linked with administrative data and completed years of education. The sample is selected to represent a consistent geographical area throughout the period but is weighted to reflect the ratio of rural to urban inhabitants in Norway. Specifically, I estimate the following Mincer equation separately for every year:

$$\ln(\mathrm{Y})_i = \alpha + \beta \mathrm{education}_i + \gamma_1 \mathrm{experience}_i + \gamma_2 \mathrm{experience}_i^2 + \epsilon_i \tag{8}$$

where $\ln(\mathrm{Y})_i$ is log income for person i restricted to males between the ages of A and B. experience is the potential years of experience (Mincer, 1974).[9] The coefficient of interest is the returns to years of education β, which can be interpreted as the percent increase in income from a one-year increase in education.

The returns to education from estimating Equation 8 are shown in Figure A.13. The returns to an additional year of education fluctuated between more than 15 percent in the mid-1930s and almost 5 percent in 1980. The majority of this fall came in the years between 1935 and 1942. At the start of this period, the economy had started growing again and the Norwegian Labor Party came into power. During the beginning of the German occupation 1940-42 the returns to years of education started dropping sharply, following a similar dramatic fall in income inequality (Abel, Abramitzky and Salvanes, 2023). The second fall happened between 1970 and 1980, typically seen

[9] Calculated as experience = age − 7 − years of education.

as the end of the post-war area and the start of a period characterised by a political right-wing wave. Like in other Western countries, I find that the returns to education started rising in the 1980s and are today back at the levels last seen in the decades following World War II.

It is unclear if we should think of the development in the returns to education since 1935 as the start of a new area or the return to one with low levels of returns to education. I find relatively low levels of returns in the 1920s, which increased sharply between 1925 and 1935. The 1920s was a particularly turbulent period in the Norwegian economy, starting with a post-war recession, severe deflation and a debt crisis. Ties had turned by 1930, the strongest boom for the Norwegian economy ever recorded, with GDP levels more than 8 percent above trend (Eika, 2008). Considering the turbulent times and the findings by Minde (1998) of declining returns to education since the late 1800s, it is likely that returns before and during World War I were significantly higher than in the 1920s, and we should consider the levels during the 1930s as closer to the 'old' normal.

5.2.2 The Impact of Returns To Education on Intergenerational Mobility

Separating the direct and indirect effects of changes in the returns to education is challenging because we usually do not have exogenous variation in the returns to schooling across cohorts or years. I use the sharp drop in the returns to education in 1940 to estimate the direct – or mechanical – effect of a decrease in the returns to years of education on intergenerational mobility. I estimate intergenerational mobility separately for each income year between 1935 and 1935 and keep the cohorts fixed to control for any responses to the investment in human capital. This ensures that incomes for the son's generation, their levels of education and the father's education are constant, allowing only changes in the father's income. Observations on fathers are always from the historical tax data, while observations on sons are from the modern administrative data.

I plot estimates of persistence in Figure A.14 with polynomial trends fitted pre- and post-shock and excluding 1940, which is partly impacted by the shock[10]. I find that rank-rank persistence in income decreased from about 11 percent to 8 percent after 1940 – representing a 27 percent drop in persistence. Levels are lower than my main estimates because it only uses one year's income. Shocks as large as this are rare, but it shows clearly how income inequality and the returns to education can impact intergenerational mobility.

Importantly, we should not think of the shock in 1940 as a pure shock to the returns to years of education but as a complex shock to the economy hitting different skill and education groups differently and not always proportional to years of educational attainment. See Abel, Abramitzky and Salvanes (2023) for more details on this specific shock or Goldin and Margo (1992) on a similar fall in the returns to education in the US.

[10]The occupation started in April 1940 and Norwegian forces finally capitulated June 1940. The Norwegian economy was at a standstill during this period, but the activity in the economy increased substantially during the fall of 1940.

6 Conclusion

In cross-country comparisons, the Nordic welfare states outrank most other countries on measures of relative intergenerational mobility in income - meaning that relative incomes in Norway, Denmark and Sweden are less tied to that of their parents than in almost any other country. This has led to strong interest from researchers and policymakers in understanding the causes behind the high mobility rates. Existing evidence shows that trends in intergenerational mobility income in the Scandinavian countries have been flat for male cohorts born between 1951 and 1979 (Ahrsjö, Karadakic and Rasmussen, 2023). If the welfare state increased mobility, it must have happened for earlier cohorts.

This paper studies long-run trends in intergenerational mobility in income for cohorts born from 1910 to 1980 and examines its determinants. I find that persistence in income rank across generations for cohorts born in the 1910s and 1920s was more than twice as high as today's – with a rank-rank persistence of about 0.55 compared to 0.20 today. The decrease in persistence is isolated to cohorts born in the early 1920s to 1940s, and the level has remained relatively constant. This shows that Norway has not always been the mobile society it is today and that this mobility must have been caused by something happening primarily in the first part of the 1900s. Cohorts born in cities from 1920 to 1935 earn better than those growing up in rural areas, even if their fathers have the same income. Given a set of assumptions, this reduction has contributed to a 16 percentage point reduction in rank-rank persistence in this period.

The education gap between rural and urban areas was large and relatively stable until the 1920s, when the share with more than primary schooling started increasing in rural areas relative to urban areas. Gaps in primary, secondary and higher education were closed sequentially and were virtually closed for cohorts born in the 1960s. To provide causal evidence of this mechanism, I use the 1936 rural primary school reform in Norway as causal identification and find that it increased incomes significantly and decreased persistence by more than five percentage points.

I also investigate if there are changes in educational premiums over time that could change the incentives to invest in education or how ability is priced in the labor market. Educational premiums were around 15 percent in the 1930s before dropping sharply at the start of World War II to about 8 percent. A second reduction in educational premiums followed in the 1970s, before trending upwards, as in many Western countries (Goldin and Katz, 2009; Autor, 2019). While the educational premium can impact investment in human capital, it can also impact intergenerational mobility in income more directly if ability is inherited strongly across generations. Using the drop in premiums at the start of World War II as a shock, I hold my sample fixes and find that the war significantly reduced persistence in income by around 30 percent.

Research into the long-run trends of mobility has been hampered by poor access to data. This research is made possible by extending the Norwegian tax registers almost 50 years back in time – to the early 1920s – using a novel individual-level dataset of more than 16 million observations from Norwegian tax records for the period from 1924-1964 (Abel and Salvanes, 2023). I link these observations with names to the modern administrative data, and historical population censuses to get links between family members. Previous research has used data on occupations from population censuses (Long and Ferrie, 2013), status information in first names (Olivetti and Paserman, 2015) and rear surnames (Clark, 2015). These methods typically find slow-moving trends or no trends at all, which stands in stark contrast to the dramatic changes found in this paper. More recent developments however find qualitatively similar patterns as found in this paper (Ward, 2021; Jácome, Kuziemko and Naidu, 2021).

References

Aaberge, Rolf, and Anthony B Atkinson. 2010. "Top Incomes in Norway." *Top Incomes: A Global Perspective*, 2: 448–480.

Aaberge, Rolf, Anthony B Atkinson, and Jørgen Modalsli. 2020. "Estimating Long-run Income Inequality from Mixed Tabular Data: Empirical Evidence from Norway, 1875–2017." *Journal of Public Economics*, 187: 104196.

Aaberge, Rolf, Jørgen Modalsli, and Edda Solbakken. 2018. "Measuring Long-run Wealth Inequality." *Working Paper*.

Aaronson, Daniel, and Bhashkar Mazumder. 2011. "The impact of Rosenwald schools on black achievement." *Journal of Political Economy*, 119(5): 821–888.

Abel, Eirik Berger, Aline Buetikofer, and Kjell Gunnar Salvanes. 2023. "Fertility, Family Formation and Human Capital." Working paper.

Abel, Eirik Berger, and Håkon Block Vagle. 2017. "Equality before the welfare state: the Norwegian income distribution 1892-1929." Master's diss. The Norwegian School of Economics.

Abel, Eirik Berger, and Kjell Gunnar Salvanes. 2023. "Digitizing Norwegian Tax Returns from the 20th Century: An Automated Machine Learning Approach." *Working paper*.

Abel, Eirik Berger, Ran Abramitzky, and Kjell Gunnar Salvanes. 2023. "The Making of Equality: How the Second World War Shaped the Norwegian Income Distribution." *Working paper*.

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

Acemoglu, Daron, Tuomas Pekkarinen, Kjell Gunnar Salvanes, and Matti Sarvimäki. 2021. "The Making of Social Democracy: The Economic and Electoral Consequences of Norway's 1936 Folk School Reform." National Bureau of Economic Research.

Ager, Philipp, Marc Goñi, and Kjell Gunnar Salvanes. 2023. "Gender-Biased Technological Change: Milking Machines and the Exodus of Women from Farming." *working paper*.

Ahrsjö, Ulrika, René Karadakic, and Joachim Kahr Rasmussen. 2023. "Intergenerational Mobility Trends and the Changing Role of Female Labor." *arXiv preprint arXiv:2302.14440*.

Autor, David H. 2019. "Work of the Past, Work of the Future." Vol. 109, 1–32, American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203.

Bailey, Martha J, Connor Cole, Morgan Henderson, and Catherine Massey. 2020. "How well do automated linking methods perform? Lessons from US historical data." *Journal of Economic Literature*, 58(4): 997–1044.

Becker, Gary S, and Nigel Tomes. 1979. "An equilibrium theory of the distribution of income and intergenerational mobility." *Journal of political economy*, 87(6): 1153–1189.

Becker, Gary S, and Nigel Tomes. 1986. "Human capital and the rise and fall of families." *Journal of labor economics*, 4(3): 1–39.

Berger, Thor, Per Engzell, Björn Eriksson, and Jakob Molinder. 2023. "Social mobility in Sweden before the welfare state." *The Journal of Economic History*, 83(2): 431–463.

Bhuller, Manudeep, Magne Mogstad, and Kjell G Salvanes. 2017. "Life-cycle earnings, education premiums, and internal rates of return." *Journal of Labor Economics*, 35(4): 993–1030.

Björklund, Anders, Markus Jäntti, et al. 1999. *Intergenerational mobility of socio-economic status in comparative perspective.* Social Policy Research Centre, University of New South Wales.

Black, Sandra E, and Paul J Devereux. 2011. "Recent developments in intergenerational mobility." *Handbook of labor economics*, 4: 1487–1541.

Black, Sandra E, Paul J Devereux, and Kjell G Salvanes. 2005. "Why the apple doesn't fall far: Understanding intergenerational transmission of human capital." *American economic review*, 95(1): 437–449.

Bryde, S. M. 1942. *Adressebok for Hordaland fylke og Bergen med skatteligninger.* Utg. 30(1986) med tittel: Adressebok med skatteligningstall for Hordaland fylke. 25(1967)-27(1971) med tittel: Adressebok for Hordaland fylke med firmafortegnelse for Bergen og med skatteligninger.

Bütikofer, Aline, and Kjell Gunnar Salvanes. 2020. "Disease control and inequality reduction: Evidence from a tuberculosis testing and vaccination campaign." *The Review of Economic Studies*, 87(5): 2087–2125.

Bütikofer, Aline, Antonio Dalla-Zuanna, and Kjell Gunnar Salvanes. 2022. "Breaking the links: Natural resource booms and intergenerational mobility." *Review of Economics and Statistics*, 1–45.

Bütikofer, Aline, Eirin Mølland, and Kjell Gunnar Salvanes. 2018. "Childhood nutrition and labor market outcomes: Evidence from a school breakfast program." *Journal of Public Economics*, 168: 62–80.

Bütikofer, Aline, Katrine V Løken, and Kjell Gunnar Salvanes. 2019. "Infant health care and long-term outcomes." *Review of Economics and Statistics*, 101(2): 341–354.

Bütikofer, Aline, René Karadakic, and Kjell Gunnar Salvanes. 2021. "Income inequality and mortality: a Norwegian perspective." *Fiscal Studies*, 42(1): 193–221.

Card, David, Ciprian Domnisoru, and Lowell Taylor. 2022. "The intergenerational transmission of human capital: Evidence from the golden age of upward mobility." *Journal of Labor Economics*, 40(S1): S39–S95.

Clark, Gregory. 2015. *The Son Also Rises: Surnames and the History of Social Mobility: Surnames and the History of Social Mobility.* Princeton University Press.

Corak, Miles. 2013. "Income Inequality, Equality of Opportunity, and Intergenerational Mobility." *Journal of Economic Perspectives*, 27(3): 79–102.

Eika, Torbjørn. 2008. "Det svinger i norsk økonomi." Published on December 8, 2008.

Feigenbaum, James J. 2018. "Multiple measures of historical intergenerational mobility: Iowa 1915 to 1940." *The Economic Journal*, 128(612): F446–F481.

Ferrie, Joseph P. 2005. "History lessons: The end of American exceptionalism? Mobility in the United States since 1850." *Journal of Economic Perspectives*, 19(3): 199–215.

Gerdrup, Karsten R. 1998. "Skattesystemet og skattestatistikk i et historisk perspektiv." *Statistics Norway*.

Goldin, Claudia, and Lawrence F Katz. 2009. *The race between education and technology.* harvard university press.

Goldin, Claudia, and Robert A Margo. 1992. "The Great Compression: The Wage Structure in the United States at Mid-century." *The Quarterly Journal of Economics*, 107(1): 1–34.

Grytten, Ola. 2008. "The economic history of Norway." *EH. Net Encyclopedia, edited by Robert Whaples*.

Grytten, Ola Honningdal. 2020. "The Wealth of a Nation: Norways Road to Prosperity." *NHH Dept. of Economics Discussion Paper, No. 17*.

Ingulstad, Mats, Gunnar D Hatlehol, and Hans Otto Frøland. 2017. "Regimenting Labour in Norway during Nazi Germany's Occupation." *Working Papers of the Independent Commission of Historians Investigating the History of the Reich Ministry of Labour (Reichsarbeitsministerium) in the National Socialist Period*.

Jácome, Elisa, Ilyana Kuziemko, and Suresh Naidu. 2021. "Mobility for all: Representative intergenerational mobility estimates over the 20th century." National Bureau of Economic Research.

Karadakic, René. 2023. "Consequences of Rapid Structural Change - Evidence from Hydropower Expansions." *working paper*.

Karlson, Kristian, and Rasmus Landersø. 2021. "The making and unmaking of opportunity: Educational mobility in 20th century-denmark." *working paper*.

Katz, Lawrence F, and Kevin M Murphy. 1992. "Changes in relative wages, 1963–1987: supply and demand factors." *The quarterly journal of economics*, 107(1): 35–78.

Leknes, Stefan, and Jørgen Modalsli. 2018. "Who Benefited from Industrialization? The Local Effects of Hydropower Technology Adoption." *Discussion Papers*.

Lie, Einar. 1995. *Ambisjon og tradisjon: Finansdepartementet 1945-1965.* Universitetforl.

Long, Jason, and Joseph Ferrie. 2013. "Intergenerational occupational mobility in Great Britain and the United States since 1850." *American Economic Review*, 103(4): 1109–1137.

Mincer, Jacob. 1974. "Schooling, Experience, and Earnings. Human Behavior & Social Institutions No. 2."

Minde, Kjell Bjørn. 1998. "Was there a Fall in the Return to Educational Investment in Norway, 1885–1955?" *Scandinavian Economic History Review*, 46(3): 6–22.

Mjølsnes, Ingvald. 2019. "Evolution of Intergenerational Mobility in Norway between 1900 and 1945." Master's diss.

Modalsli, Jørgen. 2017. "Intergenerational Mobility in Norway, 1865–2011." *The Scandinavian Journal of Economics*, 119(1): 34–71.

Mogstad, Magne, and Gaute Torsvik. 2023. "Family background, neighborhoods, and intergenerational mobility." *Handbook of the Economics of the Family*, 1(1): 327–387.

Myhre, Jan Eivind. 2022. *Den norske modellen blir til: Norge 1814-1914.* Dreyers forlag.

Norsk Regnesentral. 2023. "The Norwegian Historical Population Register." Accessed: 2023-10-16.

Olivetti, Claudia, and M. Daniele Paserman. 2015. "In the Name of the Son (and the Daughter): Intergenerational Mobility in the United States, 1850-1940." *American Economic Review*, 105(8): 2695–2724.

Pekkarinen, Tuomas, Kjell Gunnar Salvanes, and Matti Sarvimäki. 2017. "The evolution of social mobility: Norway during the twentieth century." *The Scandinavian Journal of Economics*, 119(1): 5–33.

Pekkarinen, Tuomas, Roope Uusitalo, and Sari Kerr. 2009. "School tracking and intergenerational income mobility: Evidence from the Finnish comprehensive school reform." *Journal of Public Economics*, 93(7): 965–973.

Regidor, Enrique, Cruz Pascual, David Martínez, María E Calle, Paloma Ortega, and Paloma Astasio. 2011. "The role of political and welfare state characteristics in infant mortality: a comparative study in wealthy countries since the late 19th century." *International Journal of Epidemiology*, 40(5): 1187–1195.

Seip, Anne-Lise. 1994. *Veiene til velferdsstaten: norsk sosialpolitikk 1920-1975.* Gyldendal.

Sejersted, Francis. 2021. *The age of social democracy.* Princeton University Press.

SSB. 1940. "NOS IX 189 - Skattestatistikken for budsjettåret 1939/40."

Venneslan, Christian. 2009. "Electrification and industrialisation: An assessment of the industrial breakthrough in Norway." *Scandinavian Economic History Review*, 57(2): 124–155.

Ward, Zachary. 2021. "Intergenerational mobility in American history: Accounting for race and measurement error." National Bureau of Economic Research.

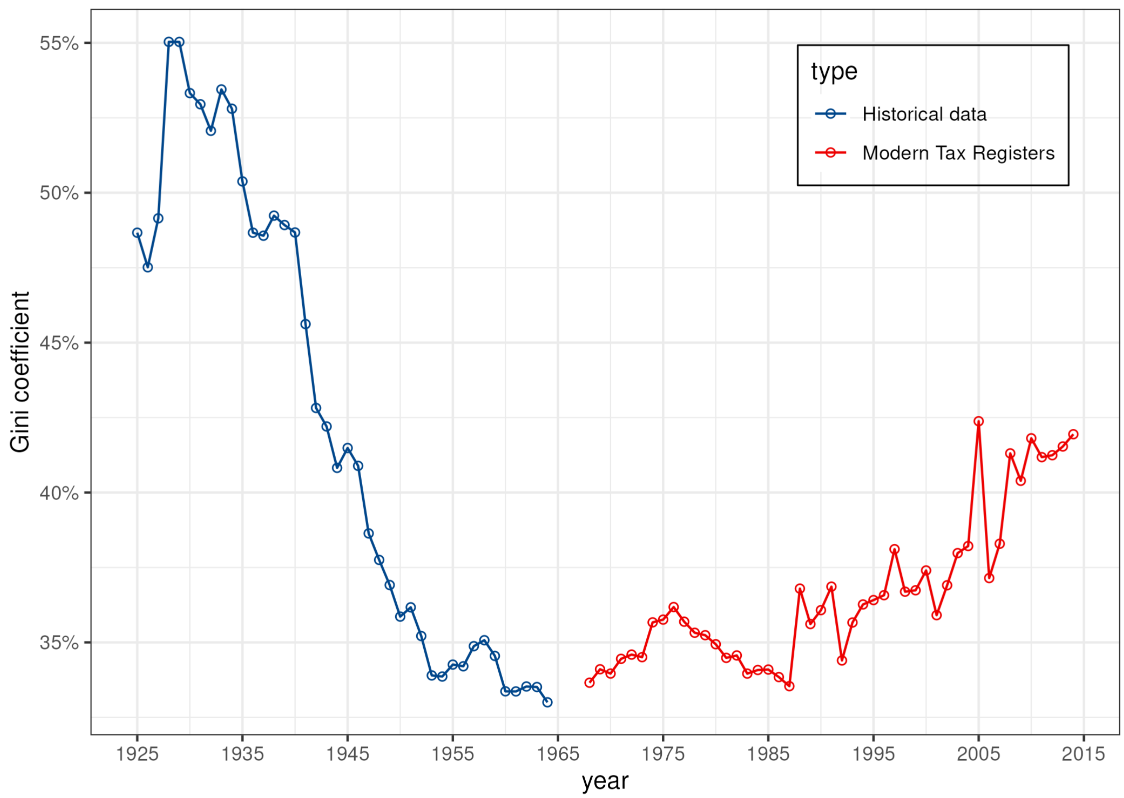
Online Appendix of:

Living the American Dream::
How Norway Became a High-Mobility Country

Eirik Berger

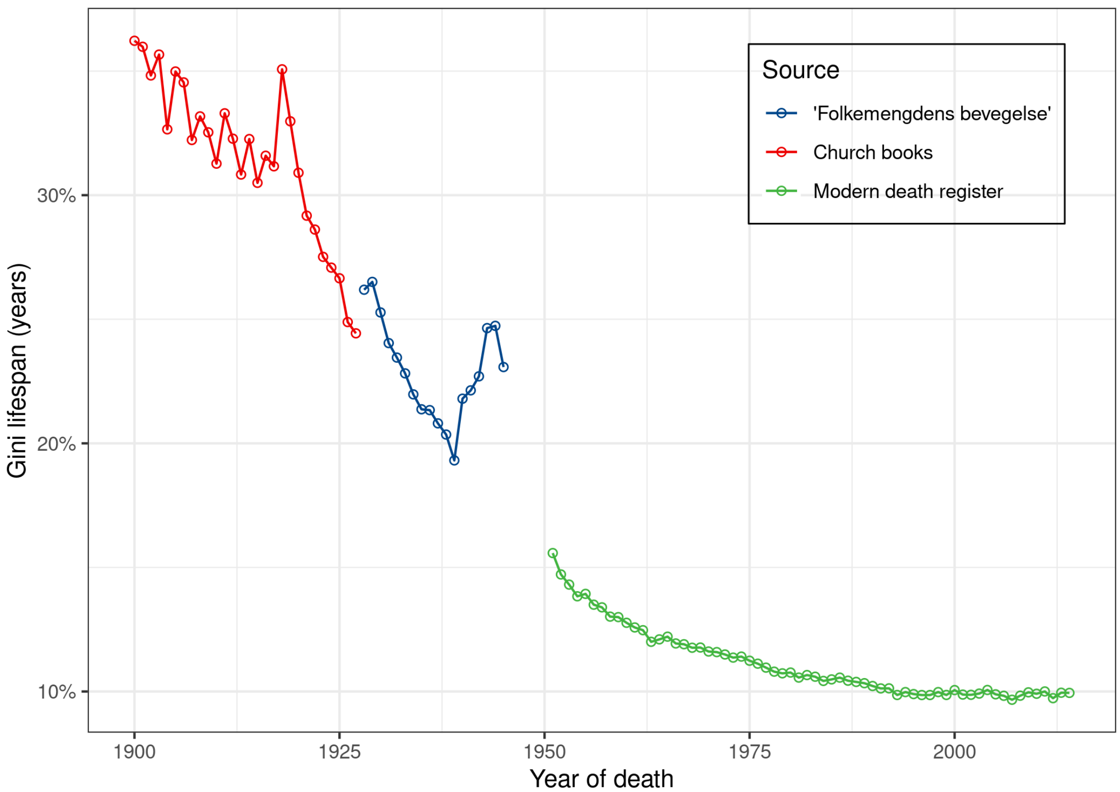
A Tables and Figures

Figure A.1: Pre-tax Income Gini Coefficient for Taxpayers in Norway



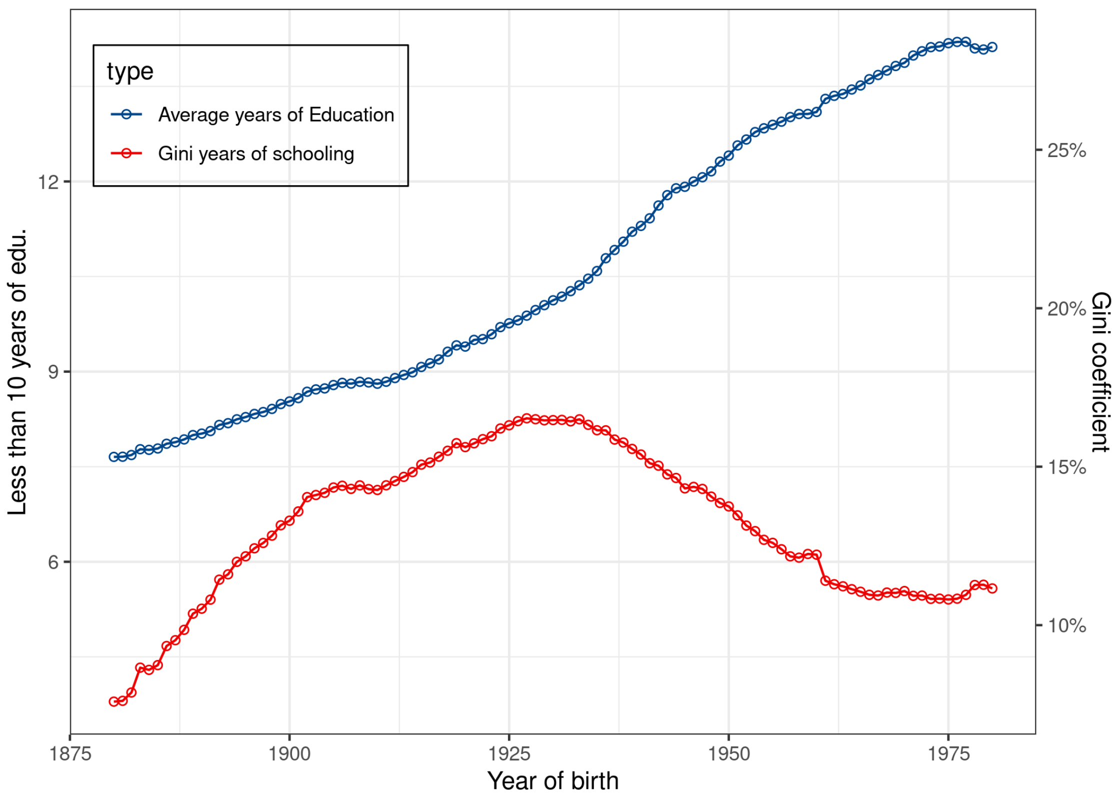
Note: The figure shows Norway's pre-tax income Gini coefficient between 1925 and 2014 from Abel, Abramitzky and Salvanes (2023). It is based on newly digitized individual-level data (1925-1964) and Norwegian administrative data from tax authorities (1967-2014).

Figure A.2: Gini Coefficient for Age at Death



Note: The figure shows the Gini coefficient for age of death for birth cohorts from 1900 to 1914. The coefficients are based on individual-level data and calculations by the author. Data from 1900-1927 originates from church books and is owned by The National Archives of Norway. Data from 1928-1945 comes from "Folkemengdens Bevegelse" by Statistics Norway. Data from 1951-2014 are from Medical Birth Registers (MBRN) at the Norwegian Institute of Public Health.

Figure A.3: Gini Coefficient for Years of Education



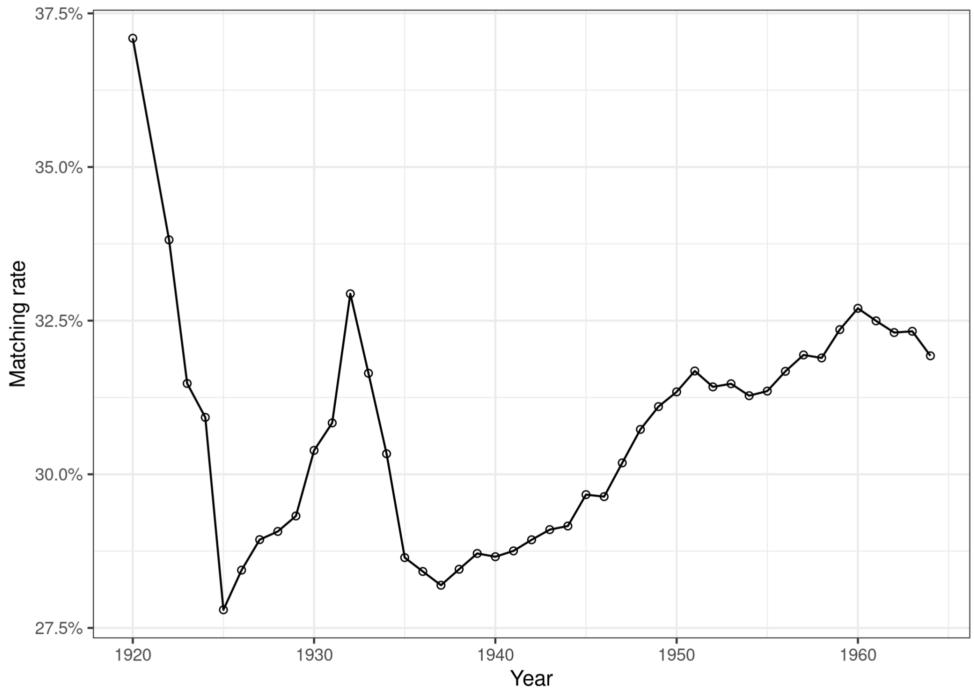
Note: The figure shows the Gini coefficient for years of education and the fraction with less than ten years of schooling. The measure of years of education is the max across the 1960, 1970, 1980 censuses and the Norwegian Education Register. See Abel, Buetikofer and Salvanes (2023) for details.

Figure A.4: Example Page from Address Books

Johannesen, Johanne F, sykegymn., C Michelsensg. 2, 0, 3800
— Johannes, disp., Engen 28, 0, 1000
— Johannes, mask., Klostret 20, 0, 5000
— Johannes, politikonst., Strømg. 28, 0, 5400
— Johannes, pølsem., Kroken 9 a, 0, 3400
— Johannes, renovatør, Roth.g. 5 b, 0, 4200
— Johannes B, løsarb., J Liesv. 40, 0, 4000
— Johannes E, portef.maker, Stubsg. 2, 5000, 5200
—Johannes M, sjåfør, Håkonsg. 11, 0, 800
— John, forsk.snekker, Klosterhaugen 9, 0, 3700
— John, konduktør, Stubsg. 3, 0, 4900
— John, konduktør, Welh.g. 57, 0, 2900
— John, snekker, Kjellersm. 9
— John M, overtrykker, Tartarg. 18, 0, 4600
— Jonas O, portier, Sigurdsg. 21, 0, 3600
— Julius, modellsnekker, Bispengg. 17, 0, 3800
— Jørgen, tømmerm., Hellev. 30 c, 0, 4000
— Kaia, bileierske, Birkebeinerg. 18, 5400, 0
— Kanutta, Sigurdsg. 30, 14,000, 2700
— Karl, skredderm., Strandg. 7, 0, 1000
— Karl, snekker, Verftsg. 3, 0, 2400
— Karl, stuert, Bredenbecksm. 3, 0, 2400
— Karl J, donkeym., N Stølen 17, 0, 3400
— Karl J, smedsv., Krohnengg. 37, 0, 4100
— Karl Johan, havnearb., Kirkeg. 33, 0, 3400
— Karl S, lagerarb., Prahlsv. 16 a, 0, 3100
— Kathrine L, husmor, Stølesm., 6200, 0
— Klara Misje, kass., Lambertsv. 6
— Knut, urmakersv., Bjørndalen 42 b, 0, 4700
— Kristian Bang, se Johannesen, Harald J
— Kristian J, K Ockensm. 4, 0, 2300
— Lars, bergingsforr., sleping, transport, pramutleie, utleie av dykkerapp., Sandviksv. 51, 6900, 7000
— Lars A M, skipssnekker, Nygårdsg. 66, 0, 2700
— Lars Aliks M, taugbåtf., Bøkkerg. 2, 0, 4900
— Lars L, skreddersv., Breistølen 27, 7000, 4250
— Laura, kjøttforr., Byens basar, bor Strandg. 25, 0, 5100
— Leif E, Roth.g. 1 b
— Leif Johan, fabr.arb., Helgesensg. 8, 0, 3550
— Leiv J, hjelpearb., Finnbergåsen 84, 0, 4200
— Lilly Bang, se Johannesen, Harald J
— Liv Holm, Rossm. 1, 0, 2800
— Louise, huseier, H Wergel.g. 7, 9000, 0
— Louise S, losji, C Ockensm. 4
— Ludolf O, mask., Kirkeg. 15 a, 0, 4000
— M A, kol.forr., Sandv.v. 63
— og Madsen (Erling Johannesen og Kristian Madsen), frukt- og grønnsaker en gros, Torvg. 1 d
— Magnus A, betj., V Holb.alm. 23, 0, 3900
Johannesen, Magnus B, lagerarb., Nordnesgt. 35, 0, 3100
— Malvin E, kontorsjef, Hellev. 33, 5000, 8300
— Margit M, syerske, Nøsteg. 71 c, 5000, 1900
— Marie, huseierske, Teaterg. 13, 9400, 0
— Marie, huseier, H Wergel.g. 12, 10,000, 1400
— Marie, kontord., Heien 1, 0, 4800
— Marie Anna, fhv. lærerinne, Holb.alm. 8, 16,00, 4600
— Marie L, huseier, Strømg. 33, 18,000, 8900
— Marius J, sjåfør, M Krohnsg. 40 b, 0, 1850
— Marius K, malersv., Lien 50, 0, 4000
— Martha, arb., Skivebk. 32 b, 17,000, 0
— Marthin, snekker, Sandv.v. 49 a, 0, 3300
— Martin B, fyrb., Fossw.g. 67, 0, 4100
— Martin Th., garver, Edvardsensg. 32, 0, 3250
— Mathias, mask., Christinegård 4, 22,000, 8450
— Mathilde, sykepl., Landåssm. 2, 0, 4200
— Mikal, bygn.tømmerm., F Meltzersg. 6, 0, 4000
— Monrad, montør, Wesselsg. 10, 0, 5100
— Mons, vaktm., C Sundtsg. 22, 21,000, 3300
— Nils, kontorist, Gimlev. 36 k, 0, 5600
— Nils A, bokb., Ø Sandv.v. 26, 0, 1400
— Nils M, lastebileier, Skostr. 9 b, 0, 2400
— Nils M, tilsynsm., Banev. 27, 7000, 5900
— Odd, kontorist, Tartarg. 15 a, 0, 3700
— Odd, sjåfør, Dokkev. 1, 0, 1800
— Odin Johan, støperiform., S Skogv. 22 b, 0, 3800
— Odvard, sjåfør, Sverresborggrend 2 a
— Olaf, bygn.snekker, Teaterg. 12, bor Kjellersm. 9, 0, 5300
— Olaf, glassliper, Øvreg. alm. 2, 0, 4700
— Olaf, prokurist, Svanev.v. 31 c, 0, 6100
— Olaf B Molvik, mask., Formannsv. 15 a, 15,000, 4900
— Olaf H, motorbåtf., Torvg. 1 a, 0, 3400
— Olaf J avd.sjef, Gimlev. 36 m, 0, 7000
— Olaf N, rigger, Klosterg. 24, 0, 3800
— Olav E, arb., Ladeg.g. 39 d, 0, 650
— Olav J, renh.arb., Skutev.tverrg. 6, 0, 4300
— Olav J, sjåfør, Damsgårdsv. 98, 0, 4100
— Ole, se Flaskecentralen
— Oleanne, strykerske, Øisteinsg. 6, 0, 2600
— Oluffa S, syerske, Prahlsv. 20 a, 0, 2100
— Oscar A, motorbåtf., Korkeg 16, 0, 3600
— Oscar H, kontorist, Lien 20 b, 0, 4100
— Otto, måleravleser, Elstersg. 2, 0, 4500
— Otto J L, posteksped., C Frimannsg. 4, 0, 4600
— Peder J, form., Bohrsg. 6, 0, 6900
— Peder J, salm, Stølesm. 4 a, 7100, 0
— Perny L, Tullinsg. 1, 0, 1700
— Peter, vaktm., Vågsalm. 1, 0, 4000
— Ragnhild, huseier, N Blekev. 4, 6500, 0

Note: The figures show a page from an address book with information from tax records. Source: *Adressebok for Hordaland fylke og Bergen med skatteligninger. 1942 Vol. 14* (1942).

Figure A.5: Matching Rates



Note: The Figure shows matching rates between the historical tax data and the administrative data as a fraction of the number of yearly observations in the historical tax data. The cutoff for the Jaro-Winkler string distance is set to 0.8.

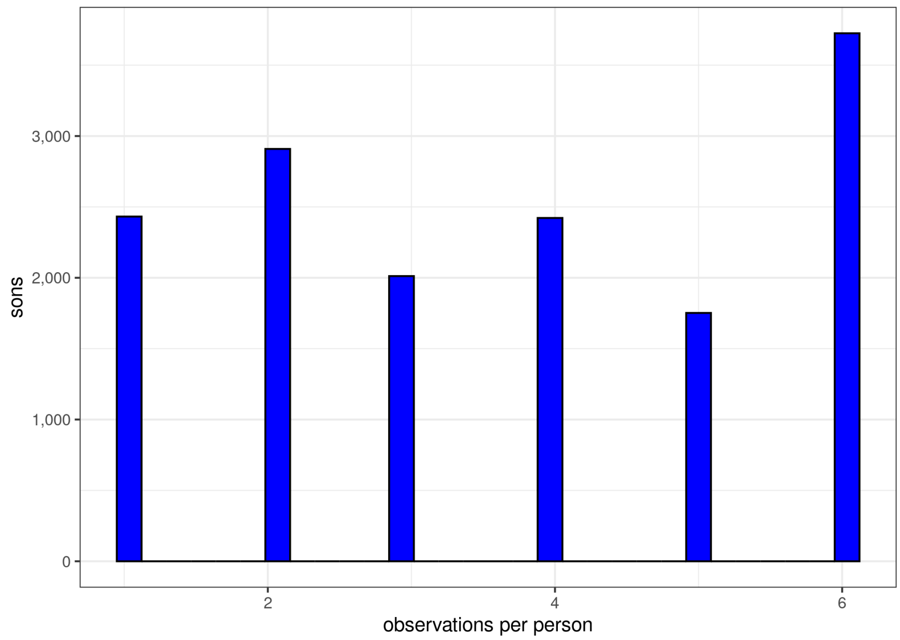
Table A.1: Summary Statistics of Key Samples

	Full	**Linked (0.8)**	**Linked (1)**	**Families**
income	3,315	3,274	3,500	4,049
wealth	6,173	6,239	6,291	3,948
urban	0.23	0.21	0.21	0.15
male	0.83	0.80	0.84	1.00
N	24,454,033	7,299,628	5,088,246	15,252

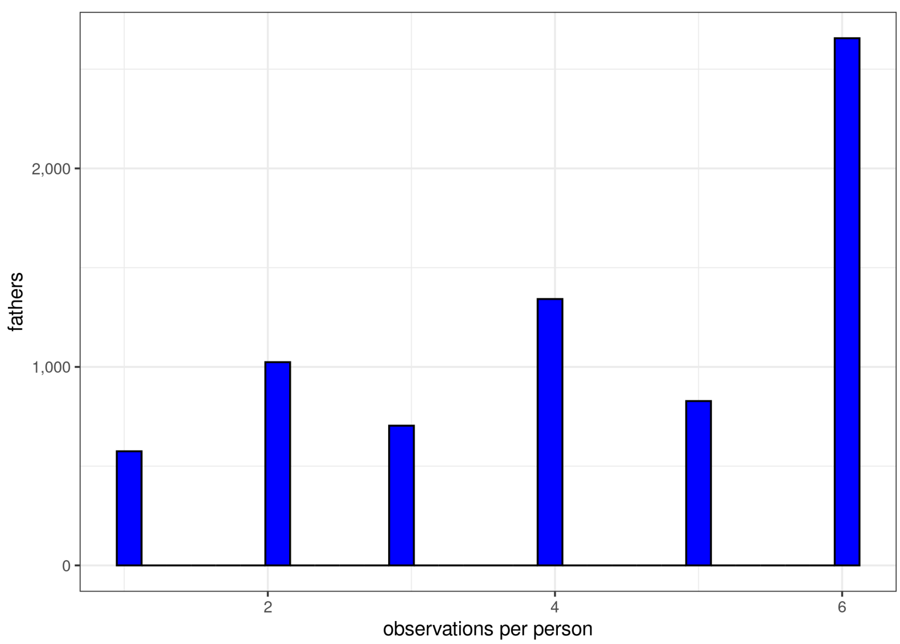
Note: This table shows descriptive statistics for the historical tax data (full), the sample linked to administrative data with a Jaro-Winkler string requirement of 0.8 and 1, and the linked sample with average incomes for sons and fathers (families). Numbers are averages for the different samples. Averages for income and wealth are calculated using inflation-adjusted numbers.

Figure A.6: Observation Per Person in the Family Sample

(a) Sons

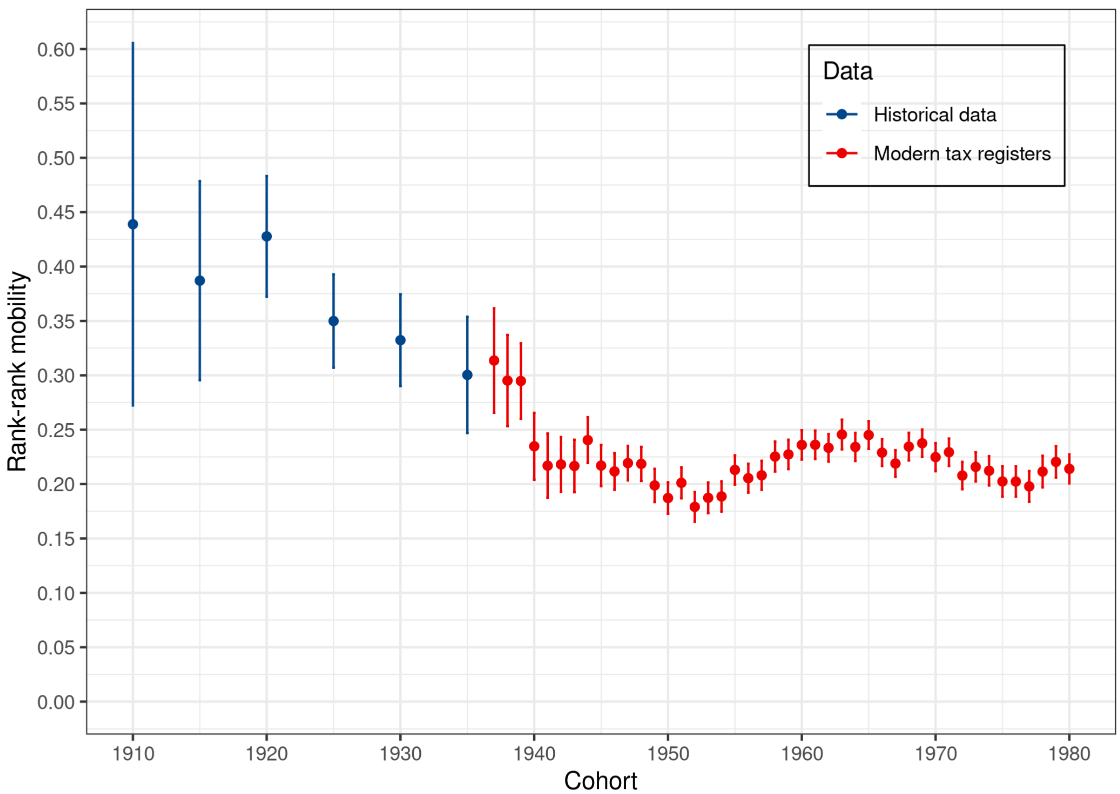


(b) Dads



Note: The Figures show the number of observations for each father and son in the family sample. Sons are measured at age 30-35, and fathers are measured at age 55-60.

Figure A.7: Relative Intergenerational Mobility in Income



Note: The figure shows long-run trends in relative intergenerational mobility in income between fathers and sons. Income is pre-tax and pre-transfers. The x-axis denotes the birth cohort of the child. See 3 for more information on empirical setup and data.

Table A.2: Re-weighting Rank-Rank Measures

	1910	1915	1920	1925	1930	1935	Full sample
No weights:							
	0.439	0.387	0.428	0.350	0.332	0.300	0.352
	(0.085)	(0.047)	(0.028)	(0.022)	(0.022)	(0.027)	(0.012)
Weights:							
by occupation	0.446	0.455	0.482	0.337	0.297	0.274	0.342
	(0.086)	(0.043)	(0.028)	(0.022)	(0.022)	(0.027)	(0.012)
by municipality	0.569	0.550	0.496	0.391	0.357	0.321	0.386
	(0.090)	(0.046)	(0.029)	(0.022)	(0.022)	(0.026)	(0.012)
by county	0.435	0.479	0.471	0.371	0.346	0.313	0.375
	(0.078)	(0.047)	(0.029)	(0.022)	(0.021)	(0.027)	(0.012)
by urban status	0.419	0.486	0.511	0.402	0.345	0.322	0.392
	(0.090)	(0.047)	(0.027)	(0.022)	(0.021)	(0.027)	(0.012)
by mun. and urban status	0.561	0.653	0.520	0.412	0.355	0.313	0.399
	(0.119)	(0.052)	(0.031)	(0.023)	(0.022)	(0.027)	(0.012)

Note: The Table presents rates of rank-rank persistence for five-year bins of the son's birth cohorts and the full sample with a selection of weighting schemes. Standard errors are in parenthesis.

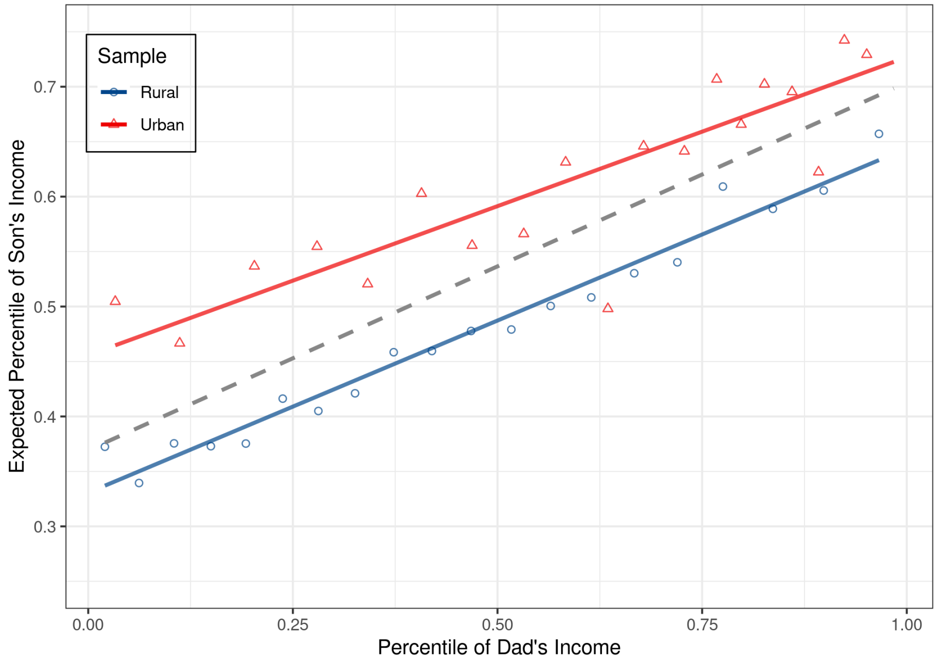
Table A.3: Sample Splits

	1910	1915	1920	1925	1930	1935	Full sample
rural	0.450	0.331	0.382	0.318	0.316	0.286	0.326
	(0.088)	(0.049)	(0.030)	(0.023)	(0.023)	(0.029)	(0.012)
urban	-0.255	0.719	0.370	0.256	0.185	0.416	0.303
	(1.080)	(0.303)	(0.117)	(0.096)	(0.072)	(0.087)	(0.044)
primary school	0.447	0.413	0.441	0.336	0.323	0.267	0.344
	(0.104)	(0.053)	(0.033)	(0.026)	(0.026)	(0.034)	(0.014)
more schooling	0.293	0.239	0.364	0.312	0.339	0.301	0.322
	(0.144)	(0.115)	(0.063)	(0.048)	(0.044)	(0.051)	(0.024)
no wealth	0.245	0.298	0.448	0.321	0.257	0.389	0.328
	(0.147)	(0.137)	(0.085)	(0.061)	(0.055)	(0.074)	(0.031)
some wealth	0.483	0.399	0.425	0.357	0.339	0.285	0.353
	(0.102)	(0.049)	(0.030)	(0.023)	(0.023)	(0.029)	(0.012)

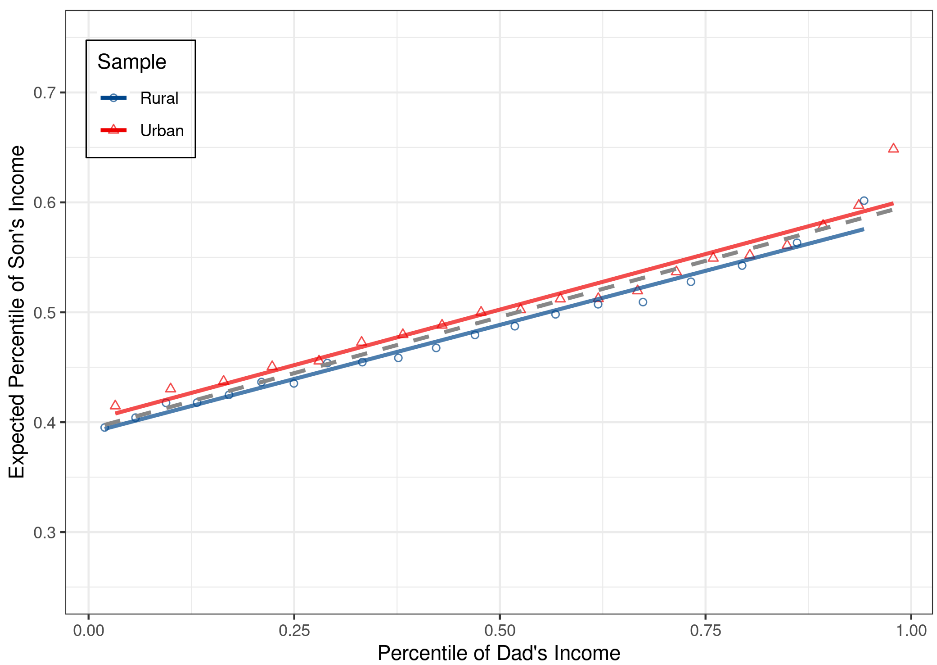
Note: The Table presents rates of rank-rank persistence for five-year bins of the son's birth cohorts and the full sample with a selection of sample selections. Standard errors are in parenthesis.

Figure A.8: Rank-rank Persistence By Rural Status

(a) Cohorts 1910-1935



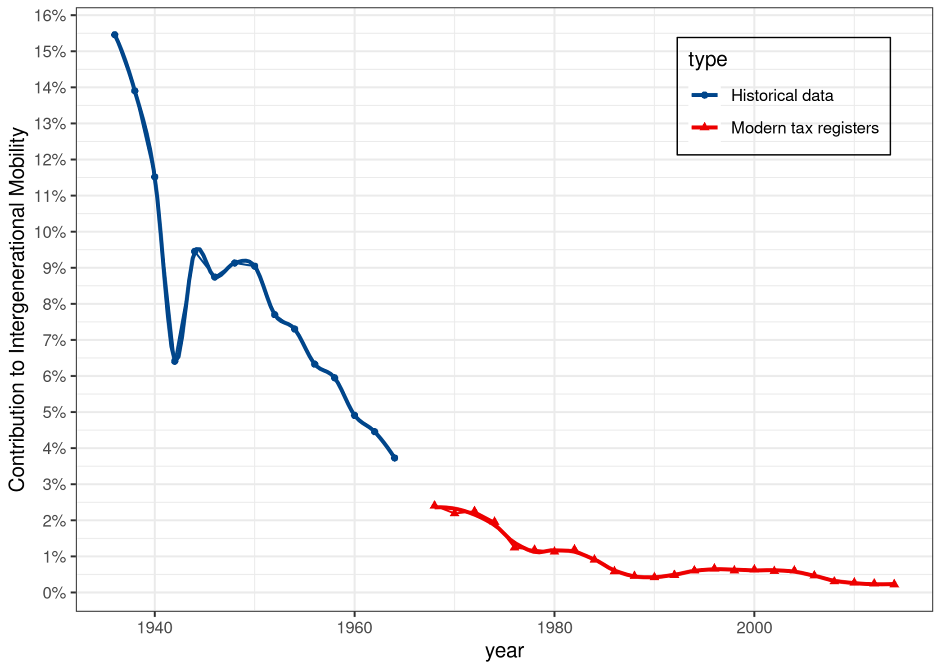
(b) Cohorts 1937-1964



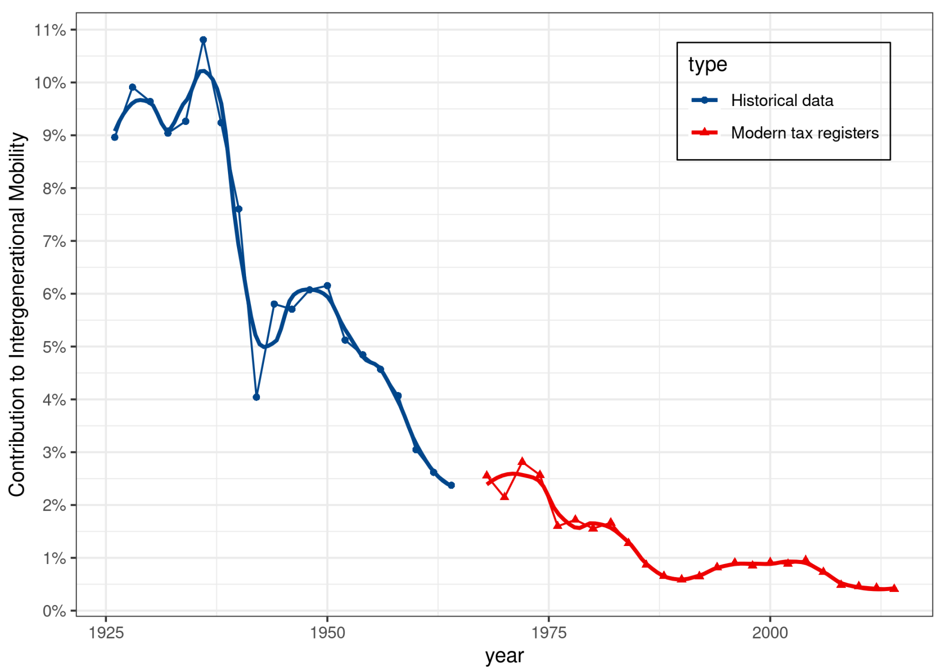
Note: The figures show average income ranks for sons by their father's place in the income distribution by father's place of residence. The two figures for cohorts born 1910-1935 and 1937-1964 represent the historical and modern samples, respectively. The grey dashed line is the join regression line, and its slope reflects the join intergenerational mobility.

Figure A.9: Contribution of Level Differences to Rank-Rank Persistence

(a) 1935-2014

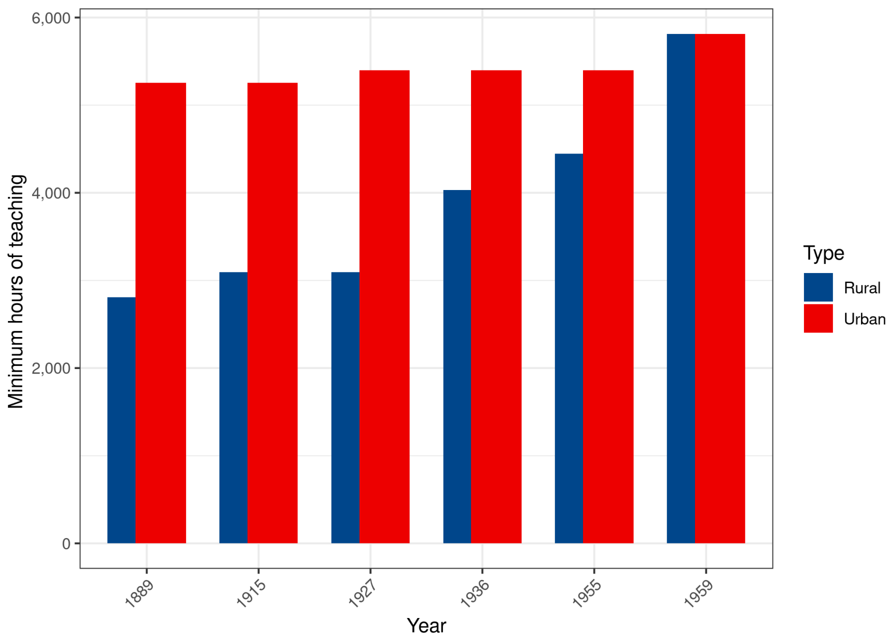


(b) 1925-2014



Note: The figure shows the contribution of the level difference in expected income rank between rural and urban areas to rank-rank persistence across generations based on a decomposition proposed by Jácome, Kuziemko and Naidu (2021). Samples are chosen to represent cover a geographical area over time. See text for more information.

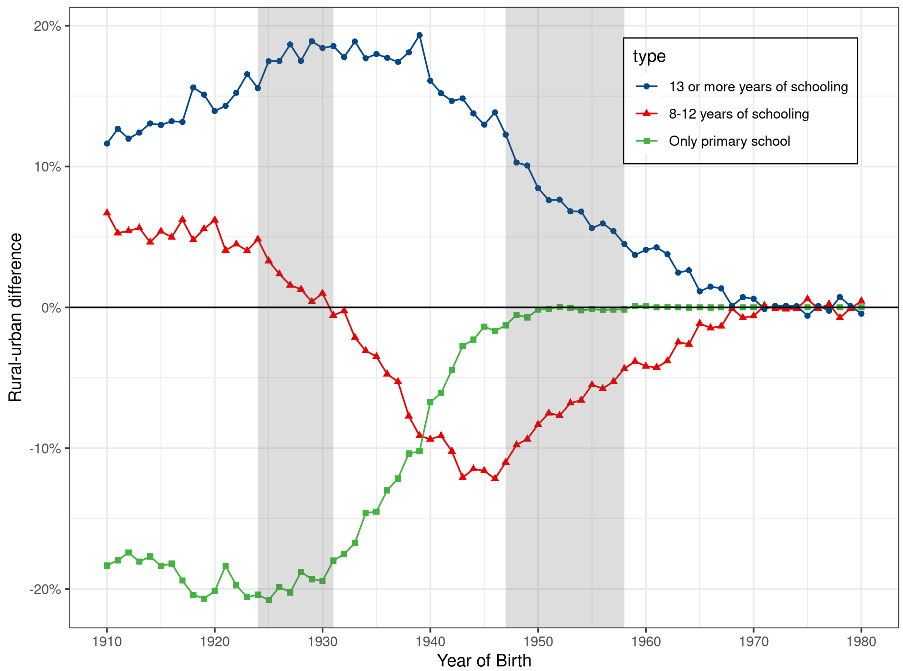
Figure A.10: Minimum Hours of Teaching During Primary School



Note: The Figures show the minimum allowed hours of schooling during primary school for rural and urban areas between 1889 and 1959. *Source*: Acemoglu et al. (2021).

Figure A.11: Rural-Urban Educational Divide

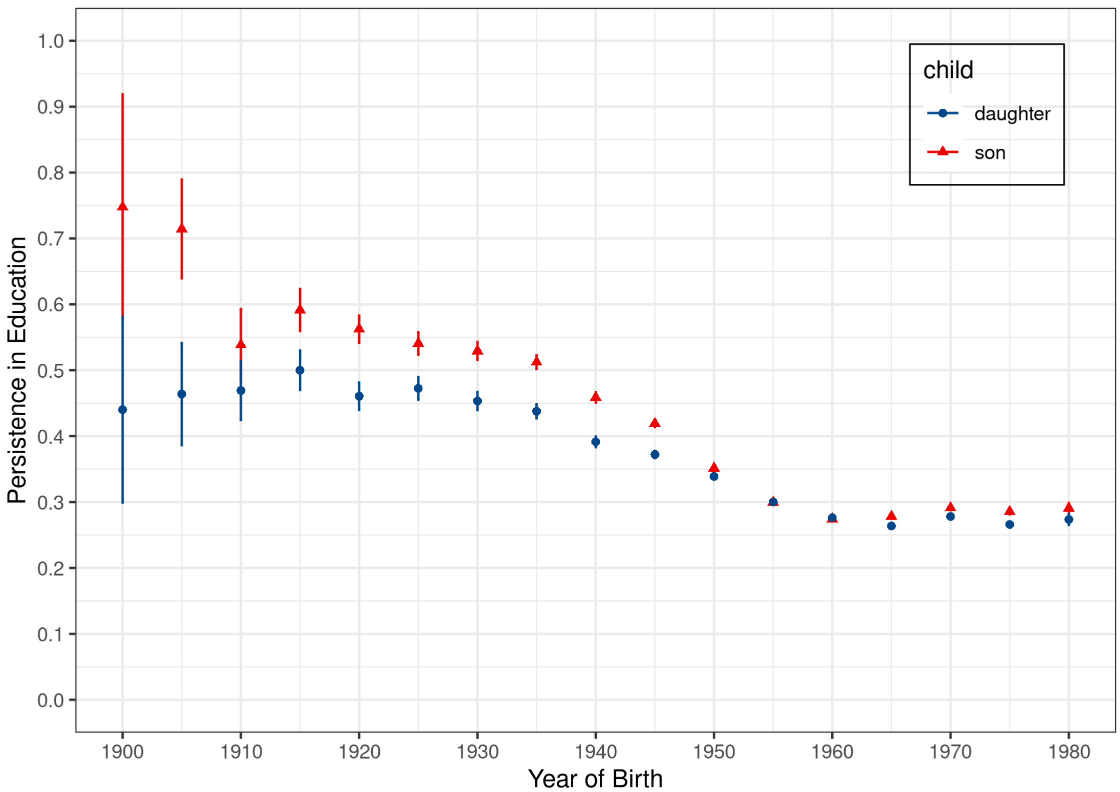
(a) Percentage points difference between rural and urban



(b) Mean difference in years of education

Note: The Figures show the gap between urban and rural areas in the percentage point with different educational attainment and the average difference in years of schooling. Individuals are characterized based on the status of their municipality of birth. Shaded areas are cohorts affected by the 1936 primary school reform and the change in compulsory schooling from 7 to 9 years in the 1960s.

Figure A.12: Relative Intergenerational Mobility in Years of Education



Note: The figure shows intergenerational mobility in years of education between fathers and children (sons and daughters) from estimating Equation 3. *Source*: Norwegian register data and own calculations.

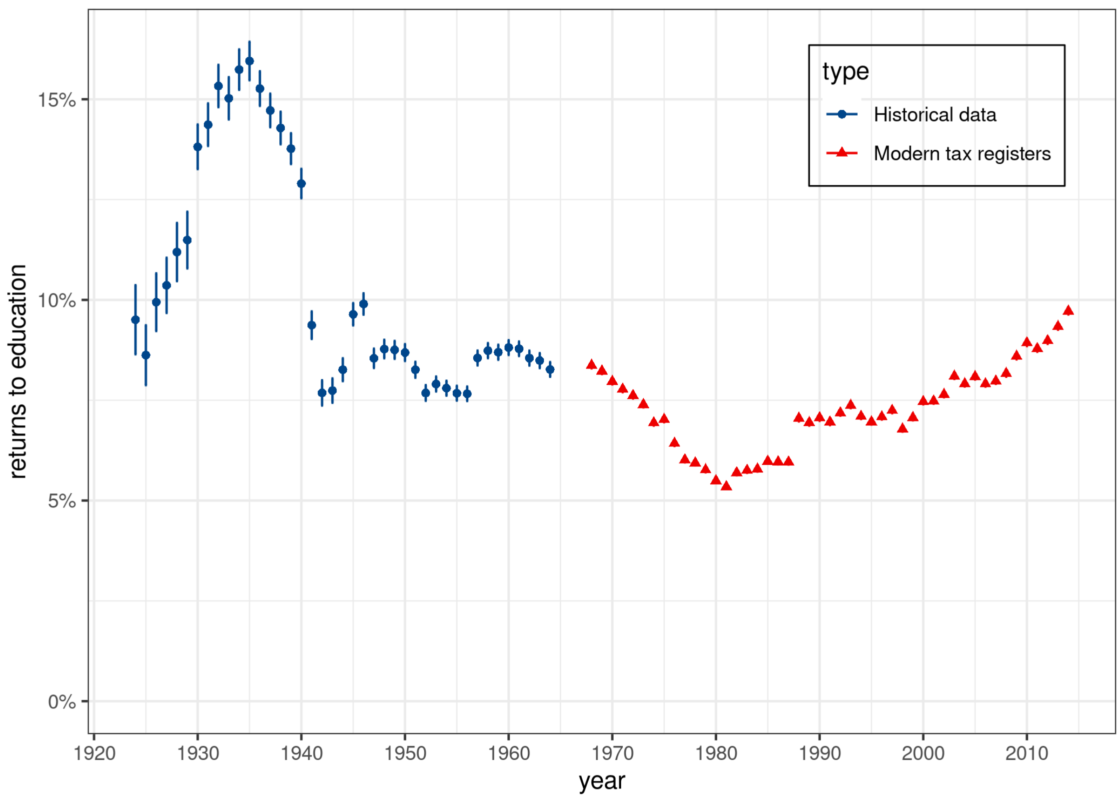
Table A.4: Causal Effect of 1936 Primary School Reform

	(1) Education child	(2) Education child	(3) Income rank child	(4) Income rank child
Reform	0.589***		0.077***	
	(0.121)		(0.012)	
Edu. dad		0.462***		
		(0.007)		
Edu. dad × reform		0.009		
		(0.015)		
Income dad				0.214***
				(0.013)
Income dad × reform				−0.053*
				(0.023)
Num.Obs.	54 519	54 067	54 552	28 246
FE: Cohort	X	X	X	X
FE: Birthplace	X	X	X	X

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

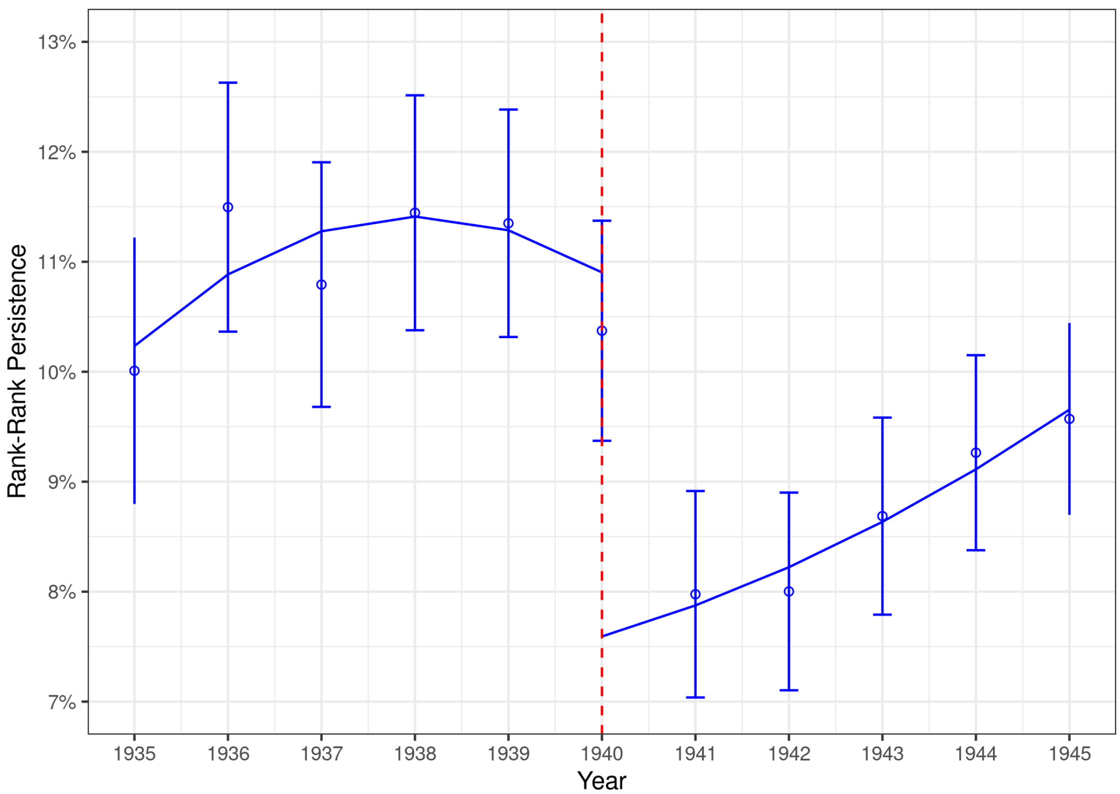
Note: The Table shows results from estimating Equation 6 and Equation 7 and years of education and income ranks. Data includes males born between 1910 and 1950, and only the dad's education and income are used. Significance levels: + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Figure A.13: The Returns to Years of Education



Note: The figure shows the estimated returns to a year of additional education from a standard Mincer equation as described in the text (see Equation 8). The sample includes data from the historical and modern tax registers and includes males only.

Figure A.14: Effect of a Reduction in the Returns to Education on Mobility



Note: The figure shows rank-rank persistence in income across generations for dads' incomes measured at different years. Trend lines are polynomials and fitted separately before and after the shock in 1940, and 1940 is excluded from the estimation as the shock partly impacted this year.

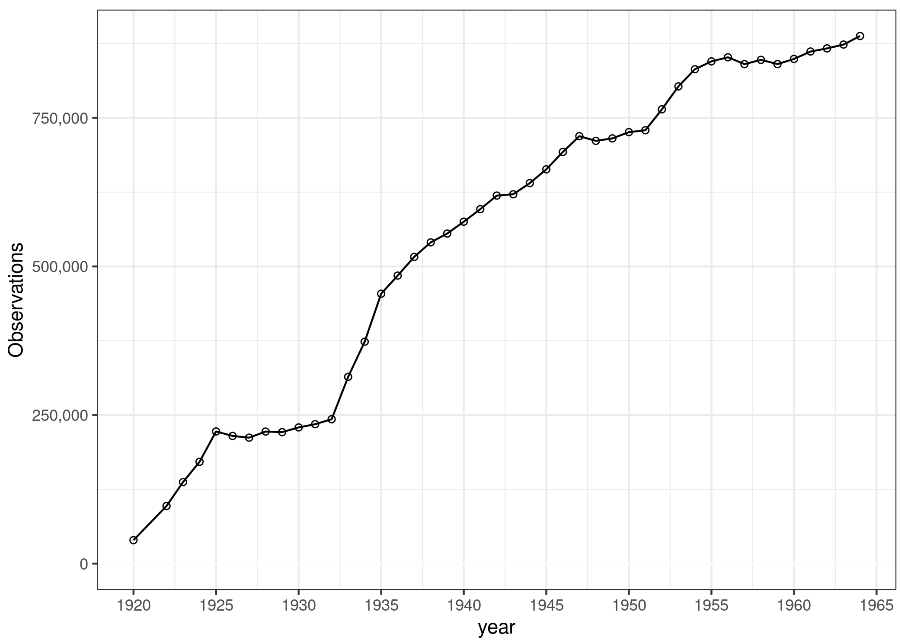
B Descriptive Statistics for Individual-Level Data

Table B.1: Descriptive Statistics

	1890	1900	1910	1920	1930	1940	1950	1960	Full sample
Missing values:									
First name	0%	2%	3%	0%	1%	0%	0%	0%	0%
Occupation	15%	35%	14%	5%	4%	4%	10%	42%	16%
Residence	77%	85%	51%	68%	27%	23%	10%	1%	17%
Municipality	100%	100%	74%	0%	0%	0%	0%	0%	2%
Gender	12%	22%	22%	8%	5%	4%	2%	2%	3%
Unique names:									
Full name	22,728	144,167	65,643	311,224	522,670	673,894	862,642	708,386	2,242,872
First name	14,745	82,406	41,631	180,098	122,632	159,567	213,245	177,731	672,359
Middle name	5,227	21,408	9,134	43,958	14,754	20,137	27,686	23,408	109,252
Last name	3,094	31,496	11,249	41,387	77,106	88,979	99,401	78,931	200,509
Unique values:									
Occupations	1,961	10,472	9,651	12,803	17,771	23,082	32,239	22,455	95,732
Address	2,877	16,246	37,723	51,870	168,353	253,983	449,764	476,631	1,089,511
Municipality	1	1	66	440	643	712	741	681	745
County	1	1	2	9	17	20	20	21	21
Books	2	25	5	31	66	70	94	68	361
Value is zero:									
Income	1%	0%	0%	1%	1%	1%	1%	1%	1%
Wealth	44%	46%	58%	40%	52%	52%	53%	42%	48%
N	33,642	289,431	131,513	745,384	2,146,802	3,320,663	5,244,074	3,990,862	15,902,371

Note: Descriptive statistics from our full dataset. The first section presents the fraction of each variable that is missing. A first name is regarded as missing if it contains only one letter. The second and third sections give the number of unique observations for a given variable. The last section gives the fraction of observation with income and wealth equal to zero. Whenever only one number is recorded in the source, we assume the wealth equals zero. The final column provides these statistics for the full sample. See Section 3 for more details.

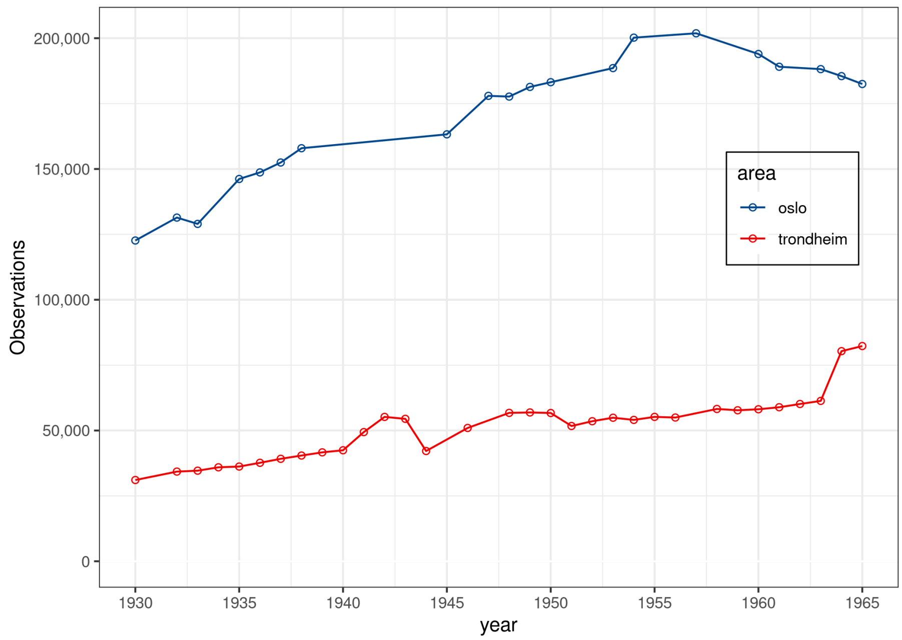
Figure B.1: Yearly Observations From Tax Registers



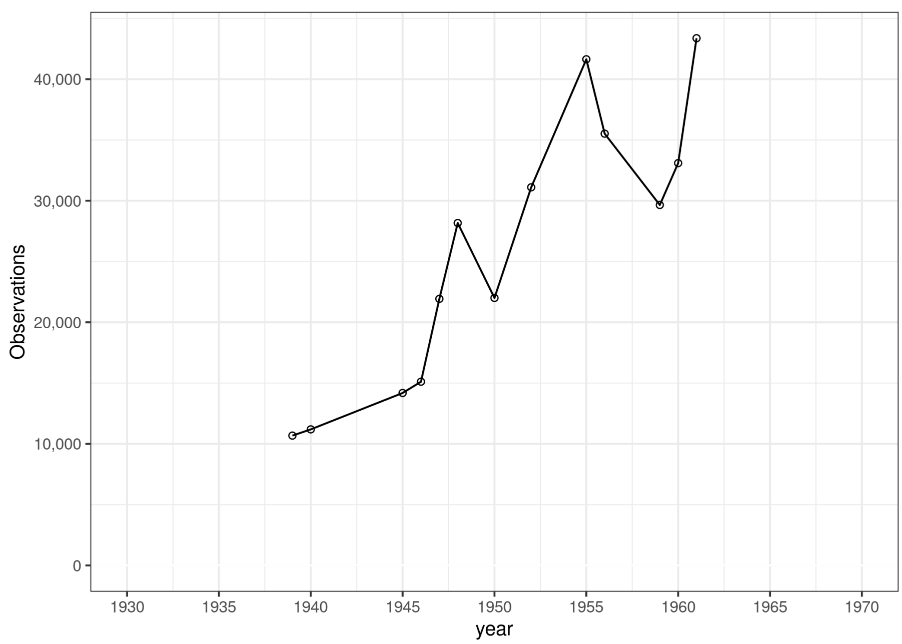
Note: The figures show the number of observations digitized from 'tax record based' address books, meaning address books with information from tax records.

Figure B.2: Observations in Additional Datasets

(a) 'Normal' Address Books



(b) Top Taxpayers Oslo

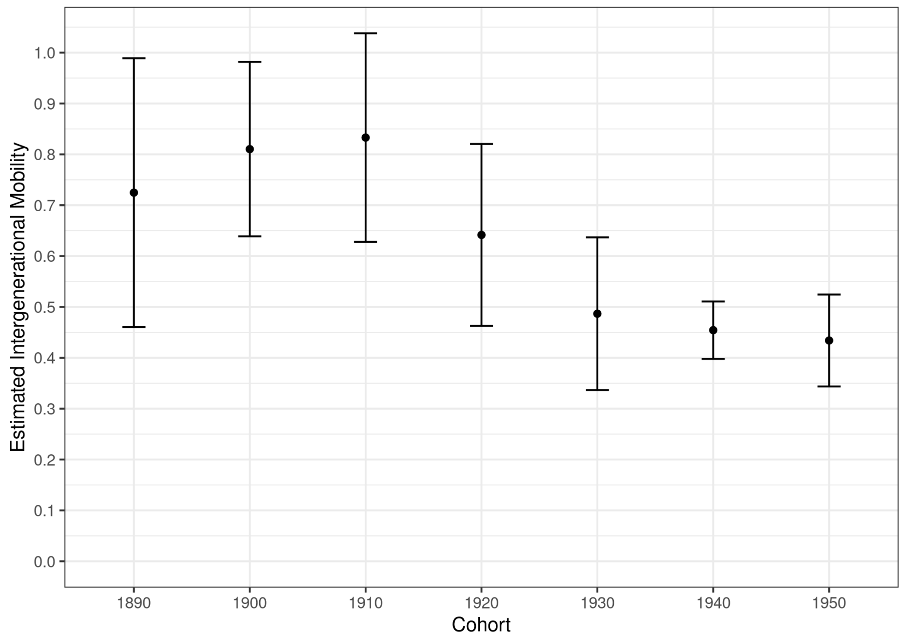


Note: The figures show the number of observations digitized first from 'normal' address books, meaning address books without information from tax records, and then from the book series on top taxpayers in Oslo and the surrounding areas.

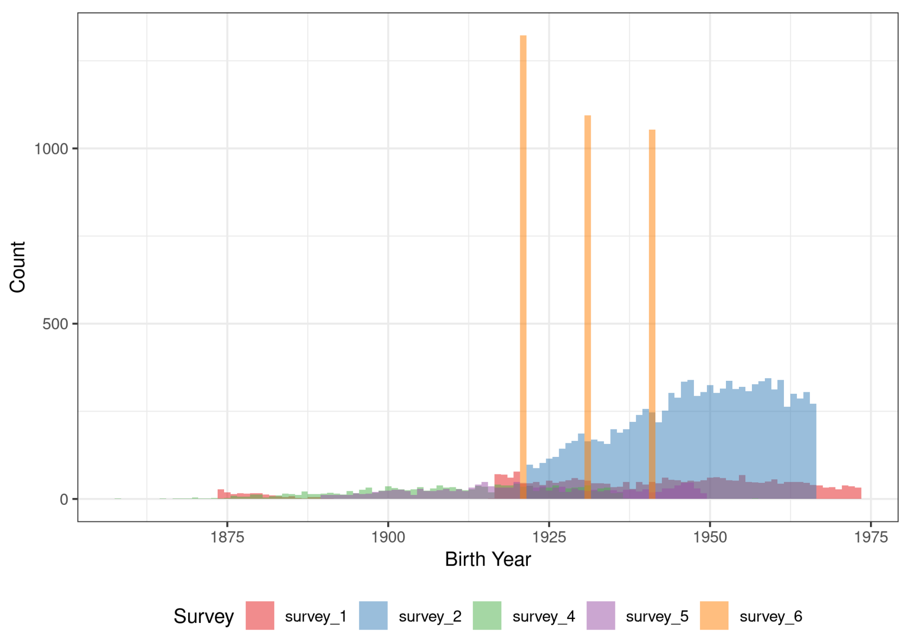
C Survey Based Estimates of Intergenerational Mobility in Education

Figure C.1: Replications of Mjølsnes (2019)

(a) Survey Based Measures of Intergenerational Mobility



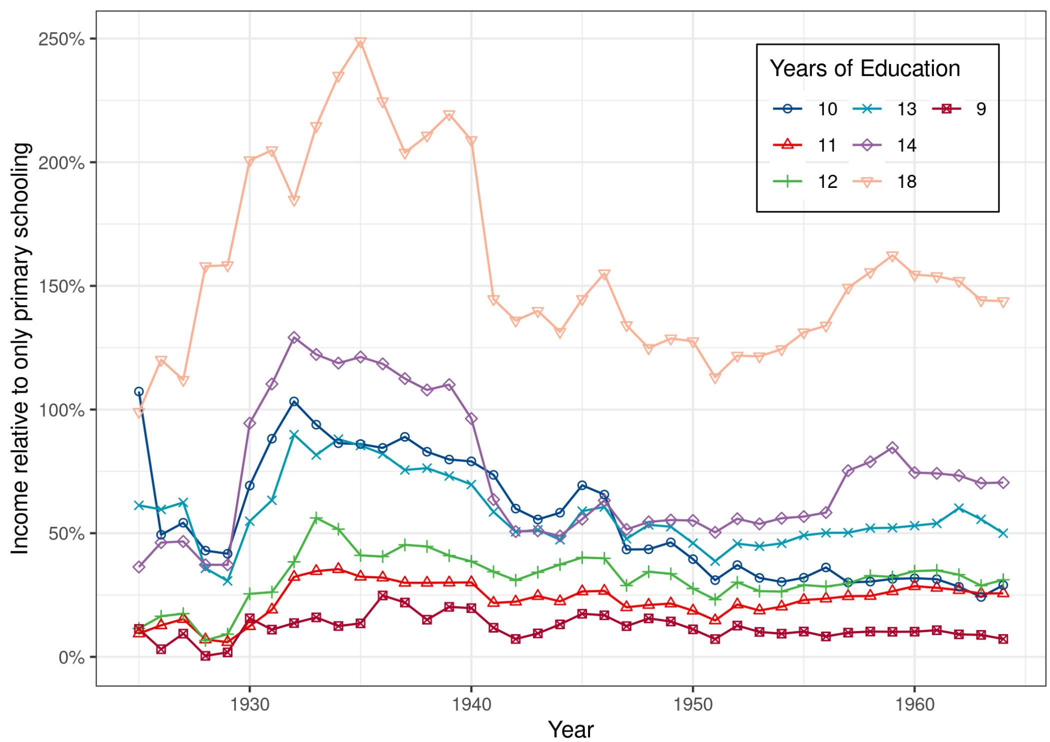
(b) Histogram of Surveys



Note: The Figures show rates of intergenerational mobility in years of education and a histogram of observations from six separate surveys. The sample includes dads and sons only. Estimates replications of Mjølsnes (2019) done by the author.

D Additional Figures

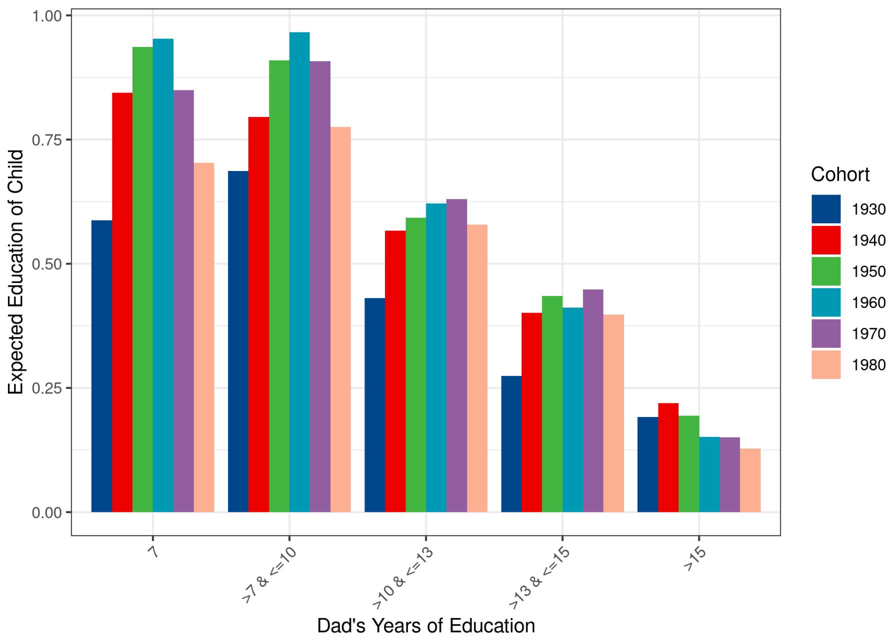
Figure D.1: Income Relative to Primary School



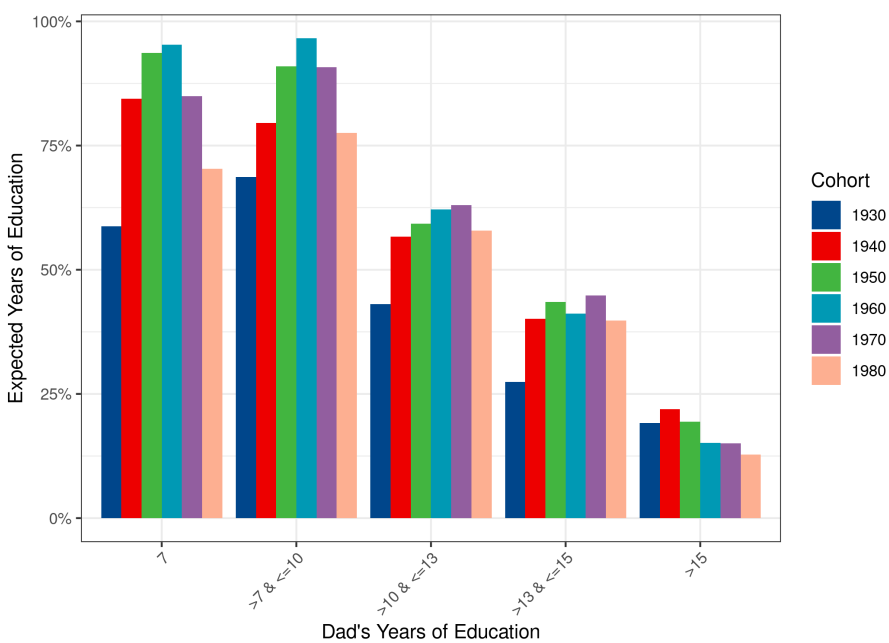
Note: The Figure shows average incomes for different completed years of education relative to those with only primary school. The sample males only.

Figure D.2: Absolute Mobility in Years of Education

(a) Expected Child Education by Dads Education by Decade



(b) Fraction with Higher Education than Father by Decade



Note: The Figures show average years of education and absolute mobility of sons based on fathers' educational attainment.