RESEARCH INSTITUTE OF INDUSTRIAL ECONOMICS

IFN Working Paper No. 1516, 2025

Cities and the Rise of Working Women

Thor Berger, Mounir Karadja and Erik Prawitz

Research Institute of Industrial Economics
P.O. Box 55665
SE-102 15 Stockholm, Sweden
info@ifn.se
www.ifn.se

Cities and the Rise of Working Women*

Thor Berger† Mounir Karadja‡ Erik Prawitz§

January 8, 2025

Abstract

We document that large cities were instrumental in shaping women's work and family outcomes in the early 20^{th} century. We focus on migrants to Stockholm, Sweden's largest city, using representative, linked census data. Female migrants to Stockholm saw persistent changes in work and family outcomes over the life-cycle. Migrants were approximately 50 percentage points more likely to enter the labor force and less likely to marry or have children than their sisters migrating to rural areas. They experienced skill-upgrading and higher real incomes, without adverse mortality effects. Early structural shifts towards services partly explain these patterns.

*We thank Martin Önnerfors for his contributions to the development of this project, as well as Martín Fernández Sánchez, Katrine Løken, Raoul van Maarseveen, JF Maystadt, Zach Ward, and several conference and seminar participants for helpful discussions and suggestions. Funding from Jan Wallanders och Tom Hedelius stiftelse and Riksbankens Jubileumsfond is gratefully acknowledged.

†Swedish Collegium for Advanced Study (SCAS), Uppsala University; Department of Economic History, School of Economics and Management, Lund University; CEPR; and Research Institute of Industrial Economics. E-mail: thor.berger@ekh.lu.se

‡Department of Economics, Uppsala University. E-mail: mounir.karadja@nek.uu.se

§Department of Economics and Statistics, Linnaeus University; and Research Institute of Industrial Economics. E-mail: erik.prawitz@lnu.se

1 Introduction

Women's entry into the labor force is one of the most fundamental economic and social changes of the 20^{th} century. An influential body of work documents female labor force participation (FLFP) trends across countries and examines the role of culture, institutions, or technology in accounting for the transition (Goldin, 1990, 1995; Costa, 2000; Olivetti, 2013). However, aggregate trends may conceal spatial variation in women's economic opportunity both historically and today (Fogli and Veldkamp, 2011; Chetty et al., 2014). Consider Figure 1, which plots FLFP by regional population rank using census data from early-20^{th} century Britain, Sweden, and United States.[1] In each country, the relationship between FLFP and population is essentially flat for most of the population distribution. Yet, all countries also exhibit a discontinuity such that FLFP is considerably higher in the most populous areas, suggesting that the very largest cities provided substantial labor-market opportunities for American and European women already in the early 20^{th} century.

The importance of cities in shaping women's economic opportunities during industrialization has long been recognized in historical and qualitative research.[2] Yet, there has been little systematic and quantitative analysis of such patterns using large-scale data. This paper aims to bridge this gap by rigorously analyzing the role of large cities in women's economic and family outcomes in the early 20^{th} century. To do so, we turn to studying urban migrants. Studying how migration affected women in a historical setting is particularly challenging, as they can seldom be linked over time in census data. Therefore, we focus on the case of Sweden, where censuses recorded women's maiden names consistently even after marriage. Using Swedish censuses between 1880 and 1910, we create a large and representative linked sample of women (and men) using automated record-linkage methods, which allows us to track women from their childhood households to their eventual migration to other rural and urban locations.

We first show that women migrating to the largest cities of Sweden experience substantially higher levels of labor force participation compared to non-migrants. In sharp contrast, female migrants to destinations outside the top decile of the population distribution show no significant changes in their employment status, whereas male migrants experience similarly high returns to migration, regardless of their destination.

Building on these results, our main analysis focuses on the largest city – Stockholm – around the turn of the 20^{th} century. Stockholm is an emblematic example of an emerging metropolis in a rapidly industrializing country: between 1860 and 1910, the city tripled in size as Sweden

[1] In the following, we use FLFP to denote market-oriented full-time work, which is better measured for women in historical censuses compared to a broader definition including, for example, unpaid labor conducted on family farms, which tends to be underreported for women in census data (Goldin, 1990; Stanfors, 2014). We address well-known issues related to the underenumeration of informal female work in historical census data below.

[2] See for example Tilly and Scott (1978), Fuchs and Moch (1990), Hill (1994), Engel (1994), Deutsch (2000), and Pooley and Turnbull (2005).

was transformed from a backward agricultural economy to a modern industrial and urban nation. Throughout this period, more than half of Stockholm's population consisted of migrants, who were disproportionately female. To estimate the city's role in shaping female work, our main empirical strategy consists of comparing female migrants to Stockholm with their sisters migrating to other destinations. This strategy has two main benefits. First, it allows us to control for all characteristics that are shared between siblings and that may influence both migration decisions and labor-market outcomes, such as parental socio-economic status and social networks. We also conduct a battery of robustness tests and extensions to investigate potential individual-level confounders. Second, only comparing *migrant* sisters enables us to tease out the role of the largest cities rather than migration itself.

Following this approach, we find that women who migrate to Stockholm are about 50 percentage points more likely to be in the labor force compared to their sisters migrating to a rural area. Migrants to urban areas other than Stockholm are also more likely to enter the labor force, but we estimate an almost four times larger difference among those moving to Stockholm. A larger effect on FLFP from moving to the largest Swedish city is consistent with the pattern in Figure 1 showing that FLFP in Europe and the United States was especially high at the top of the population distribution. As a test the for generalizability of this finding, we make use of recently generated links that uses genealogical information and machine learning algorithms to track individuals over time in US census data (Buckles et al., 2023). Using data from the 1880 and 1910 US censuses, we find substantial increases in FLFP among American women moving to the largest cities compared to their sisters who migrate elsewhere. The fact that results are similar in Sweden and the United States—two countries that differed considerably in terms of culture, institutions, and urbanisation—suggests that our main results are informative also for other historical contexts.

Women's labor-force participation was closely intertwined with marriage and fertility decisions at the turn of the 20^{th} century. In particular, women typically exited the labor force upon marriage. We therefore next investigate the implications of cities for family formation. Our estimates show that the large increases in labor-force participation among female migrants to Stockholm is mirrored almost one-to-one by sharp reductions in both marriage and fertility rates. In contrast, migrants to other destinations (that saw no effects on FLFP) were more likely to marry and have children than non-migrants.[3] Moreover, while we find a substantial *relative* increase in employment for Stockholm migrants who were married in 1910, the absolute effect is small in magnitude compared to singles. Taken together, the differential patterns in FLFP across space that we uncover are thus closely related to the decision to remain unmarried.

Our empirical strategy allows us to rule out selection based on family-level characteristics, but cannot fully rule out the presence of within-family selection into migrant destinations. However,

[3] Similar patterns are also replicated in the sample of US migrants.

several complementary approaches all indicate that such selection has a limited impact on our estimates. First, we show that estimates are very similar when excluding sibling fixed effects and individual controls, thus suggesting limited selection overall. Second, for the subset of migrants that move later in life, we are able to hold constant individual-level pre-migration outcomes in terms of employment, marriage, and childbearing that may explain both selection into migration and FLFP. These controls have minor impacts on our estimates. Third, applying the method of Oster (2019), we find evidence of somewhat *negative* selection of migrants within sibling groups in terms of labor market participation. Accounting for within-siblings individual variation therefore tends to magnify estimates, suggesting that our results are, if anything, biased towards zero.

An important question is to what extent the timing of migration precedes the choice to enter the labor force or form a family. For example, if women who do not marry choose to migrate to an urban area for work, cities would not ultimately be the drivers of our findings. Using the panel structure of our data, we perform several exercises to test for timing. First, to control for pre-migration selection in terms of marriage, we restrict attention to siblings that were all unmarried in 1900, ten years prior to our main outcomes. In line with our findings when holding pre-migration outcomes constant, we find very similar results. Second, we restrict attention to individuals who had already migrated in 1900 *and* were observed as unmarried, such that we compare sisters who made the decision to migrate prior to marriage. Even within this subsample, we estimate large and consistent differences in all our main outcomes for Stockholm migrants. Together, these tests indicate that our estimates are driven by changes occurring after, rather than before, migration. While we do not claim that our estimates capture an unbiased causal effect of urban migration – indeed we believe they are important also in a descriptive sense – the large magnitudes and limited evidence of family- or individual-level selection suggests that large cities did have an impact on women's early increases in market-oriented work.

We next explore the dynamics of urban migration. First, using a sample of early migrants, we test for persistence across census rounds. For those who migrate to Stockholm by 1900, we find that employment, marriage, and childbearing outcomes exhibit a large degree of persistence ten years later. However, these long-run estimates are approximately 30 percent smaller, indicating that a fraction of women marry and leave the labor force over time. Second, we investigate how migrants' outcomes unfold over the life cycle. Exploiting the age-differences within the sample and the fact that migrants are between 20 and 46 years old when we observe them (either in 1900 or 1910), we show that Stockholm migrants have substantially different life-cycle patterns of employment and family formation compared to rural migrants. In particular, while rural migrants have relatively high FLFP in their early 20s, this quickly subsides from above 50 percent to less than 20 percent by age 30. By contrast, Stockholm migrants display persistently high rates of FLFP throughout their 30s at approximately 80 percent, with only moderate decreases occurring after the age of 35. In their mid-forties, Stockholm migrants are still about 30 percentage points more likely to participate in the labor force, and correspondingly less likely to be married or have children. Moreover, their likelihood of working in higher skilled occupations rises consistently with age, whereas there is no such pattern for other migrants (Glaeser and Maré, 2001). Our results indicate that migration to Stockholm thus had persistent impacts on work and family life.

Why did large cities provide economic opportunity for women? A central explanation for the aggregate rise in FLFP during the 20^{th} century is the sectoral reallocation of economies towards the service sector, which provided "respectable" jobs for women (Goldin, 1995; Olivetti, 2013). We document an overlooked spatial dimension to this shift: large cities in Europe and the United States had already experienced a shift towards the service sector by the early-20^{th} century, which would take other regions more than half a century to reach. As a case in point, the share employed in services in 1900 Stockholm would be matched by other Swedish regions only in the late 1960s. Indeed, while the manufacturing sector was initially considered the main pull-factor for migration to Stockholm, it was replaced by the fast-growing service sector in the decades after the turn of the century (Johansson and Persson, 2004).

In line with existing theories and the early shift towards the service sector, women moving to Stockholm were disproportionately more likely to transition into services jobs than industry. Comparing sisters migrating to more or less service-based economies, we also find a strong positive association between service sector size at destination and FLFP. Female migrants were more likely to take up service jobs across the skill spectrum, with the bulk of migrants being concentrated in relatively low-skilled service jobs. Lastly, the increase in FLFP in Stockholm does not seem to have come at the price of lower income and unhealthy working conditions in the city. Measuring economic returns using occupational income scores reveals that female migrants to Stockholm saw substantial increases in real incomes and intergenerational mobility. Moreover, female migrants to Stockholm lived as long as their sisters migrating elsewhere, which sharply contrasts the case of male migrants to Stockholm who experienced a steep health penalty.

Our paper provides new evidence of large spatial heterogeneities in FLFP within Europe and the United States in the early 20^{th} century. Although industrialization is often associated with an aggregate decline in FLFP, we show that large cities during industrialization provided unique opportunities for female employment. In that sense, our paper contributes to an influential literature that studies aggregate trends in FLFP, documenting that it follows a U-shape over the course of economic development (Boserup, 1970; Goldin, 1990, 1995; Olivetti, 2014).[4] Prior work on FLFP during industrialization has emphasized the role of manufacturing in cities and towns in providing employment (Tilly and Scott, 1978; Goldin, 1980; Goldin and Sokoloff, 1982; Kim, 2005). We

[4]While this literature often focuses on married women, Mammen and Paxson (2000) find the U-shape in panel data on adult women regardless of their marital status.

contribute by documenting an important role of service sector jobs in the largest cities. Thus, our results are consistent with evidence that the development of the service sector, broadly defined, is a key determinant of female labor supply over the past century (Goldin, 1990, 1995; Lee and Wolpin, 2006; Akbulut, 2011; Ngai and Petrongolo, 2017; Bridgman et al., 2018; Buera et al., 2019; Cerina et al., 2021; Ngai et al., 2022). Another stream of work documents an important role for social and cultural influences on FLFP (Fernández et al., 2004; Fernández, 2013; Olivetti et al., 2020). While the development of the service sector is a key explanation behind the increase in FLFP in our data, a significant part of this increase remains unexplained, possibly due to differential cultural and social norms in the largest cities.

With our focus on internal migrants, we also provide novel evidence on the historical returns to migration for women using established automated record-linkage methods (Ferrie, 1996; Abramitzky et al., 2021). Young women can typically not be linked over time in historical census data using standard techniques since most women changed their surname upon marriage. Therefore, women are typically completely excluded from linked samples using historical data such as the pioneering Census Linking Project (Abramitzky et al., 2020). Some recent papers have relied on marriage records (Craig et al., 2019; Withrow, 2021) or genealogical data (Feigenbaum and Gross, 2020) to follow women over time in census data.[5] Although such methods are feasible to study subsets of women, obtaining large and representative linked samples of women is typically not possible given data constraints. A recent contribution by Buckles et al. (2023) applies manually collected genealogical information as training data to link both women and men over time in US censuses, obtaining a larger and more representative sample of the U.S. population than in the previous literature. As noted above, we use their data to replicate our main analysis for the U.S. setting.

Lastly, we show that focusing on the returns for female migrants alters several stylized facts regarding the returns to migration. There is a vast literature studying male migrants (Long, 2005; Kennan and Walker, 2011; Abramitzky et al., 2012; Collins and Wanamaker, 2014; Ward, 2020), including the returns to rural-to-urban migration (Young, 2013; Bryan et al., 2014; Hamory et al., 2021). Perhaps most notably, while male migrants faced a trade-off between improved economic outcomes and worse health outcomes, female migrants saw improved economic and unchanged health outcomes in Stockholm. More fundamentally, we show that migration to the emerging metropolises of the early 20^{th} century shaped women's employment, marriage, and fertility decisions in line with the "independent female worker"-phase of the evolution towards modern female labor force participation (Goldin, 2006).

[5] One solution is devised by Olivetti and Paserman (2015) who creates pseudo links to include women in their study of intergenerational mobility, which is possible since their method does not require the actual linking of individuals over time.

2 Background and institutional setting

2.1 Cities, Services, and FLFP

Figure 1 documents the stylized fact that motivates our analysis: FLFP varied substantially within countries in the early 20^{th} century and was significantly higher in the most populous places in both Europe and the United States. To construct the figure, we use individual-level data on millions of women enumerated in the population censuses of England & Wales (1911), Sweden (1910), and the United States (1910).[6] Focusing on paid, market-oriented work, we calculate the FLFP rate among prime-aged (20–55) women by their place of residence. We then divide counties and parishes into deciles based on their population and plot the mean FLFP rate within each bin after absorbing country fixed effects. Strikingly, there is a discontinuous increase in FLFP between the 9^{th} and 10^{th} deciles in all three countries. Thus, FLFP rates in Europe and the United States were substantially higher in the most populous areas. For example, FLFP rates in the most populous parishes in Sweden were about two thirds higher than the average below the top decile.

What explains this variation in FLFP rates? An influential literature documents that FLFP follows a U-shape with respect to economic development at the country-level (Goldin, 1995; Olivetti, 2013). As a country industrializes, FLFP declines due to the separation of work from the home and the expansion of dirty and physically demanding industrial jobs. Conversely, the rising part of the U-shape is explained by the shift towards service jobs with less of a social stigma for female workers. Consequently, the labor-market entry of women in the aggregate has been seen typically as a post-World War II phenomenon when the tertiary sector expanded. However, we find that focusing on aggregate structural change conceals significant spatial variation.

A key factor that may account for the higher observed FLFP rates in big cities is that they experienced the shift toward services earlier than less populated areas. Panel B of Figure 1 shows that the largest places in Britain, Sweden, and the United States had a substantially higher fraction of service jobs than other regions in the early-20^{th} century.[7] Again, we plot the mean share of

[6]To estimate FLFP, we use individual-level data drawn from the 100% 1911 census for England and Wales, the full-count 1910 census for Sweden, and a 1% random sample from the 1910 U.S. census (IPUMS, 2020). We assign individuals to the smallest administrative geographical unit available in the IPUMS data, excluding overseas military installations. In England and Sweden, women are assigned to their (civil) parish of enumeration where the total parish population is calculated directly from each respective census. For the U.S., we limit the sample to white women that are assigned to their county of enumeration and add data on county populations from the 1910 census (Haines, 2010).

[7]To estimate the size of the service sector, we calculate the fraction of the employed population that works in a service occupation. For Britain and Sweden, the service sector is defined as workers reporting an occupation in major groups 0/1 (*Professional, technical and related workers*), 2 (*Administrative and managerial workers*), 3 (*Clerical and related workers*), 4 (*Sales workers*), and 5 (*Service workers*) in the Historical International Standard Classification of Occupations (HISCO) scheme. For the United States, service jobs are defined as those individuals reporting their primary occupation as belonging to the major occupational categories 1 (*Legislators, senior officials and managers*), 2 (*Professionals*), 3 (*Technicians and associate professionals*), 4 (*Clerks*), and 5 (*Service workers and shop and market sales*) in the International Standard Classification of Occupations (ISCO) scheme.

employment in services across deciles of county or parish populations after absorbing country fixed effects.

The early shift toward the service sector in more populous places is also evident across regions in Europe. Figure 2, Panel A, provides systematic evidence for 12 European countries showing that the most densely populated regions on average had about 40 percent of their employment in services already in 1900, which was attained by less densely populated areas first in the 1970s.[8]

2.2 Stockholm and urban migration

At the beginning of the 19^{th} century, Sweden was a predominantly agricultural economy, with 90 percent of the population living in rural areas in 1850. The industrialization that began in the 1860s led to rapidly growing incomes and labor demand from urban areas, drawing large numbers of workers from surrounding regions. Stockholm, the capital, grew especially rapidly and between 1860 and 1910, tripled in size from 110,000 to 333,000 inhabitants. The development of the railroad network, beginning in 1860, further facilitated migration to cities.

Stockholm's growth was fueled by industries such as textiles, shoe making, and printing, creating a strong demand for semi-skilled and unskilled labor. By the turn of the century, the city was also a hub for domestic service, with urban households employing a substantial share of young female migrants. More than half of Stockholm's population consisted of migrants, a majority of whom were women. As a consequence, the city of Stockholm was also majority female, with 122 women per 100 men in 1910, a pattern consistent with contemporary trends in high-income countries, where urban areas tend to have more women and rural areas experience a surplus of men (Edlund, 2005).

During this time, both men and women typically left their parental home at a young age to take up work and build up savings before forming a family. This practice, known as life-cycle service, was common across socio-economic classes, with migration of young individuals being the norm regardless of gender. Using individual-level data for two distinct regions of Sweden, Sundvall et al. (2023) find that the median age of leaving home ranged from 19 to 23 between 1880 and 1910, with little variation across regions or urban/rural areas. Marriage typically occurred a few years after moving out, with the median age at first marriage ranging from 25 to 28. Female migrants to rural areas typically entered employment as maids in farming households, bound by one-year contracts and compensated through a combination of cash and in-kind payments. In her influential investigative work, journalist Ester Blenda Nordström documented the grueling conditions faced by these agricultural workers, including 16-hour workdays and limited leisure time, amounting to a few hours one day each week (Nordström, 1914). Marriage brought little relief, as between

[8] Appendix Figure A.1 displays the evolution of service shares in each individual country showing that the most densely populated region has a larger service sector throughout the period in all European countries in our data.

taking care of the household and performing agricultural tasks, Carlsson (1966) describes rural wives as enduring the most strenuous conditions of all social groups at the time. In contrast, maids employed in urban domestic service experienced significantly better conditions, earning higher wages and working fewer hours, which increased the appeal of urban migration in the late 19^{th} century (Vikström, 2003). Also outside domestic services, female labor paid better in the bigger cities, partly due to the fact that they tended to reward tenure (Burnette and Stanfors, 2020). These patterns align with the income data presented below, which indicates that working women in Stockholm earned 22 percent higher real incomes than their rural counterparts working in the same occupations.

Stockholm provides a fitting illustration of the general pattern shown in Section 2.1, in which the largest cities develop service sectors during early stages of industrialization. Panel B of Figure 2 shows that Stockholm county (containing the capital) had achieved an employment share in services of about 40 percent by the early 20^{th} century, which was matched by other counties only in the 1960s. The early shift towards services is also mirrored in very high FLFP rates: in 1910, more than 50 percent of women aged 20–55 in Stockholm were part of the labor force, compared to 37 and 18 percent in other urban and rural areas, respectively (see Appendix Table A.1). Apart from domestic work, Stockholm women most commonly found work as seamstresses, working proprietors as well as retail salespersons and clerks.[9]

Marriage rates also varied substantially across urban and rural areas. According to the 1910 census, fewer than 40 percent of women living in Stockholm were married or had at least one child, as shown in Appendix Table A.1. This contrasts sharply with rural areas and smaller cities, where 50–60 percent of women were married and had at least one child. While the differences between Stockholm natives and in-migrants were modest, marriage rates were slightly higher among in-migrants, reaching approximately 40 percent. Similar patterns emerge for fertility and labor force participation. These patterns align with the argument that opportunities for independent wage labor contributed to Stockholm's low marriage rates. For example, Kyle (1987) argues that women eschewed marriage because they could establish financially stable lives on their own without relying on a male household head.

3 Data and Sample

Our main data comes from full-count decennial censuses between 1880 and 1910. In Sweden, local priests were in charge of keeping registers of all inhabitants in their parish, recording demographic information such as dates of births, deaths, and marriages every year. These church books have formed the basis for the world's oldest running population records and is known for its high

[9]Appendix Table A.2 lists the most common occupations for women across different locations.

accuracy of spelling and birth years, improving accuracy of record linking across census rounds.

Demographic and occupational data. The census contains individual level data on year of birth, civil status, occupation, as well as family relationships between members of households. It also separates families within households in cases where multiple families reside together. Using this information, we identify siblings using a combination of parent and family indicators. Occupations are classified according to the HISCO system which identifies sector of work (Leeuwen et al., 2002). We also match occupations to HISCLASS, which categorizes occupations in terms of their skill requirements (Leeuwen and Maas, 2011). We use IPUMS's definition to measure labor force participation. A woman is defined as in the labor force if she has a reported occupation (excluding students and titles capturing noble ranks), her occupational title is not related to another household member (e.g., "worker's widow"), and is above 15 years old. We also construct an alternative measure including unreported female work (see Section 5.1). Using data on the family in 1880, we collect data on fathers' and mother's occupations, as well as a number of characteristics at the household level.[10] Apart from parents' occupations, we obtain for each household the number of families, generations, siblings, servants, and (un)related members present.

Income scores. As with many historical censuses, the Swedish census does not include individual level data on income. In order to circumvent this issue, we compute income scores that indicate the average income of individuals in a given occupation and location. To do so, we use data from the 1930 census, which included individual incomes.[11] We follow a non-parametric approach similar to Ward (2020). In a first step, we create cells by county and 3-digit occupations separately for men and women. When cells include at least 30 individuals, we assign the mean value of income to the county-occupation combination. When cells have less than 30 individuals, we assign the mean national income for this occupation as long as there are at least 30 individuals with this three digit occupation nationally. In cases where there are fewer than 30 individuals, we use the mean income at the one-digit occupational level.[12] In a second step, income scores are then adjusted by county-level CPI to account for regional cost differences, as well as urban-rural price differences within counties based on Collin (2016). At the same time, income scores are adjusted using estimated urban and Stockholm-specific income premia. These premia are calculated using a sample of 5,000 tax records from 1900, which include information on parish of residence, allowing us to distinguish income premia by urban and rural areas, including a specific premium for Stockholm. The premia for urban areas and Stockholm are 41 and 66 percentage points, respectively.

[10]Note, however, that less than 2% of our sample individuals' mothers have reported occupations in 1880.

[11]The full 1930 census is not publicly available and is thus not possible to use directly in this study.

[12]To adjust agricultural earnings upwards to address that some compensation was in-kind, we follow Collins and Wanamaker (2022) and inflate the earnings of both farmers and farm hands by 35 and 19 percent, respectively.

Linking procedure. Our empirical strategy relies on within-sibling comparisons. In order to identify siblings and other household information, we focus on children observed in their childhood home in 1880 and link them to the 1910 census when they are in their adulthood. To do so, we rely on probabilistic linking methods. We briefly describe the linkage procedure here and describe it in more detail in Appendix B.

We first designate index variables which have to match exactly for two records to be considered potential matches: sex, birth year, and parish of birth. The detail and accuracy of these time-invariant variables allow us to construct a relatively small set of candidate links. In particular two features stands out as favorable compared to other national censuses. First, since local priests were in charge of keeping the registers, the birthplace is recorded at the parish level, which constitutes a relatively small geographic area (there were about 2,400 parishes during our time period). Second, since the parish books were continuously updated, birth years do not suffer from recall error, something which is evident from the lack of age-heaping in the Swedish censuses.[13] The latter allows us to only consider potential matches among candidates with the same exact birth year.

In a second step of our linking procedure, we evaluate these candidate links by comparing first and last names.[14] Importantly, censuses typically recorded women's maiden names instead of or together with their married names, allowing married women to be identified over time. To assess name similarities, we employ the Jaro-Winkler algorithm, which compares two strings and assigns a similarity score between 0 (no similarity) and 1 (identical). We consider individuals linked if there is a unique identical match within the same sex$\times$birth year$\times$place of birth cell (implying a Jaro-Winkler score of 1) or a unique match in the same such cell that satisfies a Jaro-Winkler threshold of at least 0.85 for both the first and the last name. For the latter, we require that there is no close runner-up.[15]

In terms of linkage rates, we are able to identify 66 percent of individuals born 1862–1880 in the 1910 census back to the 1880 census.[16] This retrospective match rate over a period of 30 years (1880–1910) compares favorably to the existing literature using linked historical census data. For comparable time periods, Long and Ferrie (2013) links 20.3–21.9% in Britain and the US, Ward (2020) links 9.1 % of linkable sons in a triple-linked sample from US censuses 1910–1940, and Modalsli (2017) links 37% in Norway.[17] Our linkage rates are highly similar for men and women, at 65 and 67 percent, respectively.

[13] See Berger et al. (2023) for a comparison of age-heaping between different national sources.

[14] We follow Wisselgren et al. (2014) and categorize surnames as either family or patronymic names. To impute children's surname if missing, we use information of the father's surname and his first name (to construct patronymic surnames), as well as information on the mother if the father is not present. See Appendix B for more details.

[15] We set the cut-off at a distance of 0.05 units between the highest ranked candidate and the runner-up.

[16] The forward linkage rate is 43 percent.

[17] The recent Census Tree Project (see Buckles et al., 2023), which combines genealogical and more traditional linking methods, achieves rates between adjacent censuses of about 69–86% for men, and 58–79% for women.

To complement our linked sample between the 1880 and 1910 censuses, we add links from the 1880 census to 1890 and 1900 based on Wisselgren et al. (2014), obtained from IPUMS International. In order to explore the external validity of our findings, we also use linked data for the U.S. census obtained from the Census Tree Project (Buckles et al., 2023).

Linked census sample. Our sample consists of individuals from rural parishes that were between 0 and 16 years of age in 1880 and that are also observed in 1910. In our main analysis, we keep migrants who live outside of their childhood parish in 1910. Comparing only migrants allows us to study the role of the destination and, in particular, large cities, rather than the role of migration itself. As robustness, we also display results when only including urban migrants as well as an extended sample including non-migrants.

To be able to compare siblings within families, we restrict attention to individuals that have an identified same-sex sibling. After this restriction, the sample consists of 77,597 women from 34,326 origin households.[18] Summary statistics for different samples are presented in Appendix Table A.3. In terms of childhood characteristics measured in 1880, migrants with same-sex sibling have similar characteristics as those without, although their mothers are somewhat less likely to be in the labor force.

While we are able to achieve relatively large match rates, it is possible that matched individuals differ systematically from those that are unmatched, possibly yielding unrepresentative estimates. For example, it is easier to link individuals with uncommon names, and name commonality has been linked to traits such as individualism and socio-economic status. With this in mind, Appendix Table B.1 compares matched individuals to the full population in the same age cohorts on observable characteristics measured in 1880. The table shows overall small differences between the two samples, suggesting that our sample is representative of the population. Nevertheless, we show that our results are nearly identical when we use probabilistic weights, reflecting the probability of an observation being selected into the sample.

Age of death. Data on age of death is acquired by linking individuals from the 1910 census to the Death Index (*Dödboken*), which collects dates of death for all individuals that died in Sweden. We link these data with the same procedure as when linking individuals between the 1880 and 1910 censuses. The forward match rate for both men and women born 1864–1880 observed in the 1910 census is 82 percent.

[18] The male sample, used for comparison, consists of 66,070 individuals from 29,334 households.

4 Empirical Strategy

We start our analysis by estimating outcomes of migrants to destinations of different population size, using the full sample of women originating from a rural area, whether or not they migrate:

$$Y_{if} = \sum_{p=1}^{100} \beta^p Migrant_{if}^p + \gamma \mathbf{X}_{if} + \delta \mathbf{W}_f + \varepsilon_{if} \tag{1}$$

where Y_{if} is the outcome of interest for individual i from childhood family f. $Migrant_{if}^p$ is a binary variable taking value 1 if the individual has moved to a parish at population percentile p by 1910, and value 0 if the individual remains in the origin parish. Thus, β^p is a set of coefficients estimated for each population percentile, capturing how migrants to these destinations of varying size differ from individuals staying at their origin parish. Moreover, $\mathbf{X}_{if}$ is a matrix of fixed effects for birth year, birth order, and being the eldest sister, while $\mathbf{W}_f$ is a matrix of family-level controls measured in 1880. The family-level controls include: a full set of origin parish fixed effects, father's percentile income score rank, family size, an indicator for mother's LFP, fixed effects for father's 1-digit HISCO occupation, the presence of servants in the household and whether the household consists of a married/cohabitating couple with children, a single-parent family, an extended family (relatives only), or a composite household (family and non-relatives), as well as whether the family is multigenerational.

To isolate the role of migrating to the largest urban areas, we then focus on a sample of migrants – defined as having left the childhood parish in 1910 – and estimate the returns to migration to Stockholm, the largest and fastest growing city during our period of study. To do so, we estimate

$$Y_{if} = \beta_1 SthlmMigrant_{if} + \beta_2 OtherUrbanMigrant_{if} + \gamma \mathbf{X}_{if} + \delta \mathbf{W}_f + \varepsilon_{if} \tag{2}$$

where $SthlmMigrant_{if}$ is an indicator taking value 1 if the individual lives in Stockholm in 1910, and 0 otherwise. To differentiate between rural migrants and migrants to other urban areas, we also introduce an additional indicator variable in our main regressions denoting migrants to an urban area other than Stockholm. We can then interpret β_1 as the difference in outcomes between migrants to Stockholm and migrants to rural parishes. In an alternative specification, we also estimate models with a dummy for *ever* having migrated to Stockholm or another urban area by 1910, thus including temporary migrants. As a robustness test we further study two alternative sample definitions: (1) including non-migrants who remain in the childhood parish and (2) including only urban migrants. These changes yield similar results as our main specification.

While the above models have the advantage of controlling for a number of pre-migration variables at the individual and the family level that might influence both the decision to migrate and outcomes such as employment or income score, there may still be many unobserved factors that can lead to bias. In order to improve on this, we can limit the analysis to comparing outcomes between sisters who live in the same household in 1880.[19] By restricting attention to within-siblings variation, we hold constant all common parental influences on children, whether environmental or inherited, that may otherwise have a direct effect on both migration and our outcomes of interest. The empirical model becomes

$$Y_{if} = \beta_1 SthlmMigrant_{if} + \beta_2 OtherUrbanMigrant_{if} + \gamma \mathbf{X}_{if} + \phi_f + \varepsilon_{if} \quad (3)$$

where ϕ_f is a fixed effect common for siblings from family f, which we henceforth denote a "sibling fixed effect". Since the sibling effect absorbs any variation that does not vary within the family, the family-level controls are subsumed in this specification. All models use cluster-robust standard errors at the (childhood) family level.

Selection into migration Our main analysis compares individuals who self-select into different migration destinations. To gauge the importance of selection bias in our estimates, Panel A of Appendix Figure A.2 shows how a variety of pre-migration variables correlate with the decision to move to either Stockholm or a minor city, relative to a rural area.[20] Interestingly, we find that selection is overall considerably more pronounced for migrants to urban areas *other* than to Stockholm. Because our results below indicate a much larger impact of migration to Stockholm than to smaller cities, this is a first indication that selection on observed individual and household characteristics are not important drivers of our results.[21] In terms of family characteristics, we find that migrants to Stockholm and other urban areas are less likely to have a father working in the agricultural sector, with the difference more than twice as large for the latter group. However, we find no evidence of selection for Stockholm migrants in terms of the father's real income score in 1880, nor family size. The parental household's composition is predictive of migration to urban areas in general, but much less so for Stockholm. Finally, at the individual level, migrants are less likely to be the eldest sister, more likely to have a higher birth order, and are more likely to be younger. However, similar to before, these factors are more predictive of migration to urban areas other than Stockholm. Thus, if these factors were important in determining our results, we should find the largest results for urban areas other than Stockholm, while we in practice find the opposite.

As a second comparison, we show correlates of migration decisions for the male sample in Panel B of Figure A.2. Again, we find that selection is more prominent for migrants to urban

[19]This strategy has been applied in a variety of studies of historical migration such as Abramitzky et al. (2012), Collins and Wanamaker (2014), and Ward (2020).

[20]See also Appendix Table A.3 for summary statistics across destination categories.

[21]While this argument applies to the characteristics that we observe, we also follow Oster (2019) to account for unobserved heterogeneity, finding evidence of selection tending to underestimate our results. See Section 5.1.

areas other than Stockholm. Estimated magnitudes of selection for both destination types are very similar to those of female migrants. In spite of this, our results below indicate large differences in outcomes across gender, which we take as a further indication that selection into migration on these characteristics is not a major driver of our results.

In the within-siblings model, we absorb any selection in terms of family characteristics and control for observable individual characteristics included in Appendix Figure A.2. The identifying assumption is that then migration status in 1910 should be as good as randomly assigned across individuals of the same sex within a family, conditional on birth year, birth order and an indicator capturing if the individual is the eldest sister. However, it remains possible that migrants to Stockholm are inherently different from their siblings migrating in other dimensions, such as taste for independent wage work or marriage. In order to account for such possibilities, we also estimate models that control for adult pre-migration outcomes. Making use of individuals linked to the 1900 census, we estimate models that restrict attention to individuals who migrate after 1900, allowing us to control for lagged outcomes in that year, when our sample is between 20 and 36 years old. The model then becomes:

$$Y_{if} = \alpha Y_{if,(t-1)} + \beta_1 SthlmMigrant_{if} + \beta_2 OtherUrbanMigrant_{if} + \gamma \mathbf{X}_{if} + \phi_f + \varepsilon_{if} \quad (4)$$

where $Y_{if,t-1}$ is the lagged outcome variable. If it is the case that there is within-family selection of migrants such that for example those with greater taste for work are both more likely to migrate to Stockholm and to work post-migration, these models will take such selection into account. Given that our sample individuals are 20–36 of age in 1900, they are well into adult life and would have made decisions regarding labor supply and family formation.

The above approach notwithstanding, we cannot fully exclude the influence of unobservable individual characteristics on our estimates. Our strategy takes as given the migration choices of individuals and aims to find the best possible counterfactual for migrants by comparing their outcomes to their sisters and controlling for individual characteristics such as pre-migration outcomes.[22] The purpose of our strategy is to improve on the naive comparison of movers and stayers to obtain a more accurate estimate of the treatment-on-the-treated effect for Stockholm migrants. Yet, we believe that there is an interest in documenting these patterns among female migrants even in a descriptively. Nevertheless, while our strategy does not directly account for all possible confounders, we also apply the method of Oster (2019) in order to quantify the magnitude of bias that may be caused by unobservable characteristics after controlling for all observables, including lagged outcomes. These tests are also reassuring, indicating that our results are biased toward zero, if

[22]We also make use of a sample of twins to account for additional unobserved within-sibling differences.

anything. See the discussion of robustness in Section 5.1 for further details.

5 Results

5.1 Employment, Marriage, and Fertility

Figure 3 displays estimates of Equation (1) on labor force participation with separate coefficients by population percentile of the destination parish. The figure shows that female migrants moving to the most populous parishes were considerably more likely to be in the labor force compared to migrants moving to more sparsely populated areas. Interestingly, employment is essentially identical among non-migrants and migrants up until the 90^{th} population percentile. Beyond this threshold we find a marked non-linearity, with employment at the top percentile of population being approximately 30 percentage points higher than among non-movers. The dashed shaded lines indicate estimates with sibling fixed effects, which are generally very similar to the baseline model estimates indicated by the solid lines. In sharp contrast to female migrants, the relationship between migration and labor force participation is always positive for males but remains flat across the population distribution.

Motivated by the non-linearity with respect to population, we turn to Equation (2) and estimate models of moving to Stockholm, the largest and most densely populated city at the time. To isolate the role of the city rather than migration itself, our baseline sample only includes rural-born migrants. Table 1 displays estimates of Stockholm migration on having labor force participation in 1910. Column 1 shows that female migrants to Stockholm were 49.5 percentage points more likely to be in the labor force compared to all other female migrants. To compare Stockholm migrants with women who stayed in the rural economy, Column 2 adds an indicator variable for migration to urban areas other than Stockholm. In line with Figure 3, this has a modest effect on our estimate, which increases slightly to 53.4 percentage points.[23] While migrating to another urban area is also associated with higher FLFP, it is notably smaller than the estimate for Stockholm migration. Column 3 adds fixed effects for the individuals' birth year, birth order, being the eldest sister and a variety of family characteristics in the 1880 census, including fixed effects for the childhood parish (see Section 4 for full list). Adding these controls makes only minor changes to the estimate.

Finally, columns 4 and 5 focus on the sample of sisters. Column 4 first reproduces the results from column 3 within this subsample, displaying nearly identical coefficients. Column 5 then introduces sibling fixed effects as in Equation (3), and thus bases its estimate only on comparisons of sisters who lived in the same household in childhood. Capturing all common nature and nurture

[23] In Appendix Table A.5, we alter the sample and document similar results when using an extended sample including non-migrants as well as a smaller sample using only urban migrants.

effects that may influence both migration status and the outcome, the estimated coefficient for migrating to Stockholm becomes 49.7 percentage points. Relative to the average outcomes for women at 22.8 percent, the effect of migrating to Stockholm thus entails more than a doubling of labor force participation.

Notably, these results stand in stark contrast to the relationship for men, as men migrating to Stockholm were only one percentage point more likely to be in the labor force compared to rural migrants. All results for the male sample are shown in Appendix Table A.4.

Underreporting of occupations. A general empirical concern when studying employment of women is that female work is often under-reported in historical census data (Goldin, 1990; Stanfors, 2014). This is particularly the case for married women in farming households, who were often responsible for tasks such as animal husbandry and cow-milking (Morell, 2001). This work was combined with childcare and household tasks, and for this reason rarely corresponded to full-time work. Therefore, following Stanfors (2014), our main focus is on employment as measured in the census, which provides a measure of labor force participation outside the home, in the paid market, and to a greater extent captures full-time work. Nevertheless, we address the issue of underreporting of female occupations in two ways. First, we reduce the issue of measurement by dropping women for whom labor force participation was least likely to be precisely reported. To do so, Table 1, column 6, drops women who in 1910 lived with a male household head (either her father or husband) who was engaged in agriculture in the same year. Doing so yields an estimate of 44.8 percentage points, strikingly similar in magnitude to our preferred estimate in column 5.

We also provide estimates using a conservative measure of FLFP that includes the most common possibilities for informal work in column 7. To do so, we set FLFP to be equal to one for any woman living with a male household head in 1910 who is either engaged in farming or is a working proprietor in sales or services in 1910.[24] This conservative adjustment still yields a large, positive estimate of 22.9 percentage points increase in FLFP. Although the new outcome variable deviates from our focus on full-time, market-oriented work it is noteworthy that our estimate for migration to Stockholm still has a sizeable positive magnitude. In contrast, sisters migrating to other urban areas see a negative relationship with the conservative measure of FLFP.

Marriage, fertility, and FLFP. The previous section showed that female migrants to cities in general – and Stockholm in particular – have substantially higher labor force participation rates. However, women in the early-20th century typically faced a choice between having a family and being employed, as the vast majority of women exited the labor force upon marriage.[25] Therefore,

[24] See Chiswick and Robinson (2021) for a similar approach. The adjustment for women living with working-proprietor household heads has only minor implications for the estimates.

[25] In our main sample, less than 2 percent of married women in 1910 had a reported occupation.

we next turn to investigating migrant women's family formation. Figure 4 displays estimates based on Equation (1) where the outcome is an indicator for being married (Panel A) or having at least one child (Panel B) in 1910.[26] As before, we report separate coefficients by the population percentile in the destination. The estimates in Panels A and B of Figure 4 show that female migrants moving to the most populous parishes are much less likely to marry and have a child. By contrast migrants to destinations in the bottom-90 percent of the population distribution are substantially more likely than non-migrants to marry and have at least one child, respectively. Estimates are very similar when including sibling fixed effects reported as dashed, shaded lines in both figures.

Table 2 presents regression results for migrants to Stockholm and other urban areas. Column 1 indicates that women in our within-sibling design are 50.9 percentage points less likely to be married in 1910 if they migrated to Stockholm between 1880 and 1910 as compared to sisters migrating to rural locations.[27] Relative to the average among women of 71 percent, this represents a substantial decrease of about 70 percent. Column 2 estimates effects on childbearing, showing that female migrants are 47.2 percentage points less likely to have any children by 1910, with a relative effect 70.7 percent compared to the mean. Analogous to the case of labor force participation, migrants to other urban areas see more modest decreases.

While most female migrants that took up formal employment in Stockholm at the same time remained single and childless, the increases in employment are not solely driven by the lower marriage rates among migrants. Table 2, column 3, reports estimates where we limit the sample to sister pairs where both were married in 1910. While the coefficient is much more modest as compared to the main sample, it represents an almost twofold increase from the sample mean. In column 4, we instead restrict attention to unmarried women in 1910. Conditional on being single, women in Stockholm are nearly 20 percentage points more likely to be in the labor force compared to rural migrants.

Given the strong observed link between work and family formation, a potential concern for interpreting our results is whether the timing of marriage, for example, precedes and affects the decision to migrate or not. For instance, if single sisters moved to Stockholm and married sisters to rural areas, our results would be driven by marital decisions rather than the migrant destination itself. To address such issues of timing, we make use of the time dimension of our data and focus on women observed in both 1900 and 1910. Column 5 of Table 2 studies women who are unmarried and have yet not migrated in 1900. We find that this subgroup of women who as young adults (aged 20-36) were unmarried prior to migration exhibit outcomes in 1910 that are very similar in magnitude to the unrestricted sample. If anything, their labor force participation is somewhat

[26] Appendix Figure A.3 displays corresponding figures for the male sample.

[27] Appendix Table A.6 displays the results for different samples as well as their stability to regressions with and without controls and sibling fixed effects.

higher. Moreover, marriage and childbearing rates are still substantially lower in 1910 for these Stockholm migrants compared to their sisters, as shown in Appendix Table A.7.

Nevertheless, it is still possible that women who were unmarried in 1900 could have married before migrating in 1910, when we next observe them. To fully remove the possibility that marriage occurs before migration, the last column of Table 2 therefore further restricts the sample to women who in 1900 were unmarried *and* had already migrated.[28] Even within this sample of women who migrate before (potential) marriage, we find that Stockholm migrants are substantially more likely to be in the labor force in 1910. Analogously, they are less likely to be married or have children in 1910 (see Appendix Table A.7). The somewhat smaller estimates in these specifications are driven by older unmarried individuals in 1900, as older individuals who are unmarried in 1900 are less likely to ever marry. Appendix Table A.8 shows that when restricting attention to individuals below 27 (the median age of our sample in 1900 and the age at which a majority of women are married), estimates are similar in magnitude to those in the main specifications.

Persistence and dynamics Above, we have shown cross-sectional results that give a snapshot in time of migrants' work and family outcomes. We next consider the dynamics and possible persistence of migration outcomes by destination. Table 3, columns 1, 3, and 5 tests for persistence by studying the 1910 outcomes of those who migrated by 1900. This specification allows for migrants to relocate in the intervening 10 year period, and thus captures post-migration work and family choices. Columns 1, 3, and 5 shows that even after 10 years, migrants to Stockholm are 34 percentage points more likely to be in the labor force, and nearly 38 percentage points less likely to be married or have children. In a closely related specification, columns 2, 4, and 6 instead display results of *ever* having migrated to Stockholm between 1890 and 1910, again including both individuals who remained and those who moved away over time. The associated estimates are highly similar. Taken together, the results in Table 3 show that Stockholm migrants have persistently different outcomes than other migrants, even when allowing for later relocation decisions. At the same time, the smaller estimates also indicate that temporary Stockholm migrants at least to some extent reverted towards the work and family decisions of migrants to other destinations.[29]

To further explore how migration relates to outcomes over migrants' life-cycles, we next estimate binned scatterplots by age for migrants both in 1900 and 1910, together spanning between the age of 20 (the youngest cohort among 1900 migrants) up to age 46 (the oldest cohort among 1910 migrants). Results, including controls and sibling fixed effects, are shown in Figure 5 for

[28]Specifically, we restrict the sample to migrants who were observed in the same type of destination in both 1900 and 1910: either Stockholm, another urban area, or a rural area.

[29]To closer study the outcomes of temporary migrants, Appendix Table A.9 focuses on a subsample of individuals who migrated to Stockholm by 1900 but had left by 1910. Temporary Stockholm migrants still are significantly more likely to work, and less likely to form a family in 1910 compared to other migrants. However, estimates are considerably smaller, indicating that family formation often occurred conjointly with migrating away from Stockholm.

our three main outcomes. Panel A indicates that while migrants to rural areas quickly decrease their labor force participation between the ages of 20 and 30, Stockholm migrants remain in the labor force at high rates during the same age span. After age 30, Stockholm migrants begin to decrease their labor force participation somewhat, but even by their mid-forties participation rates are substantially higher compared to their sisters migrating to rural areas. This pattern is mirrored for marriage and having any child as seen in panels B and C of Figure 5. Stockholm migrants begin forming families to some degree after age 30, roughly ten years after their sisters in rural areas. Even in their mid-forties, having surpassed the typical childbearing age, differences in marriage and childbearing rates are large.[30] Thus, these results suggest that differences in family outcomes are persistent and cannot be explained by delayed family formation.[31]

Robustness. One worry about the empirical strategy of comparing siblings is that individual differences even within sibling groups may be important determinants of migration or labor market outcomes. Although estimates in Table 1 include controls for age, birth order, and being the eldest sister, we can additionally use pre-migration employment status for the subset of individuals who move to Stockholm later in life. We show in Appendix Table A.11 that our results are robust among the subset of migrants that move after 1900 where we can directly control for pre-migration lagged outcomes that may differ between siblings. Reassuringly, adding these controls yields very minor changes on our main estimates. We then show in Appendix Table A.12 that our findings are similar when applying twin rather than sibling fixed effects to account for additional unobserved within-sibling differences.

As an additional test for individual differences between siblings, we apply the method of Oster (2019). This method compares the differences in models with and without controls to infer what the true estimate would be if one could control for all unobserved characteristics. Specifically, we hold constant the sibling fixed effect and use individual-level controls on age, birth order, being the eldest sister, as well as lagged outcomes to determine the role of any remaining unobserved within-siblings *individual* variation. Appendix Table A.13 displays our results for FLFP, marriage, and fertility. Interestingly, all six models indicate that the bias-corrected estimate is larger than or equal in magnitude to the estimate based on observable controls. Thus, we find that there is somewhat *negative* self-selection within siblings group, correspondingly pulling our estimates towards zero. Our baseline models find small differences between the controlled and uncontrolled regressions. The bias-corrected for labor force participation is 54.0 percentage points, close to

[30] Appendix Figure A.4 shows the distribution of age at first child among women living in Stockholm in 1910.

[31] Appendix Table A.10 provides results from a model with heterogeneous effects of migration to Stockholm by age in 1910. The estimates indicate that, for Stockholm migrants, an age difference of 10 years is associated with being 8 percentage points less likely to work, as well as being 5 and 4 percent more likely to be married and have a child, respectively.

our main estimate of 49.7.[32] Using our sample of migrants after 1900, which allows us to add pre-migration lagged outcomes for employment, marriage and childbearing, we find a similar pattern. The corrected estimate in column 2 is 53.7 percentage points, which can be compared to the baseline estimate of 52.4. For marriage and childbearing, we find estimates that are again similar to the baseline estimates. Thus, we conclude that our estimates are robust to taking into account potential unobserved characteristics at the individual level that may vary within sibling groups.

We address a number of additional empirical concerns in the appendix. First, Stockholm is known for having started the trend of cohabitation as an alternative to marriage (so-called *Stockholm marriages*). To the extent that Stockholm migrants chose to cohabit instead of marrying, the results for marriage in Table 2 may be misleading. However, we show in Appendix Table A.14 that the increase in cohabitation for female migrants is small in magnitude.[33] Second, our census data include information on children born out of wedlock, which can be used as an alternative measure of informal marriages. Using these data, column 3 of Appendix Table A.14 shows that female migrants are less likely to have any children even when restricting attention to those born out of wedlock.

Appendix Table A.15 displays our results when weighting observations by their likelihood of being in the sample, in order to check for issues regarding potential unrepresentativeness caused by our linking procedure. Reassuringly, results are almost identical to our baseline models.[34] Appendix Figure A.5 also documents that our results are nearly identical when accounting for: i) age differences between migrants and their siblings, ii) whether the family sent only one migrant or at least two, and iii) whether a migrant's family size is above or below the median. These tests indicate that our results are not driven by migrants tending to be systematically younger or older than their siblings, or by specific types of households.

Migrants in the United States. We lastly replicate our results using linked census data from the United States. As described above, we focus on Sweden because it provides a unique opportunity to link women in census data using standard record-linkage techniques. However, an important question is whether our results can be generalized to other contexts. While we descriptively show in Section 2.1 that FLFP rates were substantially higher in the largest cities in both Europe and the United States, it is an open question whether our individual-level results for migrants would replicate in other settings. Therefore, we make use of census links from Buckles et al. (2023)

---

[32] Computations are made using the parameter $\tilde{R} = 1.3R$ following Oster (2019), where $\tilde{R}$ is the assumed maximum R that would be explained by a model including all variables, both observed and unobserved, and R is the observed R^2. Observed and unobserved variables are assumed to have the same impact on outcomes, $\delta = 1$.

[33] The measure of cohabitation is constructed as a binary indicator variable takes the value one if an individual (i) lives in a household with only one other adult, who (ii) is of the opposite sex, and (iii) both individuals are unmarried.

[34] To calculate weights, we use the full census data to regress an indicator for being successfully linked on age, age squared, as well as fixed effects for birth order, childhood parish, and father's social class.

that use genealogical information and machine learning algorithms to track individuals over time in US censuses. Importantly for our purposes, links are available for women as well as men. By combining these links with data from the US censuses of 1880 and 1910 (IPUMS, 2020), we create a panel of American women and their sisters that matches our Swedish sample.[35]

Figure 6 displays estimates of Equation (1) showing the effect of migration on employment by the percentile population rank in the destination county among US women. Estimates including sibling fixed effects are displayed in the shaded line. Similarly to the case in Sweden, female migrants in the US are substantially more likely to transition into the formal labor market in the most populous US counties. By contrast, there is little variation in the FLFP for women moving to counties below the 90th percentile, indicating that effects are driven by the top of the population distribution in the US as well in Sweden. Panels A and B of Appendix Figure A.6 show that marriage and childbearing outcomes follow a the same pattern, with lower likelihood of family formation in the largest cities. Although magnitudes naturally differ as the countries are substantially different, the general pattern of effects is remarkably similar across the United States and Sweden.

5.2 Occupations, Income, and Health

In this section, we study what type of occupations that drive the differential patterns documented above as well as potential income differentials and health penalties associated with migrating to the city.

Sector of Employment. We start by studying what sector of employment that is associated with migration along the population distribution. Figure 4, Panel C, displays estimates from Equation (1), where the outcome is now the probability of working in a particular sector of employment. We report separate coefficients by the population percentile of the destination parish. Starting with the service sector (displayed in red), we can note that in destinations below the 90^{th} percentile, female migrants are not more likely to work in the service sector than non-migrants. In the most populated areas, however, migrants are more than 20 percentage points more likely to transition into service work. Focusing on Stockholm migrants, Table 4, column 1, shows that the probability of working within the service sector is drastically higher for women migrating to Stockholm. The point estimate indicates a 43.4 percentage point higher likelihood of service sector work among female migrants compared to their sisters moving to rural areas. This corresponds to an almost fourfold increase from the mean in our sample. The service sector alone does not explain the full increase, however, as employment increases in industrial occupations as well, by 7.8 percentage points (column 2). In contrast to the discontinuous increase in service sector employment

[35] We use similar sample restrictions, focusing on prime-aged women (aged 30–46 in 1910) who lived in a rural area in 1880.

in the most populous areas, however, the increase in industrial workers displays a more continuous increase throughout the population distribution. This is seen both in Figure 4 (displayed in blue) and in the relatively large coefficient for migrants to other urban areas in Table 4, column 2. Appendix Table A.16 breaks down the sector of employment further according to the major groups in HISCO. Apart from agriculture, which unsurprisingly displays a negative relationship to Stockholm migration, we find positive estimates for all other occupational groups.

Skills. An important question is whether female migrants to Stockholm mainly transitioned into unskilled occupations or if they also took on more qualified employment. Table 4, column 3 shows that Stockholm migrants are 7 percentage points more likely to work in higher skilled occupations, which include medium and high skilled work.[36] Column 4 shows that Stockholm migrants are 38.7 percentage points more likely to work in low-skilled occupations compared to their sisters migrating to rural locations. Lastly, column 5 shows that a much smaller, but still statistically significant increase of 0.4 percentage points in the least skilled occupations. Focusing on the ten most common occupations among migrant women in each skill group, Appendix Figure A.7 shows additional detail on the specific occupations that women take up across the skill spectrum. The highest increase is found for maids, followed by hand and machine sewers, both of whom are classified as low skilled. However, we also observe significant increases for several higher skilled occupations such as working proprietors, professional nurses, cooks, and housekeepers in private service.

We next investigate how migrants' occupational skill requirements evolve over the life-cycle. Panel D of Figure 5 displays a binned scatter plot showing the likelihood of migrants working in higher skilled occupation by age, for both Stockholm and rural destinations. The figure shows that Stockholm migrants exhibit signs of skill upgrading over time. Across the age range of 20–46 that our sample comprises, migrants in Stockholm are successively more and more likely to work in such occupations. In their 20s, migrants are about as likely to work in higher skilled occupation in Stockholm as in rural areas. However, Stockholm migrants in their 30s and 40s are considerably more likely to do higher skilled work. Turning instead to following the same individuals over time, column 4 of Table 3 shows that individuals who move to Stockholm in 1900 are on average 7.6 percentage points more likely to work in higher skilled occupations ten year years later. The estimate for ever movers to Stockholm is similar.[37] Thus, migrants to Stockholm increased their likelihood to work in both low and higher skilled occupations. Moreover, they experienced significant occupational skill-upgrading over time, a pattern not observed among migrants to rural areas.

[36] To classify occupations by their skill requirements, we use the HISCLASS scheme following Maas and van Leeuwen (2005).

[37] Appendix Table A.17 displays corresponding estimates for low and unskilled occupations.

Income. Although female migrants to Stockholm were more likely to engage in market work and earn an independent income, the lower rates of marriage could imply that they experienced lower living standards at the household level. To investigate this, Column 5 of Table 4 shows results from a regression where the outcome is log household income score per adult household member. This takes into account cases in which women do not have reported occupations and associated income scores. The estimates indicates that household income scores increase by 20.6 percent for Stockholm migrants relative to rural areas. Thus, female migrants to Stockholm appear to have increased their disposable income. Moreover, Appendix Figure A.8 shows that household income displays a clear increasing trend with age, but only for Stockholm migrants. While the estimate in column 5 assumes that households shared income equally, we show in Appendix Table A.18 that the positive estimate for Stockholm migrants is even more pronounced when down-weighting the contribution of the spouse. This suggests that the increase in household income is not driven by Stockholm migrants matching with spouses with higher income. Appendix Table A.18 also shows that the intra-household ratio of female to male income is higher for Stockholm migrants, suggesting that female migrants to Stockholm experiencing increased economic independence relative to migrants to other destinations.

Focusing instead on women with reported occupations, column 7 of Table 4 shows that female migrants have about 7.4 percent higher real income scores compared to their sisters migrating to rural locations. One worry with using income scores at the individual level is the fact that it conditions on being employed, which is likely to be highly positively selected outside of Stockholm, where female employment is more rare. To account for this, we include occupation code fixed effects to compare individuals doing similar work. The estimate in column 8 shows that the impact on real income scores is approximately three times larger in this specification, at 22.2 percent. Thus, when comparing individuals within the same occupation, female migrants to Stockholm obtained even higher real income scores than in the unrestricted model.[38] Regardless of level of analysis, however, our findings indicate a considerable increase in real income scores for migrants to Stockholm.

Intergenerational mobility. Given the increase in real income scores, a related question is whether Stockholm migrants also experienced higher rates of intergenerational mobility. Appendix Figure A.10, Panel A, displays the association between children's and parents income ranks. Along the horizontal axis, we plot the father's income rank based on occupation in the 1880 census. On the vertical axis, we display the mean income rank attained by daughters in adulthood. Red dots indicate the average income ranks attained by Stockholm migrants, while blue dots correspond to

[38] Instead of conditioning on the exact choice of occupation, we can instead control for 1-digit occupation codes to broadly capture the choice of sector. This also yields a larger estimate for income scores at 16.7 percent.

the mean income ranks of rural migrants. Migrants on average attain higher income ranks conditional on their father's income, which correspond to a higher level of absolute mobility. The figure also presents estimates of relative mobility, as captured by the rank-rank slope of daughters' and fathers' income ranks, which corresponds to the correlation between parents and children's place in the income distribution. The smaller rank-rank associations among migrants indicates a lower degree of intergenerational persistence and thus a higher level of relative mobility. In Panel B of the same figure, we find that the pattern is similar when replacing the individual income measure with household income, taking into the fact that more rural women are married and have no own reported employment.

Health. Urban areas in Europe and the United States were historically characterized by worse health and shorter lifespans, making potential welfare gains of urban migration more ambiguous (Woods, 2003; Cain and Hong, 2009). To investigate mortality across destinations in our data, we match our sample to the official death registers to obtain individuals' age at death and create indicators for having survived past ages 31 to 80. Figure 7 displays our estimates of the difference in survival between migrants to Stockholm compared to rural areas. The results in red indicate that women who had migrated to Stockholm by 1910 were at least as likely to survive into old age as their sisters migrating to rural areas.[39] Female migrants are not significantly more or less likely to have survived across the range of ages we consider. Appendix Table A.19, Column 1 further shows no significant difference in average age at death. In contrast, column 3 shows that mortality among male migrants to Stockholm was considerably higher, as they died on average 2 years younger than their brothers in rural areas. Estimates in blue in Figure 7 show that survival rates among male migrants begin to diverge around age 50. By age 60, Stockholm migrants are approximately half a percentage point less likely to have survived. This survival penalty continues to grow until reaching its maximum at 8 percent points by age 70.

One potential explanation for the statistically insignificant difference in mortality between female migrants is that women in Stockholm may have transitioned into less hazardous service jobs upon migrating, thus negating any adverse health effects. To test for this, we include fixed effects at the 5-digit occupation level in column 2 of Appendix Table A.19, including an indicator for having no reported occupation. This test yields a more negative point estimate for Stockholm migrants, but it remains insignificant and relatively small in magnitude. The same test for males, in column 4, also indicates small changes to the estimate. This suggests that differences in occupational health and safety hazards, as measured in 1910, are not drivers of mortality. An alternative explanation for the small mortality difference in the female sample is that women in Stockholm

[39] Note that since the sample consists of individuals aged 30 to 46 in 1910, our results are mechanically less likely to find differences in mortality before age 47.

were less likely to have children, which may in turn have reduced the risk to to die in child birth or related complications. However, the estimates in Figure 7 show that the mortality differential for female migrants is not apparently related to women's childbearing age.[40]

The differential mortality rates by sex suggest that general health hazards at the city level – such as poor sanitation, hygiene, contagious diseases, or work conditions – are not the main explanation for mortality differences between migrants. Instead, male mortality may have been higher due to lifestyle related issues, such as high consumption of alcohol and tobacco (Dribe and Eriksson, 2018).

5.3 Accounting for FLFP: the Role of Services

A long literature links the entry of women into the labor force in the 20^{th} century to the expansion of service sector occupations. To what extent can services explain the results in our study? Several pieces of evidence point to the sector's importance. Figure 1 shows that differences in FLFP across the population distribution of the US and Europe correlate with the local share of service workers in the labor force. Similar to other European capitals, Stockholm experienced a significant shift toward the service sector by the early 20^{th} century, long before other regions. And, most directly, our analysis of Stockholm migrants in Table 4, column 1, indicates that services accounted for most of the employment gains of women in Stockholm.

To asses the role of services more extensively, we use the variation in sectoral composition across all migrant destinations and control for the same set of factors as our main models. First, we provide a simple decomposition in column 1 of Table 5, which shows that migrating to a location with a 10 percent larger service sector is associated with a 3.9 percentage point increase in employment. By contrast, a larger industrial sector instead predicts lower female employment. Second, to asses the extent to which the relationship between urban migration and employment is mediated by the service sector, we include both migration dummies and various sectoral composition measures in columns 2 to 4 of Table 5. Column 2 shows that the indicator for migration to Stockholm is reduced by about 20 percentage points compared to our main estimate in Table 1 when controlling for sectoral employment shares at migrants' destinations. Similarly, the estimate for migration to other urban areas is diminished by about 10 percentage points. In order to take into account the likely non-linear relationship between population and service sector size, we include cubic polynomials in column 3 and fixed effects for each percentile of the distribution of service and industry employment shares in column 4. These additions decrease our estimates further. For example, the lower estimate of column 4 roughly corresponds to a fifty percent decrease of our main estimate

[40]Even for the sample of women who are relatively young and may not have completed their fertility by 1910, Appendix Figure A.9 shows a similar pattern of nonsignificant differences in mortality.

from Table 1. Column 5 and 6 show corresponding results for marriage and fertility.[41] Thus, the availability of service sector jobs appears to be a major determinant of our results. Nevertheless, a substantial difference in employment and family formation remains, most notably in Stockholm. This suggest the presence of a city-specific component, such as norms, favoring female work.

6 Conclusion

This study provides new evidence on substantial spatial variation in women's employment and family formation patterns in the early 20th century. Using census data, we document large differences in FLFP rates within European countries and the United States, whereby women in the biggest cities were substantially more likely to be employed in the formal labor market. To establish the role of the city in shaping women's outcomes, we focus on the case of Sweden where we can track female migrants using linked census data. Women moving to the capital and largest city, Stockholm, were substantially more likely to enter the formal labor market and to remain single and childless compared to their sisters migrating elsewhere.

An important explanation for the higher FLFP rates in large cities is the fact that they had experienced an early structural shift towards the service sector, which facilitated the entry of women into paid work. Our evidence suggest that the difference in labor market composition across space can explain about half of the increase in FLFP observed in Stockholm. While our estimation strategy cannot fully rule out the influence of selection in migrant destinations, a number of tests indicate that it is unlikely to be a key determinant of our results.

Our study provides systematic evidence that large cities provided women with substantial labor market opportunities at a time when aggregate female labor force participation was decreasing. Women migrating to Stockholm saw persistent changes in their work and family outcomes. They experienced increases in individual and household incomes, alongside occupational skill upgrading over time, without encountering an adverse health impact from residing in the city. While many of these working women chose or had to forgo marriage and childbearing, they likely laid a foundation for later advances in female employment also among married women.

[41] Appendix Table A.20 replicates also columns 1–3 for marriage and fertility outcomes.

References

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

_ , _ , _ , Santiago Pérez, and Myera Rashid, "Census Linking Project: Version 2.0 [dataset]," *Data retrieved from, https://censuslinkingproject.org*, 2020.

_ , Leah Platt Boustan, and Katherine Eriksson, "Europe's tired, poor, huddled masses: Self-selection and economic outcomes in the age of mass migration," *American Economic Review*, 2012, *102* (5), 1832–56.

Akbulut, Rahşan, "Sectoral changes and the increase in women's labor force participation," *Macroeconomic Dynamics*, 2011, *15* (2), 240–264.

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

Boserup, Ester, "Woman's Role in Economic Development," 1970.

Bridgman, Benjamin, Georg Duernecker, and Berthold Herrendorf, "Structural transformation, marketization, and household production around the world," *Journal of Development Economics*, 2018, *133*, 102–126.

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

Buckles, Kasey, Adrian Haws, Joseph Price, and Haley EB Wilbert, "Breakthroughs in Historical Record Linking Using Genealogy Data: The Census Tree Project," Technical Report, National Bureau of Economic Research 2023.

Buera, Francisco J, Joseph P Kaboski, and Min Qiang Zhao, "The rise of services: the role of skills, scale, and female labor supply," *Journal of Human Capital*, 2019, *13* (2), 157–187.

Burnette, Joyce and Maria Stanfors, "Understanding the Gender Gap Further: The Case of Turn-of-the-Century Swedish Compositors," *The Journal of Economic History*, 2020, *80* (1), 175–206.

Cain, Louis and Sok Chul Hong, "Survival in 19th century cities: The larger the city, the smaller your chances," *Explorations in Economic History*, 2009, *46* (4), 450–463.

Carlsson, Sten, *Den sociala omgrupperingen i Sverige efter 1866*, Almqvist & Wiksell, 1966.

Cattaneo, Matias D., Richard K. Crump, Max H. Farrell, and Yingjie Feng, "On Binscatter," *American Economic Review*, May 2024, *114* (5), 1488–1514.

Cerina, Fabio, Alessio Moro, and Michelle Rendall, "The role of gender in employment polarization," *International Economic Review*, 2021, *62* (4), 1655–1691.

Chetty, Raj, Nathaniel Hendren, Patrick Kline, and Emmanuel Saez, "Where is the land of opportunity? The geography of intergenerational mobility in the United States," *The Quarterly Journal of Economics*, 2014, *129* (4), 1553–1623.

Chiswick, Barry R. and RaeAnn Halenda Robinson, "Women at work in the United States since 1860: An analysis of unreported family workers," *Explorations in Economic History*, 2021, *82*, 101406.

Collin, Kristoffer, *Regional wages and labour market integration in Sweden, 1732–2009* 2016.

Collins, William J and Marianne H Wanamaker, "Selection and economic gains in the great migration of African Americans: new evidence from linked census data," *American Economic Journal: Applied Economics*, 2014, *6* (1), 220–52.

Collins, William J. and Marianne H. Wanamaker, "African American Intergenerational Economic Mobility since 1880," *American Economic Journal: Applied Economics*, July 2022, *14* (3), 84–117.

Costa, Dora L, "From mill town to board room: The rise of women's paid labor," *Journal of Economic Perspectives*, 2000, *14* (4), 101–122.

Craig, Jacqueline, Katherine Eriksson, and Gregory T Niemesh, "Marriage and the intergenerational mobility of women: Evidence from marriage certificates 1850-1910," *April. https://niemesgt. github. io/files/WomenMobilityClio. pdf*, 2019.

Deutsch, Sarah, *Women and the city: Gender, space, and power in Boston, 1870-1940*, Oxford University Press, 2000.

Dribe, Martin and Björn Eriksson, "Socioeconomic status and adult life expectancy in early 20th-century Sweden: Evidence from full-count micro census data," *Lund papers in economic demography*, 2018, *2*.

Edlund, Lena, "Sex and the City," *Scandinavian Journal of Economics*, 2005, *107* (1), 25–44.

Enflo, Kerstin, Martin Henning, and Lennart Schön, "Swedish regional GDP 1855–2000: Estimations and general trends in the Swedish regional system," in "Research in Economic History," Emerald Group Publishing Limited, 2014.

Engel, Barbara Alpern, *Between the fields and the city: women, work, and family in Russia, 1861-1914*, Cambridge University Press, 1994.

Feigenbaum, James and Daniel P Gross, "Answering the Call of Automation: How the Labor Market Adjusted to the Mechanization of Telephone Operation," Technical Report, National Bureau of Economic Research 2020.

Fernández, Raquel, "Cultural change as learning: The evolution of female labor force participation over a century," *American Economic Review*, 2013, *103* (1), 472–500.

_ , Alessandra Fogli, and Claudia Olivetti, "Mothers and sons: Preference formation and female labor force dynamics," *The Quarterly Journal of Economics*, 2004, *119* (4), 1249–1299.

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

Fogli, Alessandra and Laura Veldkamp, "Nature or nurture? Learning and the geography of female labor force participation," *Econometrica*, 2011, *79* (4), 1103–1138.

Fuchs, Rachel G and Leslie Page Moch, "Pregnant, single, and far from home: Migrant women in nineteenth-century Paris," *The American Historical Review*, 1990, *95* (4), 1007–1031.

Glaeser, Edward and David Maré, "Cities and Skills," *Journal of Labor Economics*, 2001, *19* (2), 316–342.

Goldin, Claudia, "The work and wages of single women, 1870 to 1920," *The Journal of Economic History*, 1980, *40* (1), 81–88.

_ , *Understanding the gender gap: An economic history of American women* number gold90-1, National Bureau of Economic Research, 1990.

_ , "The U-Shaped Female Labor Force Function in Economic Development and Economic History," in T. P. Schultz, ed., *Investment in Women's Human Capital and Economic Development*, University of Chicago Press, 1995, pp. 61–90.

__ , "The quiet revolution that transformed women's employment, education, and family," *American economic review*, 2006, *96* (2), 1–21.

__ **and Kenneth Sokoloff**, "Women, children, and industrialization in the early republic: Evidence from the manufacturing censuses," *The Journal of Economic History*, 1982, *42* (4), 741–774.

Haines, Michael R., "Historical, Demographic, Economic, and Social Data: The United States, 1790-2002. ICPSR Study No. 2896-v3," https://doi.org/10.3886/ICPSR02896.v3 2010.

Hamory, Joan, Marieke Kleemans, Nicholas Y Li, and Edward Miguel, "Reevaluating agricultural productivity gaps with longitudinal microdata," *Journal of the European Economic Association*, 2021, *19* (3), 1522–1555.

Hill, Bridget, "Rural-urban migration of women and their employment in towns," *Rural history*, 1994, *5* (2), 185–194.

IPUMS, "Integrated Public Use Microdata Series, International: Version 7.3; Minnesota Population Center," 2020. https://doi.org/10.18128/D020.V7.3.

Johansson, Mats and Lars Olof Persson, *Stockholmsflyttare under hundra år*, Almqvist & Wiksell, 2004.

Kennan, John and James R Walker, "The effect of expected income on individual migration decisions," *Econometrica*, 2011, *79* (1), 211–251.

Kim, Sukkoo, "Industrialization and urbanization: Did the steam engine contribute to the growth of cities in the United States?," *Explorations in Economic History*, 2005, *42* (4), 586–598.

Kyle, Gunhild, *Handbok i svensk kvinnohistoria*, Stockholm: Carlsson, 1987.

Lee, Donghoon and Kenneth I Wolpin, "Intersectoral labor mobility and the growth of the service sector," *Econometrica*, 2006, *74* (1), 1–46.

Long, Jason, "Rural-urban migration and socioeconomic mobility in Victorian Britain," *The Journal of Economic History*, 2005, *65* (1), 1–35.

__ **and Joseph Ferrie**, "Intergenerational Occupational Mobility in Great Britain and the United States Since 1850," *The American Economic Review*, 2013, *103* (4), 1109–1137.

Maas, Ineke and Marco H.D. van Leeuwen, "Total and Relative Endogamy by Social Origin: A First International Comparison of Changes in Marriage Choices during the Nineteenth Century," *International review of social history*, 2005, *50* (S13), 275–295.

Mammen, Kristin and Christina Paxson, "Women's work and economic development," *Journal of economic perspectives*, 2000, *14* (4), 141–164.

Minnesota Population Center, "Integrated Public Use Microdata Series, International: Version 7.3 [dataset]," Minneapolis, MN: IPUMS 2020. https://doi.org/10.18128/D020.V7.3.

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

Morell, Mats, "Jordbruket i industrisamhället: 1870–1945," in Janken Myrdal and Mats Morell, eds., *Det svenska jordbrukets historia*, Vol. 4, Stockholm: Natur och Kultur/LT i samarbete med Nordiska museet och Stiftelsen Lagersberg, 2001.

Ngai, L Rachel and Barbara Petrongolo, "Gender gaps and the rise of the service economy," *American Economic Journal: Macroeconomics*, 2017, *9* (4), 1–44.

Ngai, Rachel, Claudia Olivetti, and Barbara Petrongolo, "Structural Transformation over 150 years of Women's and Men's Work," *Unpublished Working Paper*, 2022.

Nordström, Ester Blenda, *En piga bland pigor*, Stockholm: Wahlström & Widstrand, 1914.

Olivetti, Claudia, "The female labor force and long-run development: the American experience in comparative perspective," Technical Report, National Bureau of Economic Research 2013.

_ , "The Female Labor Force and Long-Run Development: The American Experience in Comparative Perspective," in "Human Capital in History: The American Record," University of Chicago Press, 2014, pp. 161–197.

_ **and M Daniele Paserman**, "In the name of the son (and the daughter): Intergenerational mobility in the United States, 1850–1940," *American Economic Review*, 2015, *105* (8), 2695–2724.

_ **, Eleonora Patacchini, and Yves Zenou**, "Mothers, peers, and gender-role identity," *Journal of the European Economic Association*, 2020, *18* (1), 266–301.

Oster, Emily, "Unobservable selection and coefficient stability: Theory and evidence," *Journal of Business & Economic Statistics*, 2019, *37* (2), 187–204.

Pooley, Colin and Jean Turnbull, *Migration and mobility in Britain since the eighteenth century*, Routledge, 2005.

Rosés, Joan Ramón and Nikolaus Wolf, "Regional economic development in Europe, 1900–2010: A description of the patterns," in "The Economic Development of Europe's Regions," Routledge, 2018, pp. 3–41.

Stanfors, Maria, "Women in a changing economy: the misleading tale of participation rates in a historical perspective," *The History of the Family*, 2014, *19* (4), 513–536.

Sundvall, Samuel, Christer Lundh, Martin Dribe, and Glenn Sandström, "Models of leaving home: patterns and trends in Sweden, 1830–1959," *The History of the Family*, 2023, *28* (3), 601–629.

Tilly, Louise A and Joan W Scott, *Women, work and family*, Holt, Rinehart and Winston, 1978.

van Leeuwen, Marco HD and Ineke Maas, *HISCLASS: A historical international social class scheme*, Universitaire Pers Leuven, 2011.

_ , _ , and Andrew Miles, "HISCO: Historical international standard classification of occupations," *Leuven UP*, 2002.

Vikström, Lotta, "Gendered routes and courses: the socio-spatial mobility of migrants in nineteenth-century Sundsvall, Sweden." PhD dissertation, Umeå universitet 2003.

Ward, Zachary, "Internal migration, education, and intergenerational mobility: Evidence from American history," *Journal of Human Resources*, 2020, pp. 0619–10265R2.

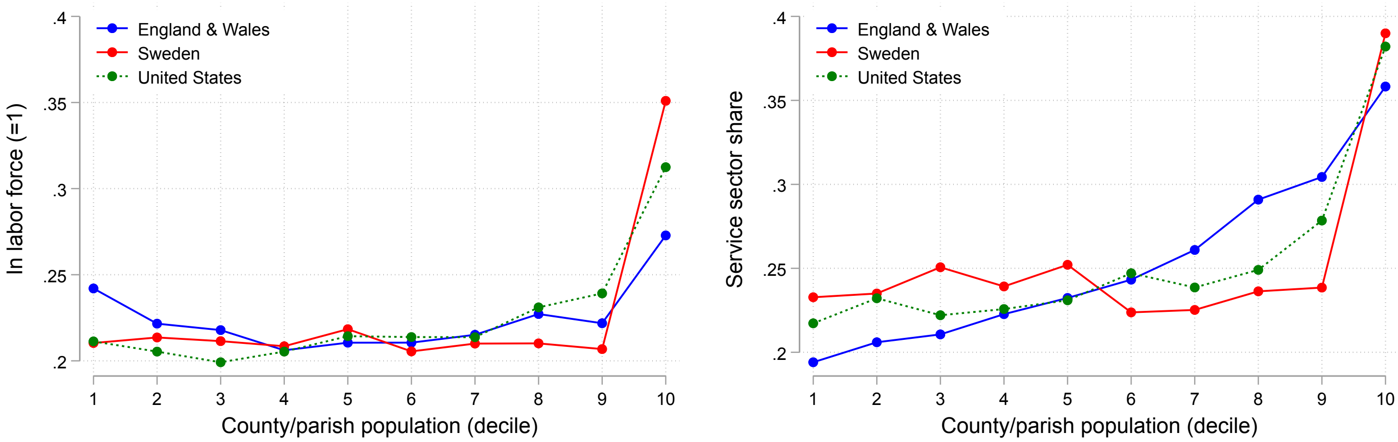
Wisselgren, Maria J., Sören Edvinsson, Mats Berggren, and Maria Larsson, "Testing Methods of Record Linkage on Swedish Censuses," *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 2014, *47* (3), 138–151.

Withrow, Jennifer, ""The Farm Woman's Problems"': Farm Crisis in the U.S. South and Migration to the City, 1920-1940," *Mimeo*, 2021.

Woods, Robert, "Urban-rural mortality differentials: an unresolved debate," *Population and development review*, 2003, *29* (1), 29–46.

Young, Alwyn, "Inequality, the urban-rural gap, and migration," *The Quarterly Journal of Economics*, 2013, *128* (4), 1727–1785.

Figures

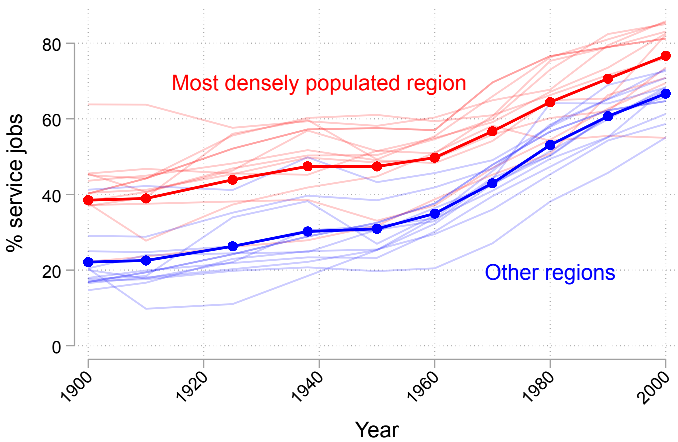


(A) IN LABOR FORCE

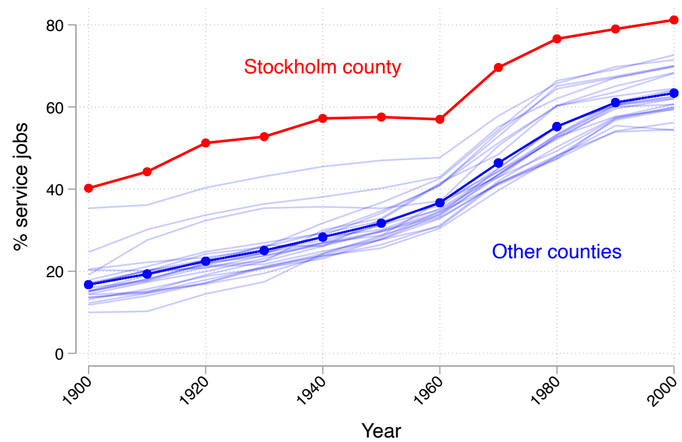
(B) SERVICE JOBS

FIGURE 1: LABOR FORCE PARTICIPATION AND THE SERVICE SECTOR IN THE EARLY 20^{th} CENTURY.

Notes: This figure shows the relationship between the population decile rank of a parish or county and its female labor force participation (Panel A) and share employed in services (Panel B). Variables on the Y-axis are residuals after absorbing country fixed effects. Census data for England and Wales are from 1911, and from 1910 for Sweden and the United States.



(A) European Regions



(B) Swedish Regions

Figure 2: Service sector size in European regions, 1900–2000.

Notes: This figure displays the evolution of the employment share of services in European regions 1900–2000. Panel A separately shows the outcome for the most densely populated NUTS-2 region within each country in red and the average share in other regions in blue. Countries included are Austria, Belgium, Denmark, Finland, France, Germany, Italy, Norway, Portugal, Spain, Sweden, and the UK. The cross-country averages for the densest and less dense regions are highlighted in bold lines. Data on the employment share of services is drawn from Rosés and Wolf (2018). Panel B displays the employment share of services across Swedish counties. Stockholm county, including the capital, is highlighted in bold red. The average for all other counties is highlighted in bold blue. Data on the employment shares for Swedish (NUTS-3) regions is drawn from Enflo et al. (2014).

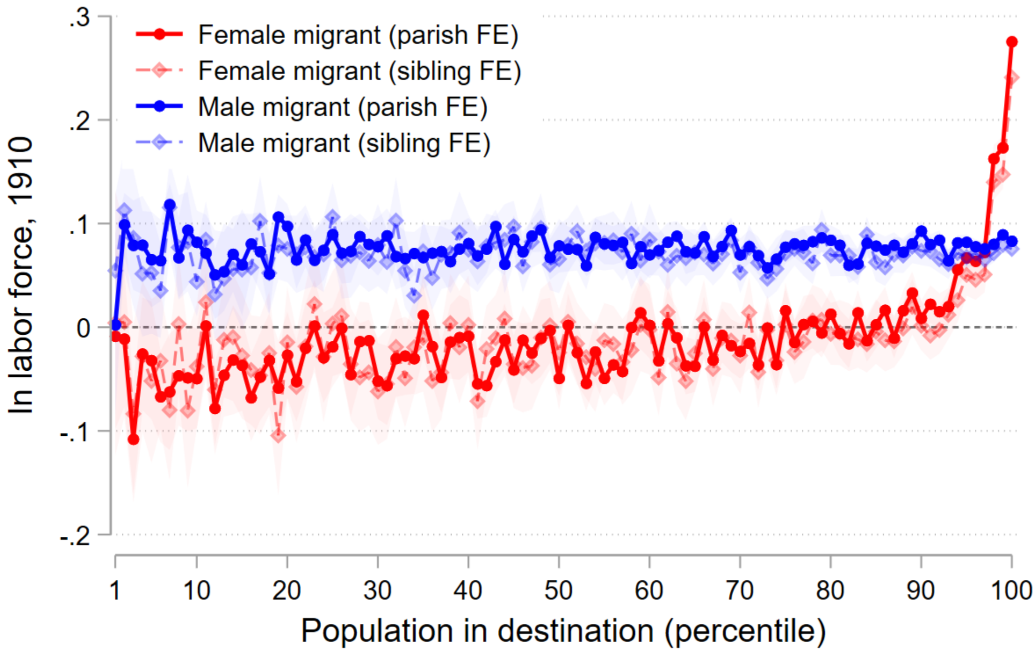
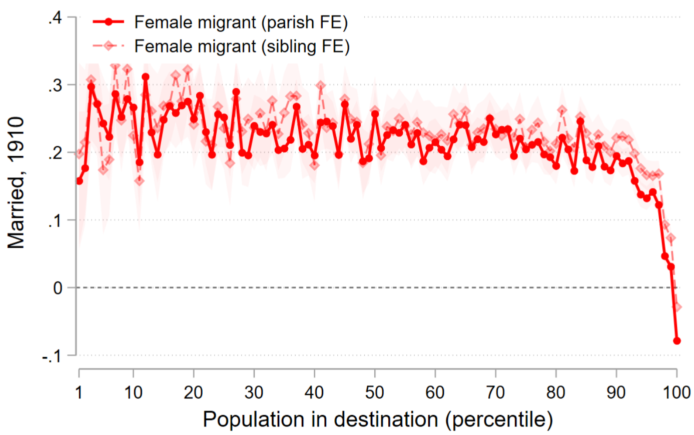
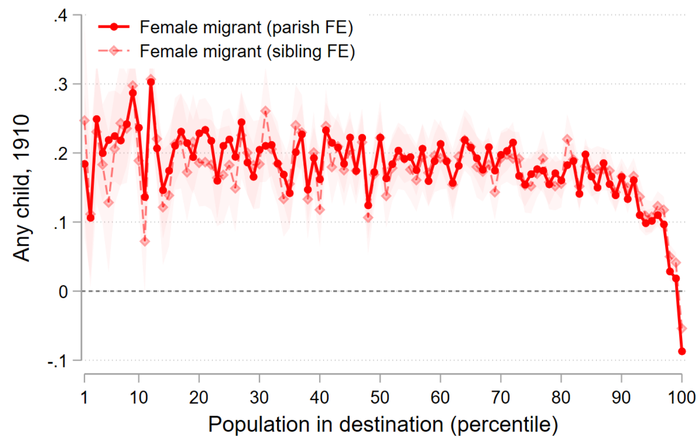


FIGURE 3: EMPLOYMENT BY POPULATION IN MIGRANT DESTINATION

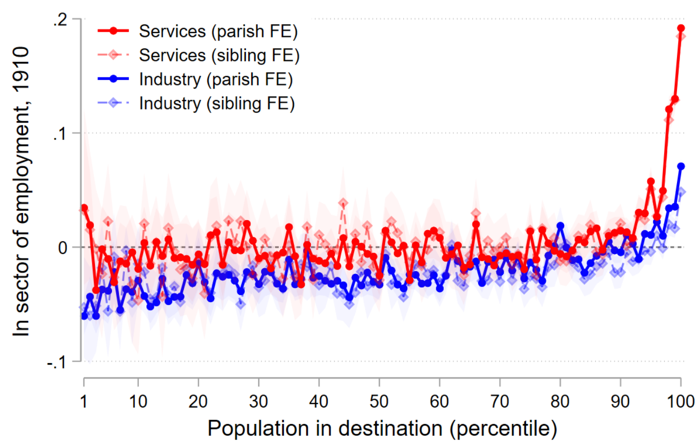
Notes: This figure displays OLS estimates of Equation (1) where the outcome is an indicator for an individual being part of the labor force in 1910. The figure plots point estimates and 95 percent confidence intervals for the returns to migration to parishes across the population percentile distribution. The sample consists of individuals born in rural parishes who have either migrated by 1910 or remain in the parish of origin (the omitted category). Estimates for females are reported in red, and for males in blue. Solid lines denote a specification using origin parish fixed effects as well as household and individual controls, while shaded lines denote estimates that include sibling fixed effects. See Section 4 for full list of control variables. Standard errors are clustered at the family level.



(A) MARRIAGE



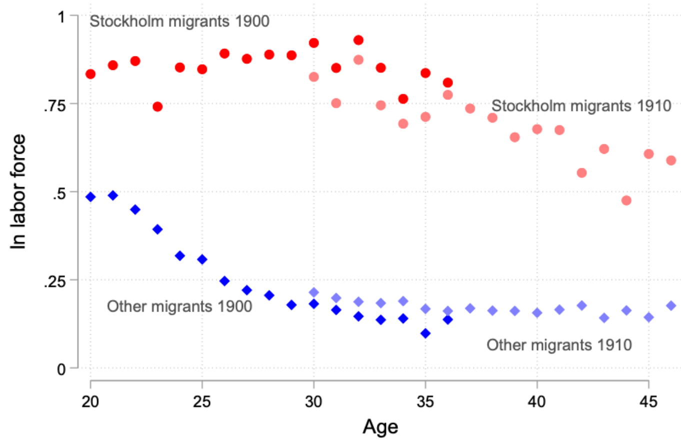
(B) ANY CHILD



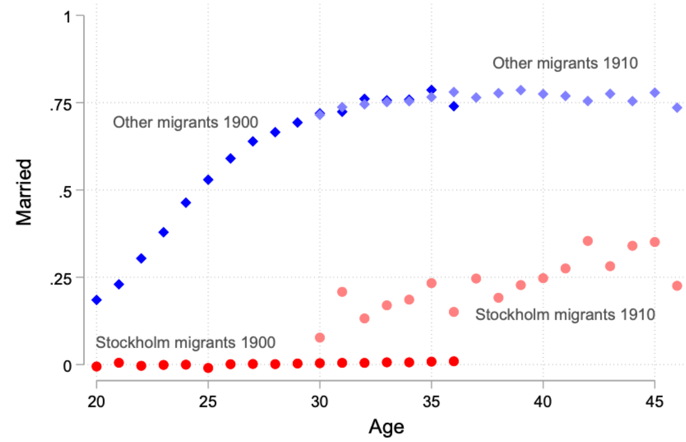
(C) SECTOR OF EMPLOYMENT

FIGURE 4: FAMILY FORMATION AND SECTOR OF EMPLOYMENT BY POPULATION IN MIGRANT DESTINATION

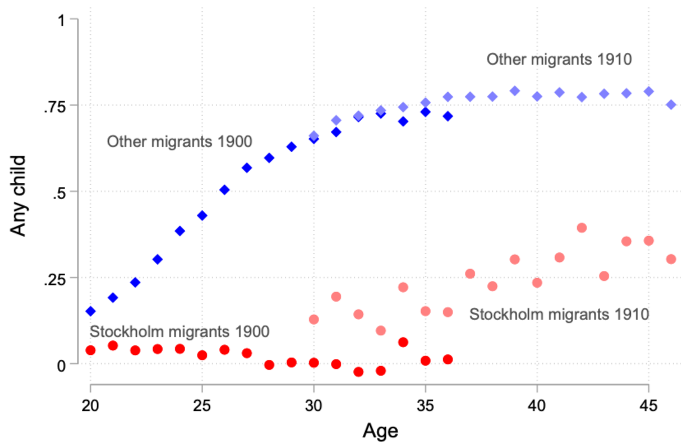
Notes: This figure displays OLS estimates of Equation (1) where the outcome is an indicator for being married (Panel A), for having any child (Panel B), or being employed in the service/industrial sector (Panel C) in 1910. The figure plots point estimates and 95 percent confidence intervals that capture the returns to migrating to different destinations ranked by their population size. Solid lines denote a specification using origin parish fixed effects as well as household and individual controls, while shaded lines denote estimates that include sibling fixed effects. See Section 4 for full list of control variables. Standard errors are clustered at the family level.



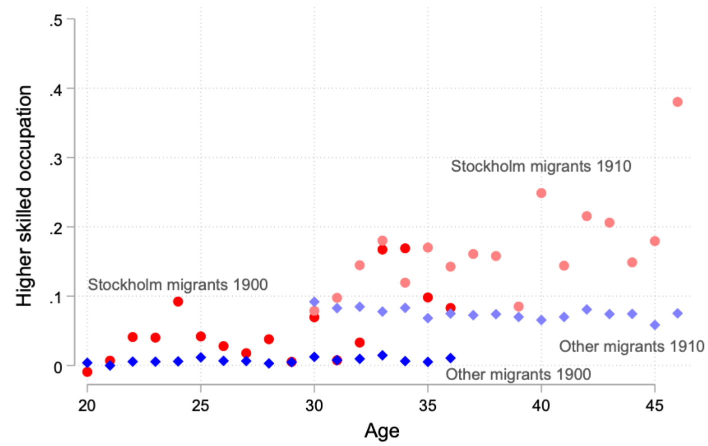
(A) IN LABOR FORCE



(B) MARRIAGE



(C) ANY CHILD



(D) HIGHER SKILL

FIGURE 5: EMPLOYMENT, FAMILY FORMATION, AND SKILLS OVER THE LIFE CYCLE

Notes: This figure displays binned scatter plots for the outcomes of migrants to Stockholm and rural areas between ages 20 to 46. *Higher skill* is an indicator for having an occupation with either medium or high skill according to HISCLASS. The graphs combine data on migrants in 1900 (when the sample is aged 20–36) and 1910 (when the sample is aged 30–46). Dark red and blue markers indicate the outcomes of migrants to Stockholm and rural areas in 1900, respectively. Light red and blue markers refer to outcomes in 1910. All estimates include for sibling fixed effects and individual controls following the method of Cattaneo et al. (2024). Section 4 for full list of control variables.

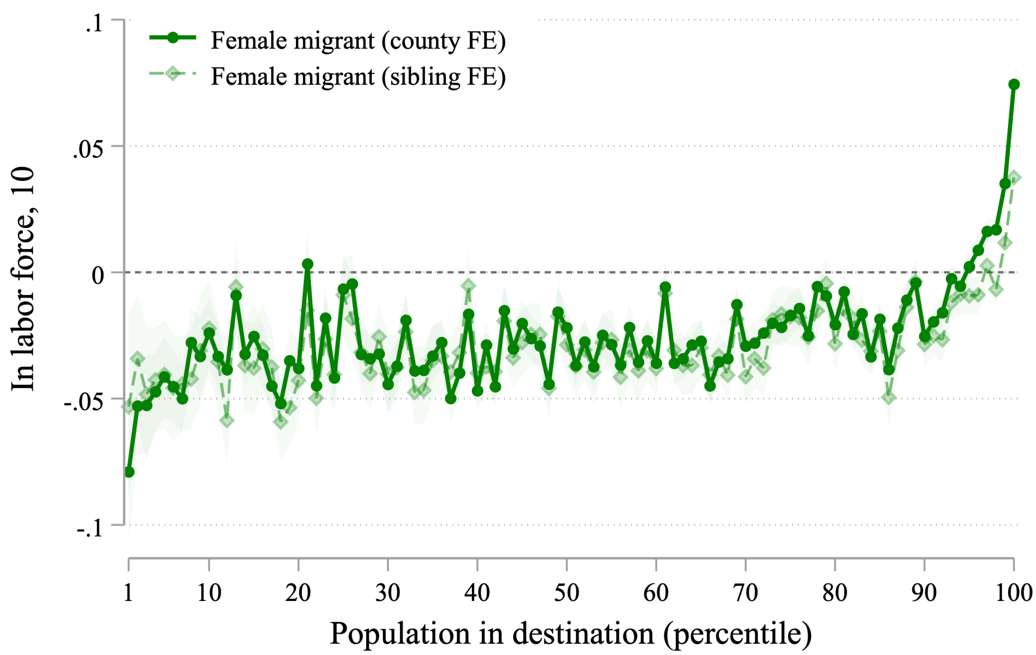


FIGURE 6: US FEMALE EMPLOYMENT BY POPULATION IN MIGRANT DESTINATION

Notes: This figure displays OLS estimates of Equation (1) where the outcome is an indicator for an individual being part of the labor force in 1910. The figure plots point estimates and 95 percent confidence intervals for the returns to migration to parishes across the population percentile distribution. The sample consists of US women individuals born in rural counties who have either migrated by 1910 or remain in the county of origin (the omitted category). Estimates for females are reported in red, and for males in blue. Solid lines denote a specification using origin county fixed effects as well as individual controls, while shaded lines denote estimates that include sibling fixed effects. Individual controls include age and indicators for being the eldest sister, literacy in 1880, and school attendance in 1880. Standard errors are clustered at the family level.

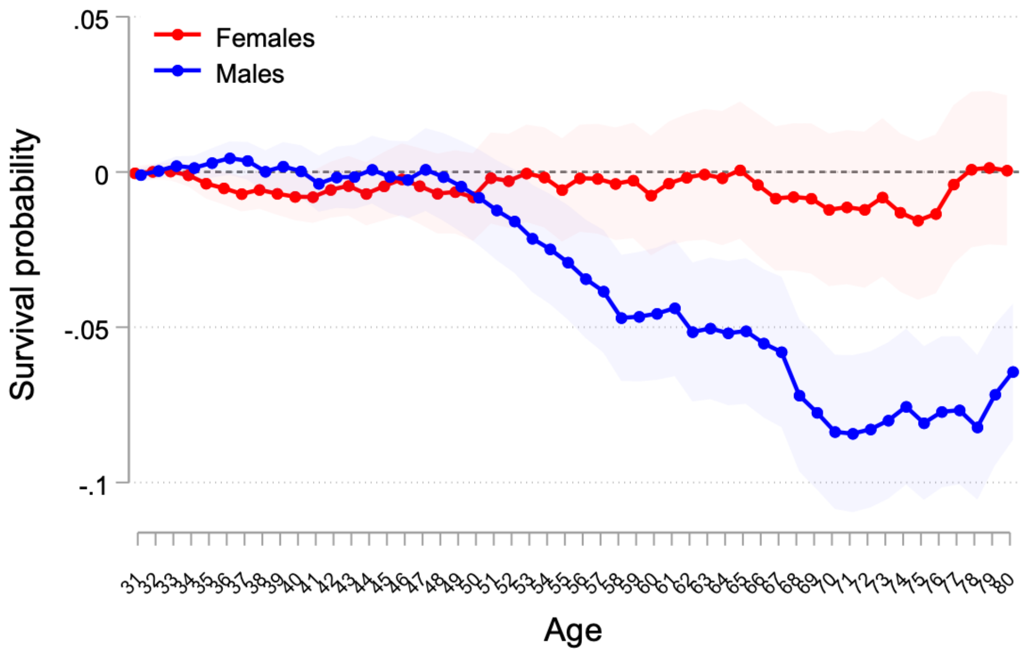


FIGURE 7: SURVIVAL PROBABILITY IN STOCKHOLM COMPARED TO RURAL AREAS

Notes: This figure displays regressions coefficients for likelihood of surviving past ages 31 to 80 in Stockholm relative to rural areas. Each coefficient is from a separate regression with siblings fixed effects and individual controls. See Section 4 for full list of control variables. Standard errors are clustered at the family level. Data on death age are obtained from the Death Index.

Tables

TABLE 1: FEMALE EMPLOYMENT BY MIGRANT DESTINATION

Dependent variable:	In labor force (=1)						
						HHH not farmer	Conservative definition
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Migrant: Stockholm (=1)	0.495***	0.534***	0.544***	0.545***	0.497***	0.448***	0.229***
	(0.005)	(0.005)	(0.005)	(0.007)	(0.010)	(0.012)	(0.011)
Migrant: other urban area (=1)		0.162***	0.157***	0.165***	0.133***	0.089***	-0.094***
		(0.002)	(0.003)	(0.004)	(0.005)	(0.007)	(0.006)
Individual controls	No	No	Yes	Yes	Yes	Yes	Yes
Family 1880 controls	No	No	Yes	Yes	No	No	No
Sibling fixed effects	No	No	No	No	Yes	Yes	Yes
Observations	185228	185228	185228	77597	77597	48619	77597
Mean outcome	0.208	0.208	0.208	0.228	0.228	0.311	0.492

Notes: OLS regressions. *In labor force* is an indicator variable taking value one if the individual is in the labor force in 1910. *Migrant: Stockholm* is an indicator taking value one if the individual lives in Stockholm city in 1910. *Migrant: other urban area* is an indicator taking value one if the individual lives in an urban area other than Stockholm in 1910. The sample in columns 1–3 consists of rural-born women who were aged 0–16 in 1880 and have left their parish of origin by 1910. Columns 4–5 restrict the sample to women who have at least one sister. Column 6 further excludes women living with a household head (either father or husband) who is a farmer. *Conservative definition* defines women living with a farmer household head (either father or husband) as being in the labor force. *Individual controls* include fixed effects for birth year, birth order, and an indicator for being the eldest sister. *Family 1880 controls* include origin parish fixed effects and a number of characteristics of the 1880 household: father's income score percentile, family size, as well as indicators for mother's employment, father's major 1-digit HISCO group, the presence of servants in the household, the household being is a married or cohabitating couple with children, a single-parent family, an extended family (relatives only), or composite family (with non-relatives), or a multigenerational family. *Sibling fixed effects* is a fixed effect for same-sex siblings. Standard errors, in parentheses, are clustered at the family level. *** - $p < 0.01$, ** - $p < 0.05$, * - $p < 0.1$.

TABLE 2: FAMILY FORMATION AND EMPLOYMENT BY MIGRANT DESTINATION

Dependent variable:	Married (=1)	Any child (=1)	In labor force (=1)			
			Married 1910	Unmarried 1910	Unmarried 1900 pre-migration	Unmarried 1900 post-migration
	(1)	(2)	(3)	(4)	(5)	(6)
Migrant: Stockholm (=1)	-0.509***	-0.472***	0.031***	0.191***	0.548***	0.367***
	(0.010)	(0.010)	(0.009)	(0.019)	(0.057)	(0.042)
Migrant: other urban area (=1)	-0.121***	-0.104***	0.016***	0.123***	0.161***	0.163***
	(0.006)	(0.006)	(0.003)	(0.016)	(0.034)	(0.030)
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes
Sibling fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	77597	77597	44915	11192	3036	7144
Mean outcome	0.710	0.707	0.017	0.745	0.303	0.430

Notes: OLS regressions. All outcomes are measured in 1910. Columns 1 and 2 use the main sample. Columns 3 and 4 restricts the sample to women that are married and unmarried in 1910, respectively. Column 5 restricts the sample to women that are observed as unmarried and living in their childhood parish in 1900. Column 6 restricts the sample to women that are unmarried and living in their destination in both 1900 and 1910. *Migrant: Stockholm* is an indicator taking value one if the individual lives in Stockholm city in 1910. *Migrant: other urban area* is an indicator taking value one if the individual lives in an urban area other than Stockholm in 1910. *Individual controls* include fixed effects for birth year, birth order, and an indicator for being the eldest sister. *Sibling fixed effects* is a fixed effect for same-sex siblings. Standard errors, in parentheses, are clustered at the family level. *** - $p < 0.01$, ** - $p < 0.05$, * - $p < 0.1$.

TABLE 3: PERSISTENCE IN OUTCOMES AFTER MIGRATION

Dependent variable:	In labor force (=1)		Married (=1)		Any child (=1)		Higher skill (=1)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Migrant in 1900: Stockholm (=1)	0.342***		-0.375***		-0.377***		0.078***	
	(0.016)		(0.016)		(0.015)		(0.011)	
Migrant in 1900: other urban area (=1)	0.091***		-0.092***		-0.091***		0.017***	
	(0.008)		(0.008)		(0.008)		(0.005)	
Ever migrated: Stockholm (=1)		0.363***		-0.096***		-0.371***		0.061***
		(0.009)		(0.014)		(0.009)		(0.006)
Ever migrated: other urban area (=1)		0.099***		-0.094***		-0.082***		0.022***
		(0.005)		(0.005)		(0.005)		(0.003)
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sibling fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	42691	88000	42691	88000	42691	88000	42691	88000
Mean outcome	0.218	0.222	0.712	0.707	0.714	0.706	0.080	0.083

Notes: OLS regressions. All outcomes are measured in 1910. Odd numbered columns display results for 1900 migrants in 1910. Even numbered columns focus on individuals who have ever migrated to Stockholm or another urban area. *Migrant in 1900: Stockholm* is an indicator taking value one if the individual lives in Stockholm city in 1900. *Migrant in 1900: other urban area* is an indicator taking value one if the individual lives in an urban area other than Stockholm in 1900. *Ever migrated: Stockholm* is an indicator taking value one if the individual was ever observed living in Stockholm in the 1890, 1900, or 1910 censuses. *Ever migrated: other urban area* is an indicator taking value one if the individual was ever observed living in another urban area than Stockholm in 1890, 1900, or 1910. *Higher skill* is an indicator for having an occupation with either medium or high skill according to HISCLASS. *Individual controls* include fixed effects for birth year, birth order, and an indicator for being the eldest sister. *Sibling fixed effects* is a fixed effect for same-sex siblings. Standard errors, in parentheses, are clustered at the family level. *** - $p < 0.01$, ** - $p < 0.05$, * - $p < 0.1$.

TABLE 4: SECTOR, SKILL, AND INCOME BY MIGRANT DESTINATION

Dependent variable:	Sector of employment		Occupational skill			Occ. income score		
	Services	Industry	High-skill	Low-skill	Unskilled	ln(H. income)	ln(Income)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Migrant: Stockholm (=1)	0.403***	0.078***	0.070***	0.387***	0.038***	0.206***	0.074***	0.222***
	(0.009)	(0.006)	(0.008)	(0.010)	(0.004)	(0.011)	(0.023)	(0.012)
Migrant: other urban area (=1)	0.094***	0.046***	0.023***	0.093***	0.013***	-0.077***	-0.181***	-0.170***
	(0.004)	(0.003)	(0.004)	(0.004)	(0.002)	(0.006)	(0.018)	(0.007)
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sibling fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Occupation fixed effects	No	No	No	No	No	No	No	Yes
Observations	77597	77597	77597	77597	77597	67216	8134	8019
Mean outcome	0.115	0.061	0.085	0.134	0.019	6.847	7.002	7.000

Notes: OLS regressions. All outcomes are measured in 1910. *Migrant: Stockholm* is an indicator taking value 1 if the individual lives in Stockholm city in 1910, and 0 if not. *Migrant: other urban area* is an indicator taking value 1 if the individual lives in an urban area other than Stockholm in 1910, and 0 if not. *Individual controls* include fixed effects for birth year and birth order, and an indicator for eldest sister. *Sibling fixed effects* is a fixed effect for same sex siblings. *Occupation fixed effects* is a fixed effect for the occupational code. Standard errors, in parentheses, are clustered at the 1880 household level. *** - $p < 0.01$, ** - $p < 0.05$, * - $p < 0.1$.

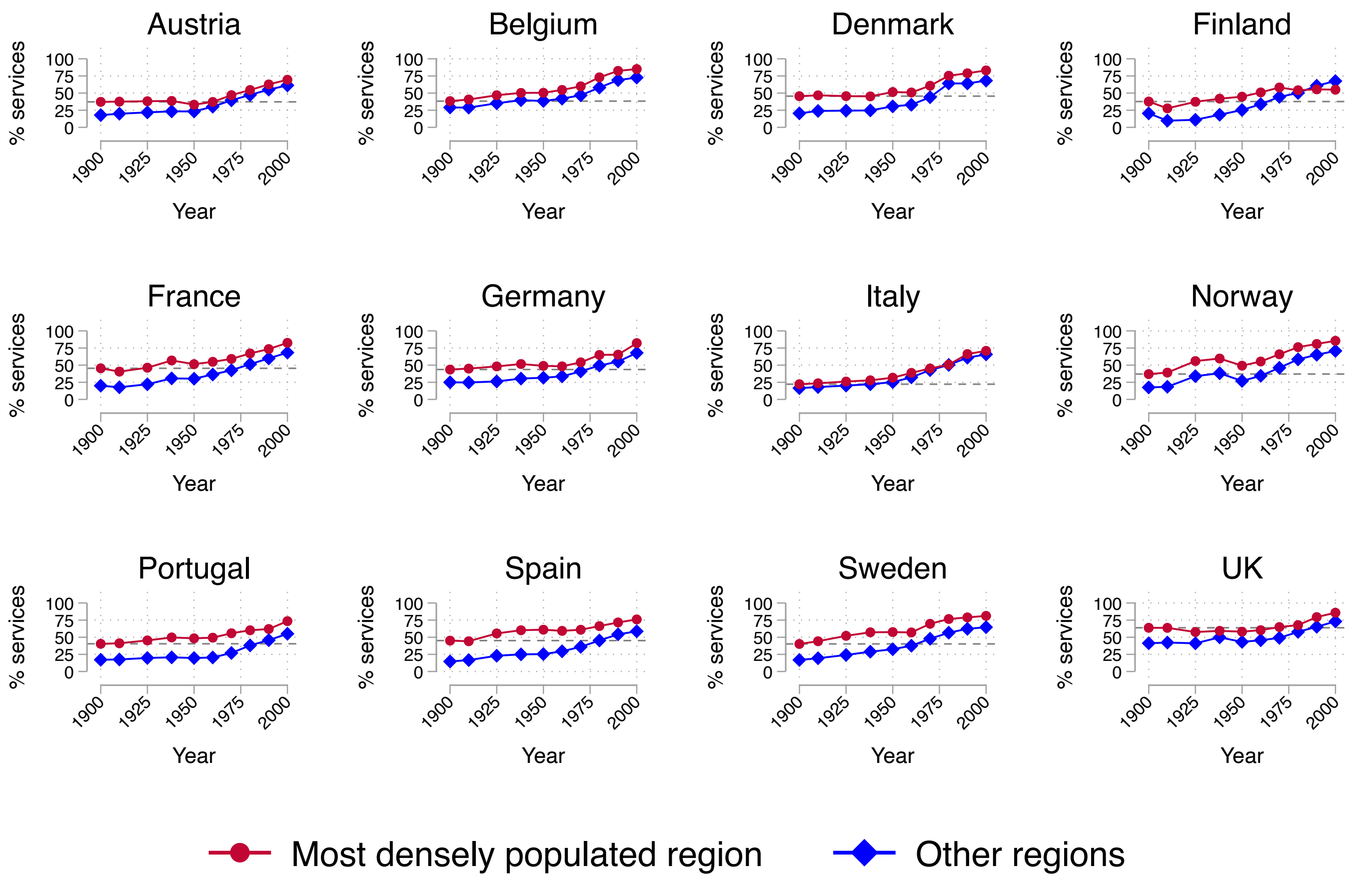
TABLE 5: THE ROLE OF SERVICES AND LABOR MARKET COMPOSITION

Dependent variable:	In labor force (=1)				Married (=1)	Any child (=1)
	(1)	(2)	(3)	(4)	(5)	(6)
Migrant: Stockholm (=1)		0.294***	0.284***	0.263***	-0.250***	-0.293***
		(0.017)	(0.018)	(0.027)	(0.029)	(0.029)
Migrant: other urban area (=1)		0.035***	0.034***	0.033***	-0.008	-0.002
		(0.008)	(0.009)	(0.011)	(0.011)	(0.012)
Destination service share	0.387***	0.337***				
	(0.028)	(0.034)				
Destination industry share	-0.149***	-0.007				
	(0.015)	(0.016)				
Destination workers (log)	0.063***	0.024***	0.023***	0.023***	-0.020***	-0.011***
	(0.002)	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes
Sibling fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Cubic sector shares	No	No	Yes	No	No	No
Sector share percentile FE	No	No	No	Yes	Yes	Yes
Observations	77597	77597	77597	77597	77597	77597
Mean outcome	0.228	0.228	0.228	0.228	0.710	0.707

Notes: OLS regressions. All outcomes are measured in 1910. *Migrant: Stockholm* is an indicator taking value 1 if the individual lives in Stockholm city in 1910, and 0 if not. *Migrant: other urban area* is an indicator taking value 1 if the individual lives in an urban area other than Stockholm in 1910, and 0 if not. *Individual controls* include fixed effects for birth year and birth order, and an indicator for eldest sister. *Sibling fixed effects* is a fixed effect for same sex siblings. *Cubic sector polynomial* includes cubic polynomials in the share working in services and in industry. Standard errors, in parentheses, are clustered at the 1880 household level. *** - $p < 0.01$, ** - $p < 0.05$, * - $p < 0.1$.

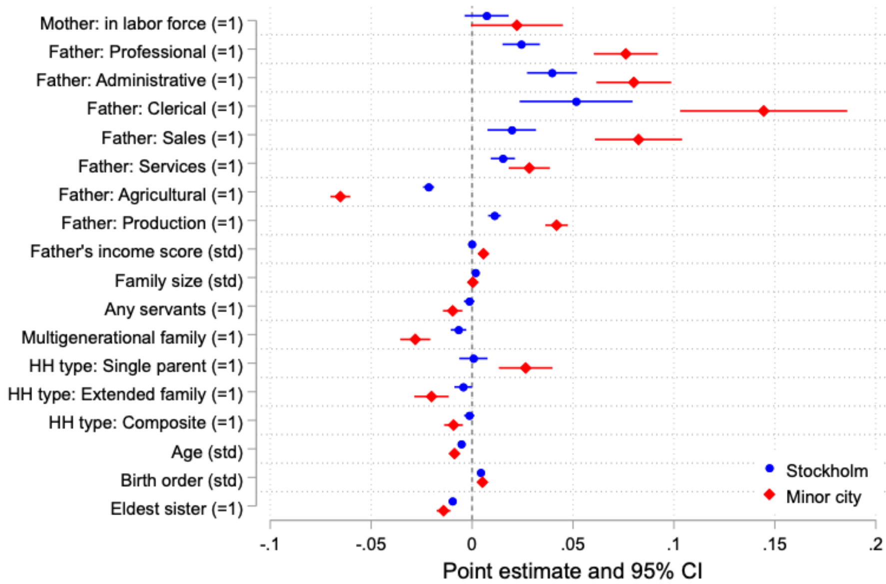
Online Appendix – not for publication

A Robustness and Additional Material

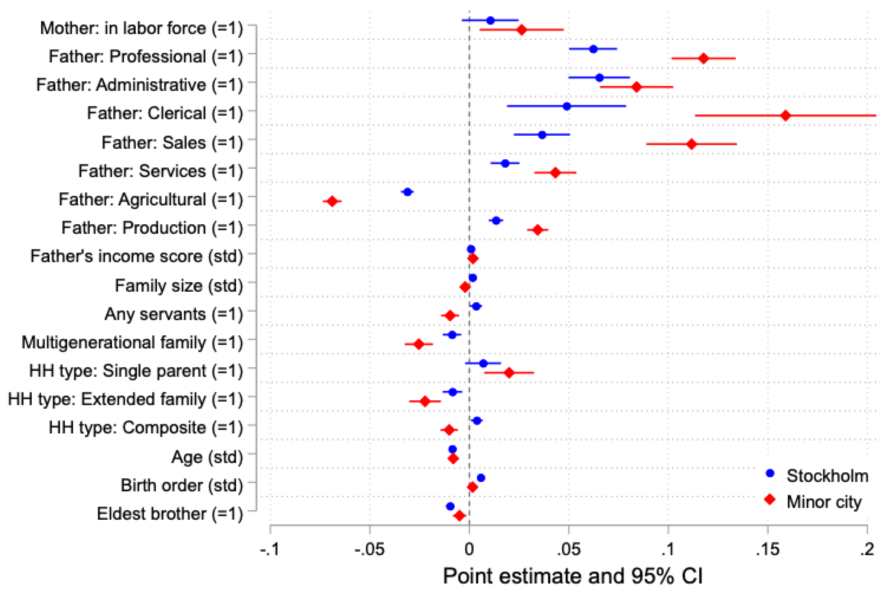


Notes: This figure shows that the most densely populated region in all European countries had a higher share of employment in the service sector compared to other regions. Data on the employment share of services is drawn from Rosés and Wolf (2018). For each country, we report the employment share in services in the most densely populated region and the unweighted average across all other regions in each country. A horizontal dashed line denotes the share of employment in services in the most densely populated region in 1900.

FIGURE A.1: SERVICE JOBS IN EUROPEAN (NUTS-2) REGIONS, 1900–2000.



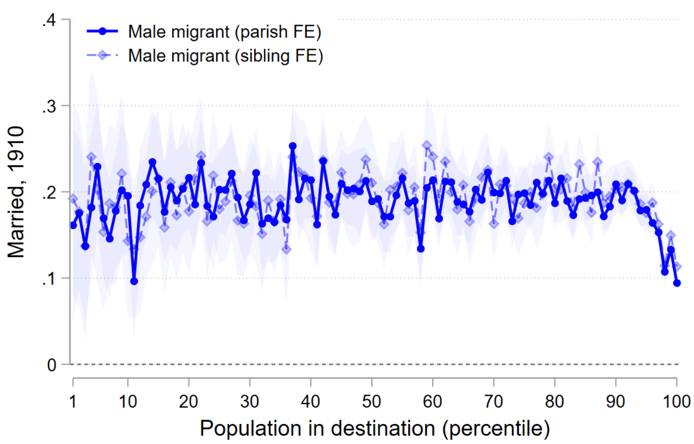
(A) WOMEN



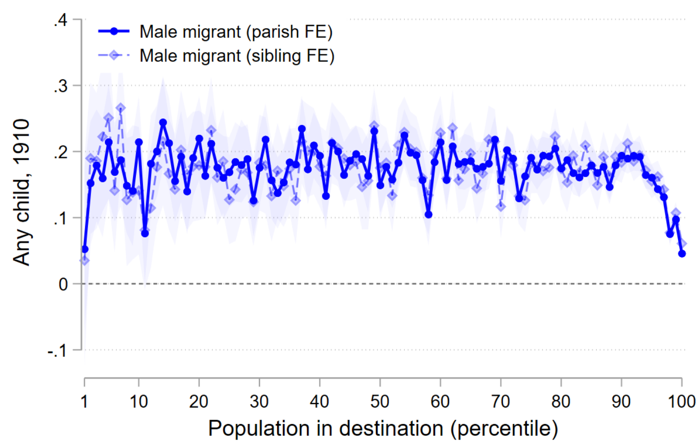
(B) MEN

FIGURE A.2: SELECTION INTO MIGRATION

Notes: This figure displays OLS estimates and 95 percent confidence intervals from separate regressions where the outcome is an indicator for migration to Stockholm (in blue) or to a minor city (in red), respectively, on different observables in the 1880 census. The omitted category is migrating to a rural area. The sample consists of women born in rural areas and aged 0–16 in 1880 who left their parish of origin by 1910 and have at least one migrating sister. Variables denoted with "(std)" are standardized. All regressions include origin parish fixed effects. Standard errors clustered at the family level.



(A) MARRIAGE



(B) ANY CHILD

FIGURE A.3: FAMILY FORMATION AMONG MALE MIGRANTS BY POPULATION IN DESTINATION

Notes: This figure displays OLS coefficients from separate estimations of Equation (1) on the sample of male migrants, where the outcome is an indicator for being married (Panel A) and having any child (Panel B) in 1910. The figure plots point estimates and 95 percent confidence intervals that capture the returns to migrating to different destinations ranked by their population size. Solid lines denote a specification using origin parish fixed effects as well as household and individual controls, while shaded lines denote estimates that include sibling fixed effects. See Section 4 for full list of control variables. Standard errors are clustered at the family level.

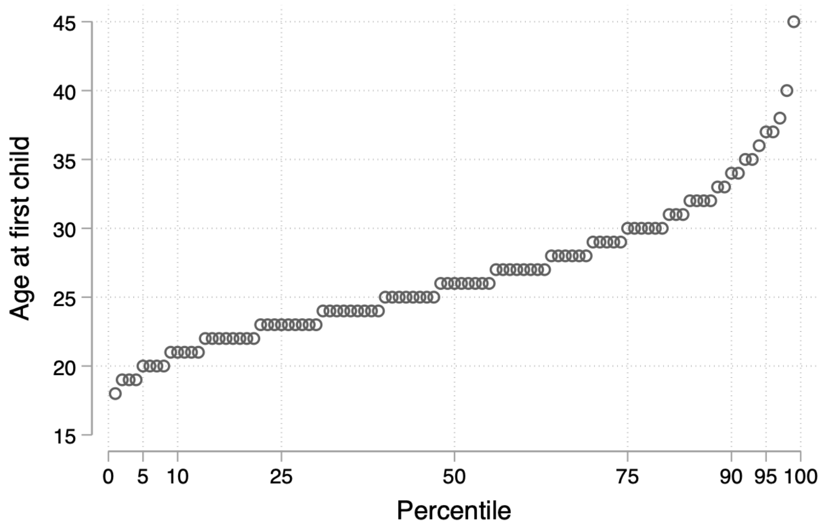
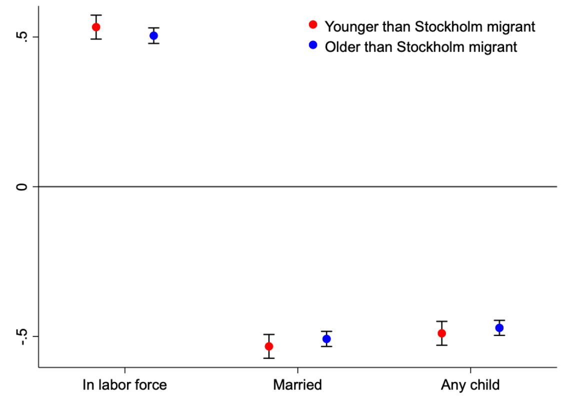
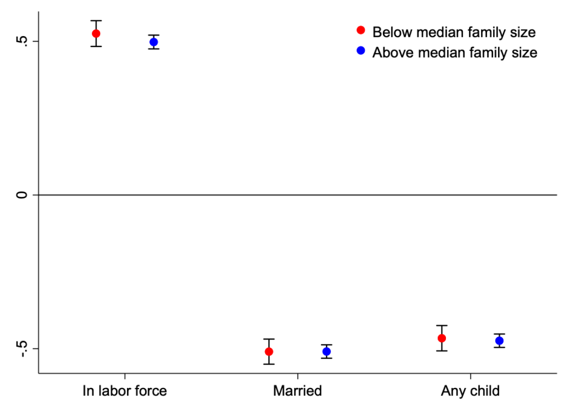


FIGURE A.4: AGE AT FIRST CHILD IN STOCKHOLM

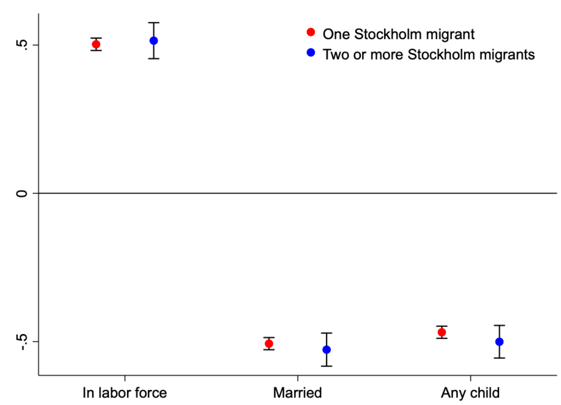
Notes: This figure displays the distribution of age at first child among women living in Stockholm in the 1910 census, based on women with a newborn child and no other own children in their household.



(A) Siblings younger/older than Stockholm migrant



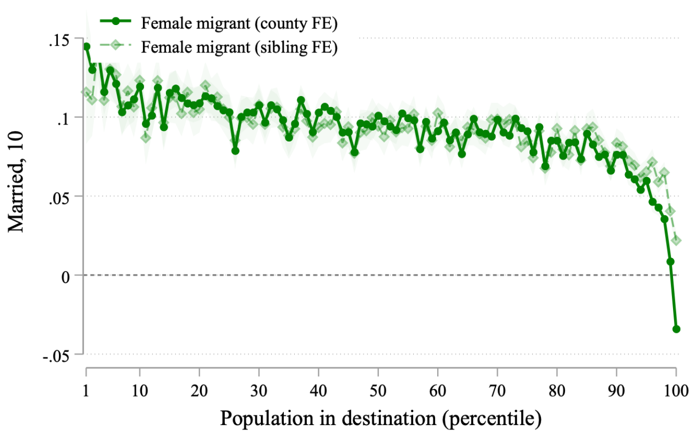
(B) Below/above median family size



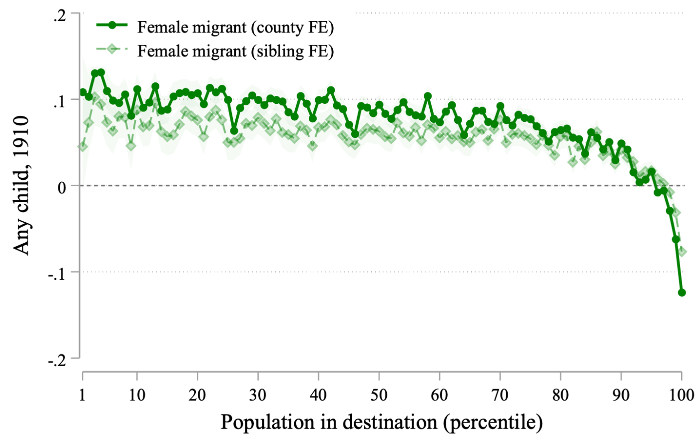
(C) One/more Stockholm migrants

Figure A.5: Splitting samples

Notes: This figure displays regression coefficients when splitting the main sample along three dimensions. In Panel A, coefficients in red indicate households in which non-Stockholm migrants in the family are all younger than the Stockholm migrant, whereas blue coefficients show the opposite scenario. In Panel B, red coefficients are estimated on the subsample of families that are smaller than the median in 1880, whereas blue coefficients are based on above median sized families. In Panel C, red coefficients are estimated when dropping families sending more than one migrant to Stockholm, whereas coefficients in blue are estimated when dropping families sending only one migrant. All outcomes are measured in to 1910. Specifications correspond to that of Column 5, Table 1. Standard errors clustered at the family level.



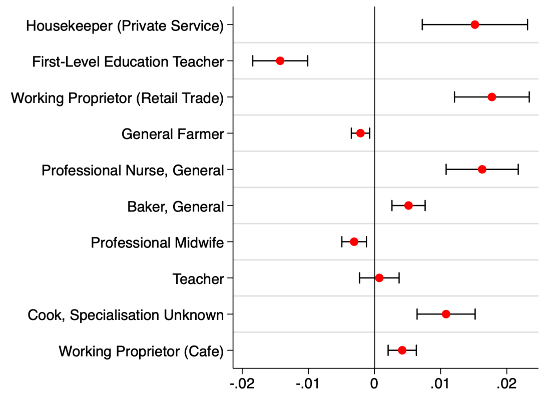
(A) MARRIED



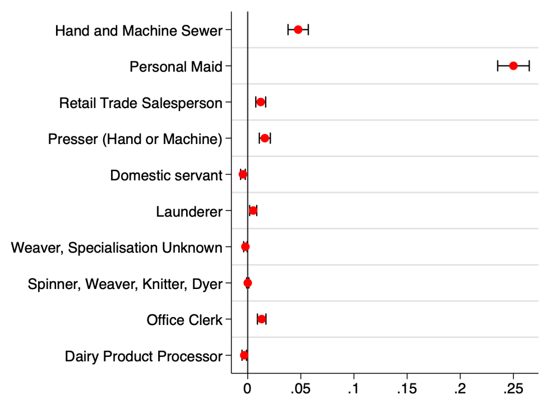
(B) ANY CHILD

FIGURE A.6: MARRIAGE AND CHILDBEARING FOR FEMALE US MIGRANTS

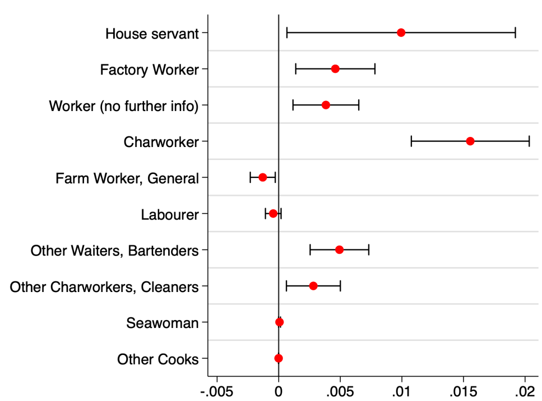
Notes: This figure displays OLS estimates of Equation (1) where the outcome is an indicator for an individual being part of the labor force in 1910. The figure plots point estimates and 95 percent confidence intervals for the returns to migration to parishes across the population percentile distribution. The sample consists of US women individuals born in rural counties who have either migrated by 1910 or remain in the county of origin (the omitted category). Dark green lines denote a specification using origin county fixed effects as well as individual controls, while light green lines denote estimates that include sibling fixed effects. Individual controls include age and indicators for being the eldest sister, literacy in 1880, and school attendance in 1880. Standard errors are clustered at the family level.



(A) HIGHER SKILLED



(B) LOWER SKILLED



(C) UNSKILLED

FIGURE A.7: FEMALE EMPLOYMENT IN THE MOST COMMON OCCUPATIONS

Notes: This figure displays OLS estimates and 95 percent confidence intervals from separate regressions where the outcome is an indicator variable equal to one if working in the denoted occupation, and zero otherwise. Occupations are defined using the full digit HISCO code. All outcomes refer to 1910. Red circles denote regression coefficients for living in Stockholm 1910. Specifications correspond to that of Column 5, Table 1. Standard errors clustered at the family level.

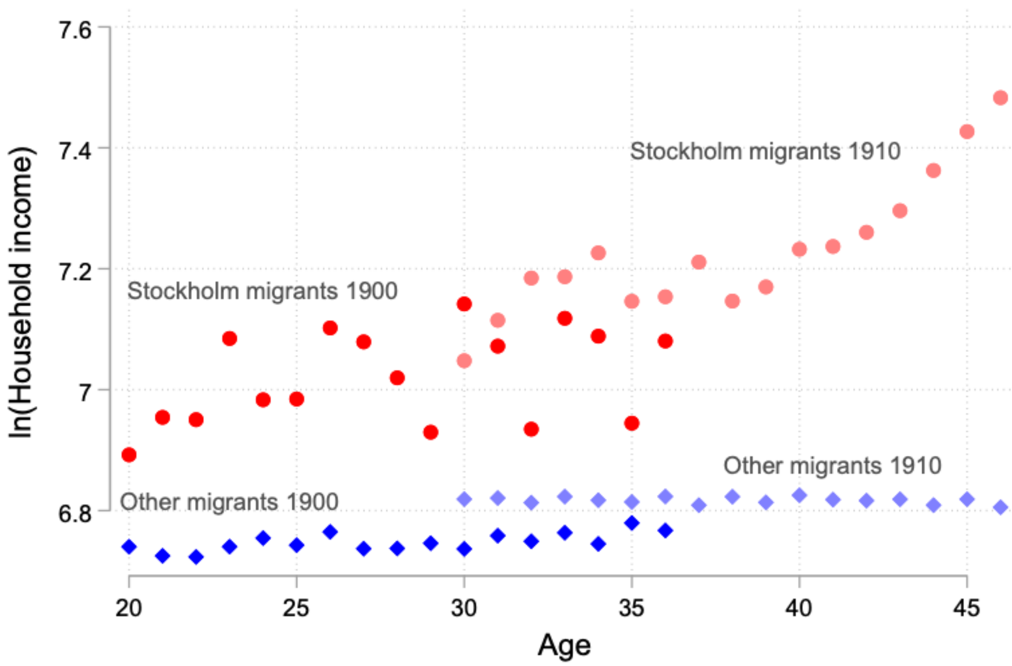


FIGURE A.8: HOUSEHOLD INCOME OVER THE LIFE CYCLE

Notes: This figure displays a binned scatter plot for the household income of migrants to Stockholm and rural areas between ages 20 to 46. The graph combines data on migrants in 1900 (when the sample is aged 20–36) and 1910 (when the sample is aged 30–46). Dark red and blue markers indicate the outcomes of migrants to Stockholm and rural areas in 1900, respectively. Light red and blue markers refer to outcomes in 1910. All estimates include for sibling fixed effects and individual controls following the method of Cattaneo et al. (2024). Section 4 for full list of control variables.

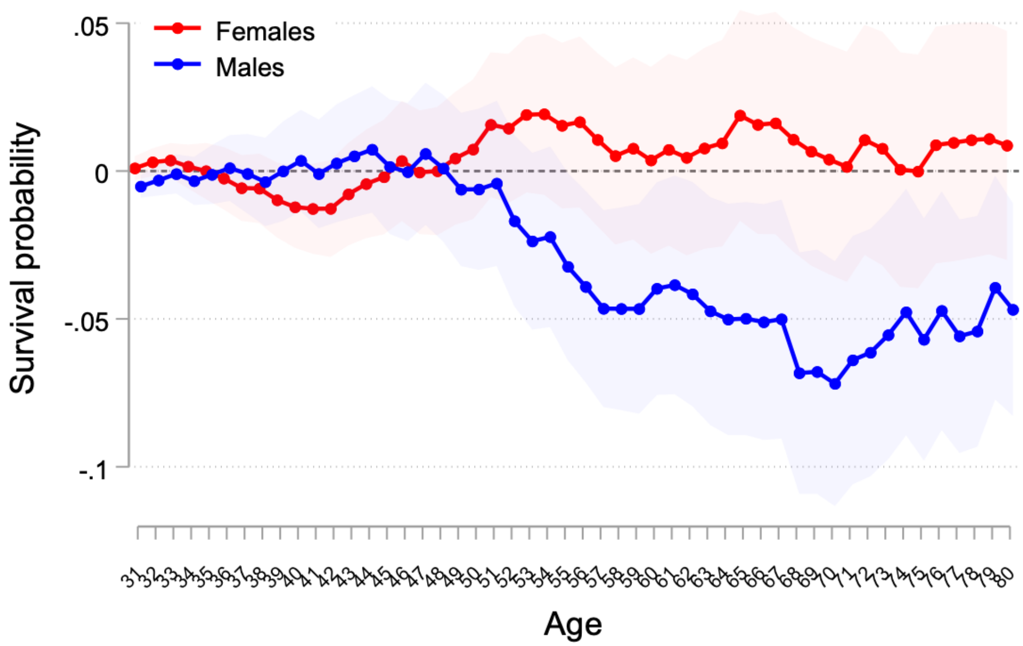
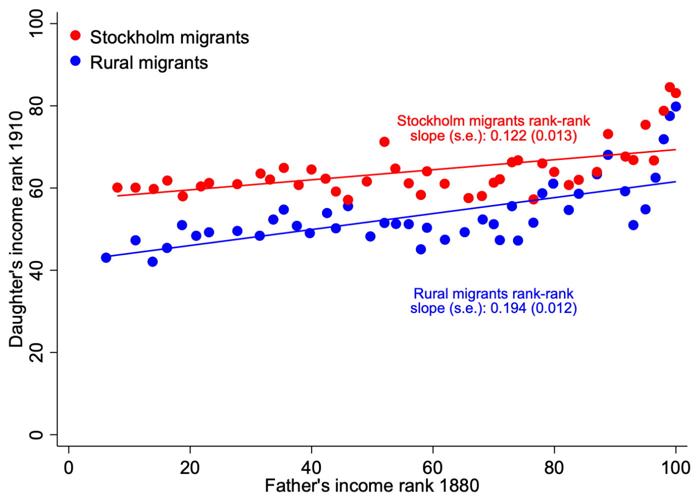
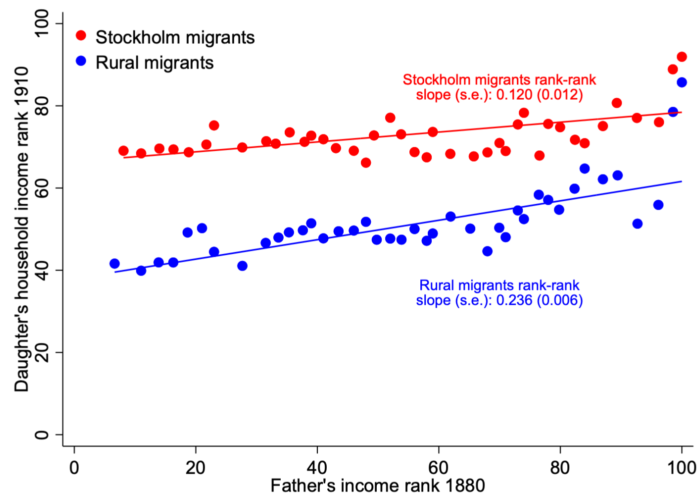


FIGURE A.9: SURVIVAL PROBABILITY IN STOCKHOLM COMPARED TO RURAL AREAS

Notes: This figure displays regressions coefficients for likelihood of surviving past ages 31 to 80 in Stockholm relative to rural areas. The sample consists of men and women who were between 30 and 37 years old in 1910. Each coefficient is from a separate regression with siblings fixed effects and individual controls. See Section 4 for full list of control variables. Standard errors are clustered at the family level. Data on death age are obtained from the Death Index.



(A) INCOME SCORE RANK



(B) HH INCOME SCORE RANK

FIGURE A.10: INTERGENERATIONAL INCOME MOBILITY

Notes: This figure displays a binned scatter plot of the association between father's and daughter's income ranks. Ranks are based on the average income for the occupation held by the father in 1880 and daughters in 1910. Red dots indicate the correlation among migrants to Stockholm. Blue dots indicate the correlation among rural migrants.

TABLE A.1: DESCRIPTIVE STATISTICS FOR PRIME-AGED (20–55) FEMALES IN 1910 CENSUS

	(1) Stockholm mean	(2) Other urban area mean	(3) Rural area mean	(4) Stockholm (migrant) mean	(5) Stockholm (native) mean
FLFP	0.51	0.37	0.18	0.51	0.50
Married	0.39	0.49	0.59	0.40	0.37
Any child	0.36	0.50	0.61	0.37	0.35
Age	34.73	35.05	35.92	35.57	33.25

Notes: Summary statistics for prime-aged (20–55) females in the 1910 population census for different categories: 1) Stockholm inhabitants; 2) Urban individuals (outside of Stockholm); 3) Rural individuals; 4) Migrants living in Stockholm; and 5) Stockholm natives.

TABLE A.2: DESCRIPTIVE STATISTICS FOR THE MOST COMMON FEMALE OCCUPATIONS

Stockholm:

	Freq.	Percent	Cum.
personal maid	21286	24.68	24.68
hand and machine sewer	7479	8.67	33.35
working proprietor	4455	5.17	38.52
house servant	4418	5.12	43.64
retail trade salesperson	4358	5.05	48.69
office clerk	3344	3.88	52.57
housekeeper	3013	3.49	56.07
worker	2863	3.32	59.39
charworker	2263	2.62	62.01
factory worker	1492	1.73	63.74
presser	1311	1.52	65.26
professional nurse	1305	1.51	66.77
cashier	1204	1.40	68.17
telephone switchboard operator	1191	1.38	69.55
teacher	1026	1.19	70.74
janitor	897	1.04	71.78
cook	815	0.95	72.72
launderer	800	0.93	73.65
bookbinder	780	0.90	74.56
cigar maker	770	0.89	75.45

Urban (excluding Stockholm):

	Freq.	Percent	Cum.
house servant	24184	16.88	16.88
hand and machine sewer	12345	8.62	25.50
personal maid	10480	7.32	32.82
housekeeper	7958	5.56	38.38
working proprietor	7786	5.44	43.81
factory worker	6785	4.74	48.55
retail trade salesperson	5814	4.06	52.61
worker	3078	2.15	54.76
general farmer	2800	1.95	56.71
first-level education teacher	2665	1.86	58.57
spinner	2519	1.76	60.33
office clerk	2422	1.69	62.02
charworker	2123	1.48	63.51
domestic servant	2122	1.48	64.99
launderer	2121	1.48	66.47
presser	2047	1.43	67.90
professional nurse	1933	1.35	69.25
weaver	1924	1.34	70.59
baker	1838	1.28	71.87
teacher	1818	1.27	73.14

Rural:

	Freq.	Percent	Cum.
house servant	83860	26.01	26.01
general farmer	46372	14.38	40.40
housekeeper	21538	6.68	47.08
small subsistence farmer	16931	5.25	52.33
hand and machine sewer	16429	5.10	57.42
farm worker	14500	4.50	61.92
first-level education teacher	10630	3.30	65.22
worker	10480	3.25	68.47
personal maid	8736	2.71	71.18
working proprietor	6206	1.93	73.10
domestic servant	5747	1.78	74.89
factory worker	4226	1.31	76.20
retail trade salesperson	3687	1.14	77.34
spinner	3151	0.98	78.32
teacher	2848	0.88	79.20
labourer	2678	0.83	80.03
weaver	2643	0.82	80.85
other military ranks	2553	0.79	81.65
professional nurse	2553	0.79	82.44
professional midwife	2355	0.73	83.17

Notes: The tables displays the frequency, percentage, and cumulative percentage of the most common female occupations in 1910 by location: Stockholm, urban, and rural.

TABLE A.3: SUMMARY STATISTICS OF MIGRANT SAMPLE

	All	Sibling sample			
Panel A: Women		All destinations	Stockholm	Other urban	Rural parish
Age 1880	7.243	7.148	6.644	6.977	7.257
Mother in labor force 1880	0.014	0.008	0.009	0.009	0.008
Father's income score 1880	7.483	7.498	7.523	7.522	7.487
Birthorder	2.717	3.103	3.213	3.132	3.083
Eldest sister	0.482	0.311	0.271	0.297	0.319
In labor force 1910	0.208	0.228	0.686	0.323	0.151
Married 1910	0.734	0.710	0.237	0.626	0.785
Any child 1910	0.728	0.707	0.253	0.627	0.778
Observations	185228	77597	5100	19011	53486

Panel B: Men					
Age 1880	7.101	7.008	6.408	6.821	7.157
Mother in labor force 1880	0.014	0.009	0.010	0.010	0.008
Father's income score 1880	7.480	7.494	7.525	7.535	7.475
Birthorder	2.702	3.067	3.199	3.094	3.039
Eldest brother	0.466	0.304	0.272	0.299	0.310
In labor force 1910	0.965	0.966	0.974	0.970	0.963
Married 1910	0.764	0.757	0.598	0.737	0.786
Any child 1910	0.700	0.694	0.492	0.662	0.734
Observations	160636	66070	6557	14600	44913

Notes: Summary statistics for all migrants and migrants with siblings, the latter also shown across our three different migrant categories: i) Stockholm migrant, ii) Other urban area migrant, and iii) Rural parish migrant. Panel A displays mean values for women and panel B displays mean values for men.

TABLE A.4: EMPLOYMENT, MARRIAGE, AND CHILDREN FOR MEN 1910

A. Dependent variable:	In labor force (=1)				
	(1)	(2)	(3)	(4)	(5)
Migrant: Stockholm (=1)	0.008***	0.009***	0.009***	0.010***	0.009**
	(0.001)	(0.001)	(0.002)	(0.002)	(0.004)
Migrant: other urban area (=1)		0.004***	0.005***	0.007***	0.001
		(0.001)	(0.001)	(0.002)	(0.003)
Observations	160636	160636	160634	66070	66070
Mean outcome	0.965	0.965	0.965	0.966	0.966

B. Dependent variable:	Married (=1)				
	(1)	(2)	(3)	(4)	(5)
Migrant: Stockholm (=1)	-0.190***	-0.203***	-0.200***	-0.183***	-0.173***
	(0.004)	(0.004)	(0.004)	(0.007)	(0.009)
Migrant: other urban area (=1)		-0.055***	-0.048***	-0.042***	-0.040***
		(0.003)	(0.003)	(0.004)	(0.006)
Observations	160636	160636	160634	66070	66070
Mean outcome	0.764	0.764	0.764	0.757	0.757

C. Dependent variable:	Any child (=1)				
	(1)	(2)	(3)	(4)	(5)
Migrant: Stockholm (=1)	-0.239***	-0.257***	-0.248***	-0.231***	-0.215***
	(0.004)	(0.004)	(0.005)	(0.007)	(0.010)
Migrant: other urban area (=1)		-0.075***	-0.066***	-0.062***	-0.061***
		(0.003)	(0.003)	(0.005)	(0.006)
Individual controls	No	No	Yes	Yes	Yes
Family 1880 controls	No	No	Yes	Yes	No
Sibling fixed effects	No	No	No	No	Yes
Observations	160636	160636	160634	66070	66070
Mean outcome	0.700	0.700	0.700	0.694	0.694

Notes: OLS regressions. This table shows the effect of migration on marriage, children, and labor force participation in 1910. *Individual controls* include fixed effects for birth year and birth order, and an indicator for eldest brother. *Sibling fixed effects* is a fixed effect for same sex siblings. Standard errors are given in parentheses and are clustered at the 1880 household level. *** - $p < 0.01$, ** - $p < 0.05$, * - $p < 0.1$.

TABLE A.5: MIGRANTS' EMPLOYMENT USING ALTERNATIVE SAMPLES

Dependent variable:	In labor force (=1)			
Panel A: Extended sample	(1)	(2)	(3)	(4)
Migrant: Stockholm (=1)	0.548***	0.546***	0.554***	0.514***
	(0.006)	(0.006)	(0.006)	(0.009)
Migrant: other urban area (=1)	0.175***	0.174***	0.166***	0.135***
	(0.003)	(0.003)	(0.003)	(0.004)
Observations	156525	156525	156525	156525
Mean outcome	0.188	0.188	0.188	0.188
Panel B: Migrant sample	(1)	(2)	(3)	(4)
Migrant: Stockholm (=1)	0.536***	0.533***	0.545***	0.497***
	(0.007)	(0.007)	(0.007)	(0.010)
Migrant: other urban area (=1)	0.172***	0.171***	0.165***	0.133***
	(0.004)	(0.004)	(0.004)	(0.005)
Observations	77597	77597	77597	77597
Mean outcome	0.228	0.228	0.228	0.228
Panel C: Urban migrant sample	(1)	(2)	(3)	(4)
Migrant: Stockholm (=1)	0.353***	0.352***	0.386***	0.362***
	(0.010)	(0.010)	(0.012)	(0.018)
Individual controls	No	Yes	Yes	Yes
Family 1880 controls	No	No	Yes	No
Sibling fixed effects	No	No	No	Yes
Observations	14601	14601	14601	14601
Mean outcome	0.422	0.422	0.422	0.422

Notes: OLS regressions. All outcomes are measured in 1910. *Individual controls* include fixed effects for birth year and birth order, and an indicator for eldest sister. *Family 1880 controls* include fixed effects for the following within the 1880 household: father's and mother's occupation, family size, number of families, generations, mothers, fathers, couples, servants, unrelated members, as well as an indicator for farming households. Standard errors, in parentheses, are clustered at the 1880 household level. *** - $p < 0.01$, ** - $p < 0.05$, * - $p < 0.1$.

Table A.6: Migrants' marriage and fertility using alternative samples

Dependent variable:	Married (=1)				Any child (=1)			
Panel A: Extended sample	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Migrant: Stockholm (=1)	-0.481***	-0.475***	-0.484***	-0.441***	-0.470***	-0.462***	-0.465***	-0.422***
	(0.006)	(0.006)	(0.006)	(0.009)	(0.006)	(0.006)	(0.006)	(0.009)
Migrant: other urban area (=1)	-0.082***	-0.079***	-0.068***	-0.046***	-0.090***	-0.087***	-0.071***	-0.049***
	(0.004)	(0.004)	(0.004)	(0.005)	(0.004)	(0.004)	(0.004)	(0.005)
Observations	156525	156525	156525	156525	156525	156525	156525	156525
Mean outcome	0.686	0.686	0.686	0.686	0.690	0.690	0.690	0.690
Panel B: Migrant sample	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Migrant: Stockholm (=1)	-0.548***	-0.544***	-0.551***	-0.509***	-0.525***	-0.520***	-0.522***	-0.472***
	(0.006)	(0.006)	(0.007)	(0.010)	(0.007)	(0.007)	(0.007)	(0.010)
Migrant: other urban area (=1)	-0.159***	-0.157***	-0.149***	-0.121***	-0.151***	-0.149***	-0.138***	-0.104***
	(0.004)	(0.004)	(0.004)	(0.006)	(0.004)	(0.004)	(0.004)	(0.006)
Observations	77597	77597	77597	77597	77597	77597	77597	77597
Mean outcome	0.710	0.710	0.710	0.710	0.707	0.707	0.707	0.707
Panel C: Urban migrant sample	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Migrant: Stockholm (=1)	-0.372***	-0.371***	-0.399***	-0.388***	-0.370***	-0.368***	-0.389***	-0.363***
	(0.009)	(0.009)	(0.011)	(0.018)	(0.009)	(0.009)	(0.011)	(0.018)
Individual controls	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Household controls	No	No	Yes	No	No	No	Yes	No
Sibling fixed effects	No	No	No	Yes	No	No	No	Yes
Observations	14601	14601	14601	14601	14601	14601	14601	14601
Mean outcome	0.512	0.512	0.512	0.512	0.519	0.519	0.519	0.519

Notes: OLS regressions. Panel A displays results for an extended sample where we include also non-migrants. Panel B replicates results for the main sample of migrants. Panel C restricts the sample to only include migrants to urban areas. All outcomes are measured in 1910. *Individual controls* include fixed effects for birth year and birth order, and an indicator for eldest sister. *Family 1880 controls* include i) a full set of origin county fixed effects, father's percentile income rank, and family size; and ii) a set of dummies capturing: the mother's LFP, the father's major (1-digit) HISCO group, the presence of servants in the household and whether a household is a married/cohabitating couple with children, single-parent family, extended family (relatives only), or composite (family and non-relatives) as well as whether the family is multigenerational (all measured in 1880). *Sibling fixed effects* is a fixed effect for same sex siblings. Standard errors, in parentheses, are clustered at the 1880 household level. *** - $p < 0.01$, ** - $p < 0.05$, * - $p < 0.1$.

TABLE A.7: UNMARRIED MIGRANTS

Dependent variable:	Married (=1)		Any child (=1)	
	Unmarried 1900 sample			
	pre-migration	post-migration	pre-migration	post-migration
	(1)	(2)	(3)	(4)
Migrant: Stockholm (=1)	-0.460***	-0.323***	-0.467***	-0.305***
	(0.059)	(0.037)	(0.055)	(0.038)
Migrant: other urban area (=1)	-0.143***	-0.118***	-0.135***	-0.116***
	(0.034)	(0.028)	(0.035)	(0.028)
Individual controls	Yes	Yes	Yes	Yes
Sibling fixed effects	Yes	Yes	Yes	Yes
Observations	3036	7144	3036	7144
Mean outcome	0.596	0.408	0.557	0.420

Notes: OLS regressions. All outcomes are measured in 1910. Columns 1 and 2 restrict the sample to women that are observed as unmarried and living in their childhood parish in 1900. Columns 3 and 4 restricts the sample to women that are unmarried and living in their destination in both 1900 and 1910. *Migrant: Stockholm* is an indicator taking value one if the individual lives in Stockholm city in 1910. *Migrant: other urban area* is an indicator taking value one if the individual lives in an urban area other than Stockholm in 1910. *Individual controls* include fixed effects for birth year, birth order, and an indicator for being the eldest sister. *Sibling fixed effects* is a fixed effect for same-sex siblings. Standard errors, in parentheses, are clustered at the family level. *** - $p < 0.01$, ** - $p < 0.05$, * - $p < 0.1$.

TABLE A.8: UNMARRIED EARLY MIGRANTS, BY AGE

Dependent variable:	Married (=1)		Any child (=1)		In labor force (=1)	
	Below 27	27 and over	Below 27	27 and over	Below 27	27 and over
	(1)	(2)	(3)	(4)	(5)	(6)
Migrant: Stockholm (=1)	-0.440***	-0.255***	-0.356***	-0.194***	0.519***	0.265***
	(0.073)	(0.065)	(0.073)	(0.064)	(0.075)	(0.082)
Migrant: other urban area (=1)	-0.101**	-0.171***	-0.120***	-0.082	0.127***	0.244***
	(0.046)	(0.058)	(0.046)	(0.053)	(0.046)	(0.062)
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes
Sibling fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3027	1591	3027	1591	3027	1591
Mean outcome	0.503	0.240	0.495	0.272	0.332	0.591

Notes: OLS regressions. This table shows the effect of migration on marriage, child bearing, and labor force participation in 1910. The sample is restricted to women who had already migrated and were unmarried in 1900. Columns separate the sample by age in 1900. *Individual controls* include fixed effects for birth year and birth order, and an indicator for eldest sister. *Sibling fixed effects* is a fixed effect for same sex siblings. Standard errors are given in parentheses and are clustered at the 1880 household level. *** - $p < 0.01$, ** - $p < 0.05$, * - $p < 0.1$.

TABLE A.9: TEMPORARY MIGRANTS TO STOCKHOLM

Dependent variable:	In labor force (=1)	Married (=1)	Any child (=1)	Higher skill (=1)
	(1)	(2)	(3)	(4)
Migrant in 1900: Stockholm (=1)	0.134***	-0.182***	-0.220***	0.055***
	(0.022)	(0.022)	(0.022)	(0.015)
Migrant in 1900: other urban area (=1)	0.078***	-0.082***	-0.081***	0.014***
	(0.008)	(0.008)	(0.008)	(0.005)
Individual controls	Yes	Yes	Yes	Yes
Sibling fixed effects	Yes	Yes	Yes	Yes
Observations	39225	39225	39225	39225
Mean outcome	0.187	0.743	0.744	0.075

Notes: OLS regressions. All outcomes are measured in 1910. *Individual controls* include fixed effects for birth year and birth order, and an indicator for eldest sister. *Family 1880 controls* include fixed effects for the following within the 1880 household: father's and mother's occupation, family size, number of families, generations, mothers, fathers, couples, servants, unrelated members, as well as an indicator for farming households. Standard errors, in parentheses, are clustered at the 1880 household level. *** - $p < 0.01$, ** - $p < 0.05$, * - $p < 0.1$.

TABLE A.10: HETEROGENEITY BY AGE

Dependent variable:	In labor force	Married	Any children
	(1)	(2)	(3)
Migrant: Stockholm (=1)	0.550***	-0.538***	-0.498***
	(0.016)	(0.015)	(0.015)
Stockholm migrant × Age	-0.008***	0.004**	0.004**
	(0.002)	(0.002)	(0.002)
Migrant: other urban area (=1)	0.128***	-0.107***	-0.102***
	(0.009)	(0.009)	(0.010)
Migrant: other urban area (=1) × Age	0.001	-0.002*	-0.000
	(0.001)	(0.001)	(0.001)
Individual controls	Yes	Yes	Yes
Sibling fixed effects	Yes	Yes	Yes
Observations	77597	77597	77597
Mean outcome	0.228	0.710	0.707

Notes: OLS regressions. Standard errors are given in parentheses and are clustered at the 1880 household level. *** - $p < 0.01$, ** - $p < 0.05$, * - $p < 0.1$.

TABLE A.11: CONTROLLING FOR PRE-MIGRATION OUTCOMES

Dependent variable:	In labor force			Married			Any children		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Migrant after 1900: Stockholm (=1)	0.534***	0.537***	0.524***	-0.465***	-0.445***	-0.448***	-0.477***	-0.449***	-0.449***
	(0.054)	(0.054)	(0.054)	(0.055)	(0.055)	(0.055)	(0.051)	(0.052)	(0.052)
Migrant after 1900: other urban area (=1)	0.150***	0.148***	0.136***	-0.145***	-0.124***	-0.124***	-0.127***	-0.106***	-0.105***
	(0.028)	(0.027)	(0.027)	(0.028)	(0.027)	(0.027)	(0.029)	(0.029)	(0.029)
In labor force 1900		0.134***	0.078**			-0.020			0.001
		(0.030)	(0.031)			(0.029)			(0.030)
Married 1900			-0.188***		0.275***	0.322***			0.066
			(0.037)		(0.022)	(0.040)			(0.049)
Any child 1900			0.018			-0.064		0.312***	0.256***
			(0.037)			(0.039)		(0.024)	(0.049)
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sibling fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	4276	4276	4276	4276	4276	4276	4276	4276	4276
Mean outcome	0.240	0.240	0.240	0.681	0.681	0.681	0.646	0.646	0.646

Notes: OLS regressions. Standard errors are given in parentheses and are clustered at the 1880 household level. *** - $p < 0.01$, ** - $p < 0.05$, * - $p < 0.1$.

Accounting for between-sibling variation: a twin design

There may be additional individual level variation between siblings that can bias results and that are not captured by the inclusion of sibling fixed effects. These should be smallest between twins, who either have identical or near-identical genetics. Using data on birth year and birth month, we are able to identify 1539 sets of twins. Interestingly, Stockholm migration rates in the twin sample are similar to the full sample and close to 5% for both males and females. Table A.12 replicates our results using twin fixed effects. The twin results confirm our main results. All have the same estimated sign as for the full sample. Moreover, all estimates, apart from those on income score, are larger in magnitude than the baseline estimates. Thus, if anything, these estimates indicate that innate character traits are unlikely to be biasing our results in a positive direction.

TABLE A.12: TWIN FIXED EFFECTS

Dependent variable:	In labor force	Married	Any children
	(1)	(2)	(3)
Migrant: Stockholm (=1)	0.760***	-0.744***	-0.536***
	(0.107)	(0.108)	(0.121)
Migrant: other urban area (=1)	0.054	-0.017	-0.023
	(0.079)	(0.083)	(0.086)
Individual controls	Yes	Yes	Yes
Twin fixed effects	Yes	Yes	Yes
Observations	502	502	502
Mean outcome	0.257	0.701	0.683

Notes: OLS regressions. Standard errors are given in parentheses and are clustered at the 1880 household level. *** - $p < 0.01$, ** - $p < 0.05$, * - $p < 0.1$.

TABLE A.13: INFLUENCE OF UNOBSERVED WITHIN-SIBLINGS CHARACTERISTICS (OSTER 2019)

Dependent variable:	In labor force		Married		Any child	
	(1) Baseline	(2) + lagged dep. var.	(3) Baseline	(4) + lagged dep. var.	(5) Baseline	(6) + lagged dep. var.
Beta from Oster(2019)	0.540	0.537	-0.549	-0.448	-0.496	-0.449
Uncontrolled beta	0.463	0.500	-0.481	-0.432	-0.451	-0.450
Controlled beta	0.497	0.524	-0.509	-0.448	-0.472	-0.449

Notes: This table displays coefficients for Stockholm migrants using Oster (2019). Reported coefficients refer to our main regressor of interest, *Stockholm migrant*, an indicator taking value one for individuals who live in Stockholm in 1910, and zero otherwise. All models include sibling fixed effects. Estimates from the barebones model with only sibling fixed effects are report in the row *Uncontrolled beta*. Regression estimates after including controls are shown in row *Controlled beta*. The row *Beta from Oster (2019)* displays the bias-corrected estimate for *Stockholm migrant*. This estimate is based on the difference between the controlled and uncontrolled betas as well as observed changes in R^2 across models. Columns labelled *Baseline* display estimates when adding individuals' age, birth order, and a dummy for being the eldest sister as controls. Columns labelled + *lagged dep. var.* restrict the sample to individuals who had not yet moved in 1900 and further controls for these individuals' pre-migration outcomes for employment, marriage, and childbearing in 1900. Following suggested defaults in Oster (2019), we assume that observed and unobserved variables have the same relative impact on the outcome ($\delta = 1$), and that the maximum R^2 equals 1.3 times the observed R^2 value in the controlled model. *** - $p < 0.01$, ** - $p < 0.05$, * - $p < 0.1$.

TABLE A.14: COHABITATION AND CHILDREN BORN OUT OF WEDLOCK

Dependent variable:	Cohabit	Cohabit or married	Any illegitime child	Any child
	(1)	(2)	(3)	(4)
Migrant: Stockholm (=1)	0.005*	-0.505***	-0.066***	-0.472***
	(0.002)	(0.010)	(0.004)	(0.010)
Migrant: other urban area (=1)	-0.000	-0.121***	0.004	-0.104***
	(0.001)	(0.006)	(0.003)	(0.006)
Individual controls	Yes	Yes	Yes	Yes
Sibling fixed effects	Yes	Yes	Yes	Yes
Observations	77597	77597	77597	77597
Mean outcome	0.007	0.717	0.070	0.707

Notes: OLS regressions. All outcomes are measured in 1910. *Cohabit* is an indicator for households in which there are no married adults, and exactly one single adult female and one single adult male. *Any illegit. children* is an indicator for having any children born out of wedlock. *Individual controls* include fixed effects for birth year and birth order, and an indicator for eldest sister. Standard errors are given in parentheses and are clustered at the 1880 family level. *** - $p < 0.01$, ** - $p < 0.05$, * - $p < 0.1$.

TABLE A.15: OLS REGRESSIONS: WEIGHTING REGRESSIONS WITH THE INVERSE PROBABILITY OF BEING LINKED ACROSS CENSUSES

Dependent variable:	In labor force	Married	Any children
	(1)	(2)	(3)
Migrant: Stockholm (=1)	0.504***	-0.519***	-0.483***
	(0.010)	(0.010)	(0.010)
Migrant: other urban area (=1)	0.133***	-0.122***	-0.106***
	(0.005)	(0.006)	(0.006)
Individual controls	Yes	Yes	Yes
Sibling fixed effects	Yes	Yes	Yes
Observations	75051	75051	75051
Mean outcome	0.225	0.713	0.712

Notes: OLS regressions. All outcomes are measured in 1910. Regressions are weighted by probability weights calculated from regressing an indicator for being successfully linked on age, age squared, as well as fixed effects for birth order, childhood county, and father's social class (using HISCLASS). *Individual controls* include fixed effects for birth year and birth order, and an indicator for eldest sister. Standard errors, in parentheses, are clustered at the 1880 family level. *** - $p < 0.01$, ** - $p < 0.05$, * - $p < 0.1$.

TABLE A.16: MIGRANTS' MAJOR OCCUPATIONAL (HISCO) GROUPS

Dependant variable:	Professional	Administrative	Clerical	Sales	Service	Agricultural	Production
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Migrant: Stockholm (=1)	0.013***	0.018***	0.026***	0.033***	0.344***	-0.008***	0.078***
	(0.005)	(0.005)	(0.003)	(0.004)	(0.009)	(0.001)	(0.006)
Migrant: other urban area (=1)	-0.002	0.002	0.006***	0.018***	0.070***	-0.007***	0.046***
	(0.002)	(0.002)	(0.001)	(0.002)	(0.004)	(0.001)	(0.003)
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sibling fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	77597	77597	77597	77597	77597	77597	77597
Mean outcome	0.029	0.027	0.008	0.017	0.091	0.010	0.061

Notes: OLS regressions. All outcomes are measured in 1910. *Professional* captures the major groups 0–1 in HISCO covering "professional, technical and related workers". *Administrative* captures the major group 2 covering "administrative and managerial workers". *Clerical* captures the major group 3 covering "clerical and related workers". *Sales* captures the major group 4 covering "sales workers". *Service* captures the major group 5 covering "service workers". *Agricultural* captures the major group 6 covering "agricultural, animal husbandry and forestry workers, fishermen and hunters". *Production* captures the major groups 7–9 covering "production and related workers, transport equipment operators and labourers". *Individual controls* include fixed effects for birth year and birth order, and an indicator for eldest sister. Standard errors, in parentheses, are clustered at the 1880 family level. *** - $p < 0.01$, ** - $p < 0.05$, * - $p < 0.1$.

TABLE A.17: PERSISTENCE IN OUTCOMES AFTER MIGRATION

Dependent variable:	Low skill (=1)		Unskilled (=1)		Income score		HH inc. score	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Migrant in 1900: Stockholm (=1)	0.236***		0.028***		0.030		0.157***	
	(0.014)		(0.006)		(0.037)		(0.016)	
Migrant in 1900: other urban area (=1)	0.067***		0.007***		-0.113***		-0.004	
	(0.006)		(0.003)		(0.027)		(0.008)	
Ever migrated: Stockholm (=1)		0.277***		0.023***		0.098***		0.179***
		(0.008)		(0.003)		(0.021)		(0.009)
Ever migrated: other urban area (=1)		0.067***		0.008***		-0.141***		-0.043***
		(0.004)		(0.002)		(0.015)		(0.005)
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sibling fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	42691	88000	42691	88000	4020	8890	36311	75212
Mean outcome	0.129	0.132	0.019	0.019	6.989	6.994	6.831	6.839

Notes: OLS regressions. All outcomes are measured in 1910. *Migrant in 1900: Stockholm* is an indicator taking value 1 if the individual lives in Stockholm city in 1900, and 0 if not. *Ever migrated: Stockholm* is an indicator taking value 1 if the individual lives in Stockholm city in 1900 or 1910, and 0 if not. Similarly, *Migrant in 1900: other urban area* and *Ever migrated: other urban area* are indicator variables for living in urban parishes other than Stockholm. *Individual controls* include fixed effects for birth year and birth order, and an indicator for eldest sister. *Sibling fixed effects* is a fixed effect for same sex siblings. Standard errors are given in parentheses and are clustered at the 1880 family level. *** - $p < 0.01$, ** - $p < 0.05$, * - $p < 0.1$.

TABLE A.18: MIGRANTS' HOUSEHOLD INCOME WITH DIFFERENT WEIGHTS FOR SPOUSE INCOME

Dependant variable:	Household income score					Intra-household ratio
	100	80	60	40	20	
	(1)	(2)	(3)	(4)	(5)	(6)
Migrant: Stockholm (=1)	0.206***	0.326***	0.482***	0.701***	1.075***	0.512***
	(0.011)	(0.011)	(0.012)	(0.014)	(0.020)	(0.010)
Migrant: other urban area (=1)	-0.077***	-0.048***	-0.010	0.044***	0.136***	0.126***
	(0.006)	(0.006)	(0.006)	(0.008)	(0.011)	(0.006)
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes
Sibling fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	67216	67216	67216	67216	67216	77597
Mean outcome	6.847	6.681	6.467	6.166	5.652	0.298

Notes: OLS regressions. All outcomes are measured in 1910. For completeness, we set the intra-household ratio (of female to male income) equal to one if the woman lives in her own household. *Individual controls* include fixed effects for birth year and birth order, and an indicator for eldest sister. Standard errors, in parentheses, are clustered at the 1880 family level. *** - $p < 0.01$, ** - $p < 0.05$, * - $p < 0.1$.

TABLE A.19: MIGRANTS' HEALTH OUTCOMES

Dependent variable:	Age at death			
	Women		Men	
	(1)	(2)	(3)	(4)
Migrant: Stockholm (=1)	-0.145	-0.537	-2.056***	-1.719***
	(0.352)	(0.393)	(0.351)	(0.371)
Migrant: other urban area (=1)	0.291	0.253	-1.065***	-0.739***
	(0.198)	(0.203)	(0.236)	(0.251)
Individual controls	Yes	Yes	Yes	Yes
Sibling fixed effects	Yes	Yes	Yes	Yes
Occupation fixed effects	No	Yes	No	Yes
Observations	59994	59863	47994	47857
Mean outcome	73.631	73.632	71.785	71.788

Notes: OLS regressions. *Individual controls* include fixed effects for birth year and birth order, and an indicator for eldest sister. *Family 1880 controls* include fixed effects for the following within the 1880 family: father's and mother's occupation, family size, number of families, generations, mothers, fathers, couples, servants, unrelated members, as well as an indicator for farming households. *Sibling fixed effects* is a fixed effect for same sex siblings. *** - $p < 0.01$, ** - $p < 0.05$, * - $p < 0.1$.

TABLE A.20: THE ROLE OF SERVICES AND LABOR MARKET COMPOSITION, FAMILY FORMATION

Dependent variable:	Married (=1)				Any child (=1)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Migrant: Stockholm (=1)		-0.303***	-0.287***	-0.250***		-0.290***	-0.277***	-0.293***
		(0.018)	(0.018)	(0.029)		(0.018)	(0.019)	(0.029)
Migrant: other urban area (=1)		-0.016*	-0.015	-0.008		-0.009	-0.011	-0.002
		(0.009)	(0.009)	(0.011)		(0.009)	(0.009)	(0.012)
Destination service share	-0.448***	-0.445***			-0.423***	-0.434***		
	(0.030)	(0.036)			(0.030)	(0.037)		
Destination industry share	0.178***	0.016			0.181***	0.020		
	(0.016)	(0.018)			(0.016)	(0.017)		
Destination workers (log)	-0.060***	-0.019***	-0.017***	-0.020***	-0.054***	-0.014***	-0.014***	-0.011***
	(0.002)	(0.003)	(0.003)	(0.004)	(0.002)	(0.003)	(0.003)	(0.004)
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sibling fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cubic sector shares	No	No	Yes	No	No	No	Yes	No
Sector share percentile FE	No	No	No	Yes	No	No	No	Yes
Observations	77597	77597	77597	77597	77597	77597	77597	77597
Mean outcome	0.710	0.710	0.710	0.710	0.707	0.707	0.707	0.707

Notes: OLS regressions. All outcomes are measured in 1910. *Migrant: Stockholm* is an indicator taking value 1 if the individual lives in Stockholm city in 1910, and 0 if not. *Migrant: other urban area* is an indicator taking value 1 if the individual lives in an urban area other than Stockholm in 1910, and 0 if not. *Individual controls* include fixed effects for birth year and birth order, and an indicator for eldest sister. *Sibling fixed effects* is a fixed effect for same sex siblings. *Cubic sector polynomial* includes cubic polynomials in the share working in services and in industry. Standard errors, in parentheses, are clustered at the 1880 family level. *** - $p < 0.01$, ** - $p < 0.05$, * - $p < 0.1$.

B Data Appendix: Record Linkage

Our empirical strategy follows individuals across censuses. In particular, we focus on children observed in their childhood home in 1880 and link them to the 1910 census when they are in their adulthood. To do so, we rely on probabilistic linking methods using a fully automated procedure, ensuring full replicability. We describe the linkage procedure in greater detail in this appendix.

B.1 Data sources

The data sources used in the linkage process comes from the full-count decennial censuses of 1880 and 1910 distributed through IPUMS International (Minnesota Population Center, 2020). The Swedish censuses are well-known for their high accuracy of spelling and birth years, improving accuracy of record linking across census rounds. Crucially for its high quality, enumeration was not based on self-reports to census takers. Instead, local priests were in charge of keeping registers of all inhabitants in their parish, recording demographic information such as dates of births, deaths, and marriages every year.

B.2 Children's first and last names in the censuses

The first step to prepare the census data for record linkage entails establishing and cleaning the first and last names of all individuals. For first names, we opt for using as much information as possible and make use of all reported first names in the censuses. Thus, our definition of first names includes middle names. This is motivated by the naming convention of not always placing the used given name first. Establishing last names is somewhat less straightforward in our setting due to the fact that many children living in their family households lack a registered surname in the census. As we focus on children in the 1880 census, aged 0–15, neglecting this aspect would reduce our sample considerably. In fact, only 8 percent of our sample of children in the 1880 census have a reported surname. The lack of a registered surname is due to the organization of the registers (Wisselgren et al., 2014). Luckily, by using information on the parents living in the same household, it is possible to impute the surname of children.

To complicate things, however, there existed parallel naming conventions for surnames in our time period, with both family names and patronymic names. This implies that it is not explicit which type of surname that will be adopted for a given child (with missing surname information in the census). We therefore follow Wisselgren et al. (2014) and construct three last name categories: i) family names, ii) patronymic names used as family name, and iii) true patronymic names.

For children lacking surnames and with a biological father present in the household, we impute the surname of the child. First, if the father has a family name, this is entered as a family name for the child, i.e., category ii). About 34 percent of children without an explicit surname have a father with a family name. Second, if the father has a patronymic name used as a family name, this is entered in category ii) without the suffix. About 72 percent of children without an explicit surname obtains a patronymic family name in this way. Third, we construct a patronymic name using the first name of the father and place this in category iii). This is possible for about 92 percent of children without explicit surname. For children without a father present, we construct family names and patronymic family names using information on the mother following the same procedure.

For children with explicit surnames, names not ending with "son" and "dotter" are placed in the family name category, while names including these suffixes are placed in both category ii) and iii). The suffixes "son" and "dotter" are then removed. About 69 percent of children in the 1880 census with a reported surname have a family name.

Finally, for individuals with two (explicit or constructed) family names or patronymic family names, we make use of both by constructing two versions in each category. This is of particular importance for married women, since they typically have two reported surnames of which one is the maiden name. With these surname categories in place, we move on to the record linking process.

B.3 Linkage procedure

For the linkage procedure, we start by designating stable index variables which have to match exactly for two records to be considered potential matches: sex, birth year, and parish of birth. The detail and accuracy of these time-invariant variables allow us to construct a relatively small set of candidate links. In particular, two features stands out as favorable compared to other national censuses. First, since local priests were in charge of keeping the registers, the birthplace is recorded at the parish level, which constitutes a relatively small geographic area (there were about 2,400 parishes during our time period). Second, since the parish books were continuously updated, birth years do not suffer from recall error, something which is evident from the lack of age-heaping in the Swedish censuses.[42] The latter allows us to only consider potential matches among candidates with the same exact birth year.

Next, we evaluate these candidate links by comparing first and last names. We do this in three steps. First, if two candidates have the exact same full list of first names (typically consisting of three names) and the full list of last names (typically consisting of one name), and there are no other candidates fulfilling the same criteria we consider this a match. Note that individuals with no explicit surname in the census cannot be matched in this first step. Given that children with an

[42]See Berger et al. (2023) for a comparison of age-heaping between different national sources.

explicitly reported surname are few, the amount of links established in this first step are few (only 1.6 percent of all children in the 1880 census are linked this way).

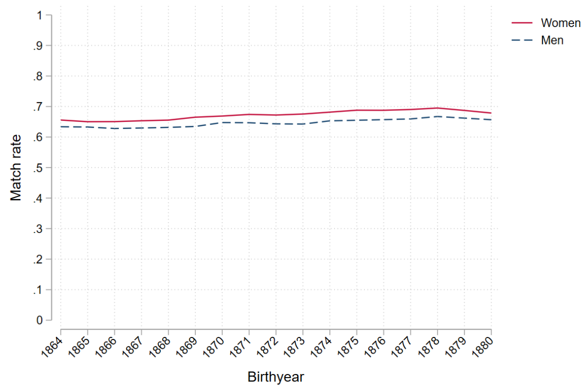
Second, if no match is found in step 1, we make use of our constructed surnames and compare surnames within each of the three categories (family names, patronymic family names, and true patronymic names). To assess name similarities within these categories, we employ the Jaro-Winkler algorithm, which compares two strings and assigns a similarity score between 0 (no similarity) and 1 (identical). For most pairs of individuals, this will produce three different scores, i.e., one for each category (but for the few pair of individuals were both have two family names and two patronymic names, we obtain twelve scores). We save the highest score for each considered pair of individuals. To compute the similarity between first names without imposing any order of first names, we calculate the mean for the n number of first name pairs with the highest Jaro-Winkler score, where n is equal to the number of first names in the record with the least number of first names in the pair. Since the used given name may be placed in any order among the reported first names, this allows for individuals with several reported first names to be matched to individuals with only one reported given name. With these two scores in hand, for each individual in census 1910, we rank all potential candidates in census 1880. We consider individuals linked if there is a match within the same sex$\times$birth year$\times$place of birth cell that satisfies a Jaro-Winkler threshold of at least 0.85 for *both* the first and the last name and require that there is no close runner-up. We set the cut-off at a distance of 0.05 units between the highest ranked candidate and the runner-up. The vast majority of matched pairs (about 85.5 percent) are matched in this way.

Second, if no match is found in step 3, we proceed with an additional attempt to match individuals. Instead of constructing the first name score as described above, we simply use the score from comparing the name similarity between the full list of first names. Since the Jaro-Winkler metric puts more weight on the first entries in a string, this puts more emphasis on the name ordered first in the full list of first names. About 11 percent of our matched individuals are obtained in this last step.

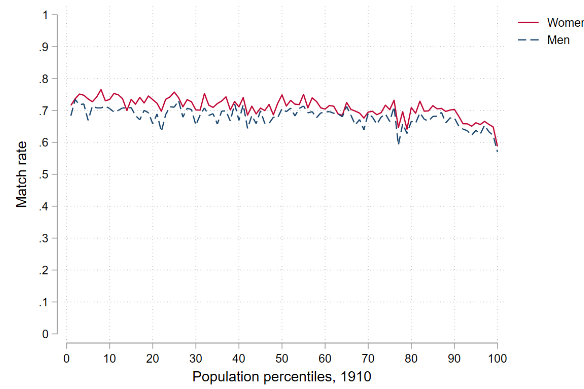
B.4 Linked sample and linkage rates

In terms of linkage rates, we are able to identify 66 percent of individuals born 1862–1880 in the 1910 census back to the 1880 census.[43] Importantly, our linkage rates of men and women are similar. If anything, we match slightly more women with a backward linkage rate of 67 percent (compared to 65 percent for men). For both men and women, linkage rates are similar across birthyears and locations (see Appendix Figure B.1), although more populated locations generally display somewhat lower name rates. The latter is expected due the fact that there are more potential

[43] The forward linkage rate is 43 percent.



(A) BIRTHYEAR



(B) LOCATION

FIGURE B.1: MATCH RATES BY BIRTHYEAR AND LOCATION FOR MEN AND WOMEN.

Notes: These figures display backward match rates between the 1910 census and the 1880 census. In panel A, match rates are displayed by birthyear and sex. In panel B, match rates are displayed by location (in terms of population percentile) and sex.

candidates in more populated parishes.

While we are able to achieve relatively high linkage rates, it is possible that matched individuals differ systematically from those that are unmatched, possibly yielding unrepresentative estimates. For example, it is easier to link individuals with uncommon names, and name commonality has been linked to traits such as individualism and socio-economic status. With this in mind, Appendix Table B.1 compares matched individuals to the full population in the same age cohorts on observable characteristics measured in 1880. The table shows overall small differences between the two samples, suggesting that our sample is representative of the population. Nevertheless, we show that our results are nearly identical when we use probabilistic weights, reflecting the probability of an observation being selected into the sample (see Appendix Table A.15).[44]

[44]To calculate these, we use the full census data to regress an indicator for being successfully linked on age, age squared, as well as fixed effects for birth order, childhood county, and father's social class (using HISCLASS).

Table B.1: Summary statistics for full population and linked sample in 1880 census

	All mean	Linked mean
Age	7.317	7.210
Birthorder	2.758	2.743
Eldest sister (=1)	0.478	0.481
Father's age	42.783	42.724
Mother's age	39.547	39.416
Mother in labor force (=1)	0.002	0.002
Father in labor force (=1)	0.922	0.929
Father white collar (=1)	0.138	0.139
Family members in household	6.241	6.196
Any servants (=1)	0.191	0.217
Multigenerational family (=1)	0.059	0.060
HH type: Extended family (=1)	0.049	0.047
HH type: Composite (=1)	0.223	0.250
Observations	589399	287456

Notes: The table displays mean values for the 1880 census and our linked sample. Both consists of girls aged 0–16 years in 1880.