Cities and the Rise of Working Women*

Thor Berger† Mounir Karadja‡ Erik Prawitz§

December 29, 2025

Abstract

This paper documents substantial spatial variation in female labor force participation (FLFP) within early 20^{th}-century Europe and the United States with remarkably high participation rates among women in large cities. To shed light on the role of cities in shaping women's work, we use linked census data to show that female migrants to large cities in Sweden and the United States experienced substantial increases in FLFP, as well as sharp reductions in marriage and fertility. When focusing on migrants to Stockholm, we find increases in FLFP of about 50 percentage points and similarly large decreases in marriage rates that persist over the life cycle. The increases in employment among migrants were primarily driven by a transition into services jobs, which was facilitated by the fact that major American and European cities had experienced a structural shift toward services already in the early 20^{th} century, up to half a century before a similar shift took place in smaller cities and rural areas.

*We thank Leah Boustan, Martín Fernández Sánchez, Raoul van Maarseveen, JF Maystadt, Zach Ward, Martin Önnerfors, 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

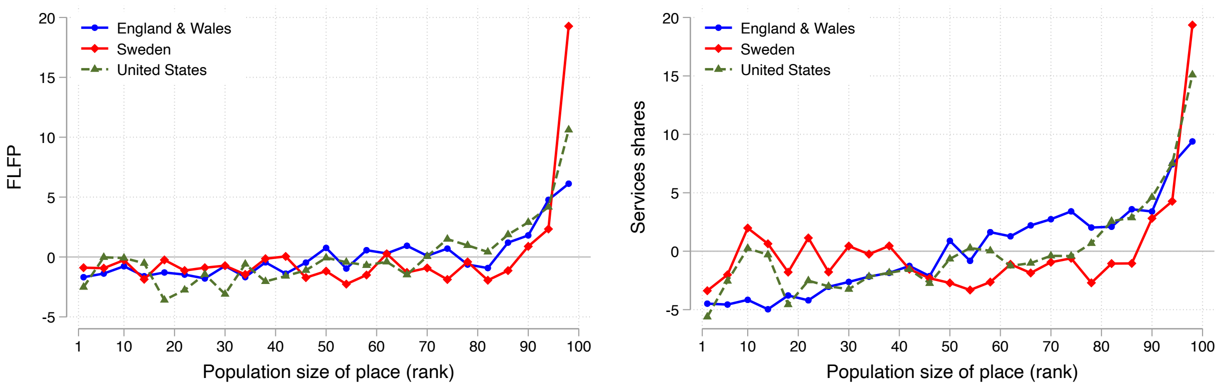
Women's entry into the formal labor market is among the defining economic and social changes of the 20^{th} century. An influential literature documents trends in female labor force participation (FLFP) across countries and highlights the role of culture, institutions, and structural change in shaping the transformation of women's work (Goldin, 1990, 1995; Costa, 2000; Olivetti, 2014). However, national aggregates can mask substantial spatial heterogeneity in women's economic opportunities, both historically and today (Fogli and Veldkamp, 2011; Chetty et al., forthcoming).

Consider Panel A of Figure 1, which displays the distribution of women's formal employment across counties and parishes in early 20^{th}–century Britain, Sweden, and the United States ranked by their population.[1] In all three countries, FLFP is uniformly low across much of the distribution but rises sharply in the most populated areas—an indication that large cities offered distinctive labor-market opportunities for women. The pattern of high FLFP rates in large cities has been recognized in historical accounts, as well as by contemporary observers: for example, a special U.S. Department of Labor report was designated to examine the causes and consequences of the rapid rise of "working women" in major cities (Wright, 1889). Yet, there is a lack of rigorous quantitative research on the role of cities in shaping women's work.

This paper shows that large cities were crucial for women's entry into the labor market in the early 20^{th} century. To do so, we study the outcomes of female urban migrants originating from rural areas. Yet, analyzing the returns to migration for women is complicated because women are notoriously difficult to link across historical censuses, due to name changes at marriage. The main analysis therefore makes use of Swedish census data 1880–1910, which consistently recorded women's maiden names, providing a unique opportunity to construct a large and representative linked sample of women who can be traced over time and across space.

We first descriptively analyze differences in FLFP among women who grew up in rural areas and later moved across Sweden's 2,400 rural and urban parishes by the early 20^{th} century. Migrants to the most populous parishes saw substantial increases in FLFP compared to stayers and migrants

[1] We use FLFP and women's formal employment interchangeably because our measure follows census definitions of "gainful" employment. We focus on formal, paid work outside the home as it reflects women's entry into labor markets and growing economic independence. While many women worked informally—for example, as homeworkers or unpaid on farms and were often not counted as in the labor force (Goldin, 1990; Stanfors, 2014; Chiswick and Robinson, 2021)—their exclusion does not affect census-based measures of formal employment (Costa, 2000, p. 103).



(A) FLFP (B) SERVICE JOBS

FIGURE 1: THE GEOGRAPHY OF FLFP AND SERVICE JOBS IN THE EARLY 20^{th} CENTURY.

Notes: This figure shows the average FLFP rate (Panel A) and share of service jobs (Panel B) in counties (United States) and parishes (England & Wales or Sweden) where counties and parishes are grouped into 25 equal-sized bins within each country based on their population. For each bin, we calculate the percentage point difference in FLFP and the share of employment in services in these places and the average FLFP rate and services share observed across all places in each country. The data is based on the 1911 census for England & Wales and the 1910 census for Sweden and the United States. See Appendix E.1 for details on data construction.

to less populous destinations, which closely mirrors the non-linear pattern in Figure 1. In contrast, female migrants to destinations outside the top decile of the population distribution exhibit no changes in FLFP, whereas male migrants experience similarly high returns to migration regardless of their destination.

Motivated by the sharp non-linearity in the returns to migration for women, we zoom in on migration to Sweden's largest city: Stockholm—an emblematic example of an emerging metropolis in a rapidly industrializing country.[2] Between the mid-19^{th} century and 1910, the city tripled in size due to substantial inward migration. Women greatly outnumbered men in Stockholm and were also overrepresented among migrants, who made up more than half of the city's residents. Notably, more than half of the prime-aged women in Stockholm were in the labor force, which is about 40 percent more than in other cities and nearly three times the rate in rural areas.

To analyze Stockholm's role in shaping FLFP, we compare the outcomes of migrants to the capital with migrants to other destinations. A simple OLS estimate shows that women who move to

[2]Our focus on Stockholm is motivated by the fact that it is broadly representative in terms of size of other major American and European urban areas in this period. Its population of 342,000 inhabitants in the early 20^{th} century made it slightly larger than rapidly growing American cities such as Minneapolis, Los Angeles, or Washington, D.C., and moderately smaller than Detroit or San Francisco.

Stockholm are about 50 percentage points more likely to enter the labor force than other migrants, compared to an overall FLFP rate of about 23 percent in the sample. However, a key concern is that women who move to Stockholm may differ systematically from those moving to other rural and urban locations. For example, if daughters from more industrious families are more likely to move to Stockholm, a simple comparison of their outcomes with migrants to other urban and rural locations would overstate the causal effect of Stockholm on women's work.

We address the issue of migrant sorting using three empirical strategies showing that the estimated impact on FLFP from moving to Stockholm remains relatively stable at around 50 percentage points across specifications. First, we control for a rich set of individual, family, and origin characteristics that accounts for sorting on observables. Second, we include family fixed effects and compare sisters migrating to different destinations, which absorbs all shared individual and family traits though still leaving scope for within-family sorting across destinations. Third, we add family-by-destination fixed effects and compare sisters to their brothers who move to the *same* destination, which limits the potential role of within-family sorting to the extent that it does not vary by sex. These strategies and a battery of robustness tests consistently suggest that both between- and within-family sorting across destinations have at most a limited effect on our results and that moving to Stockholm had a substantial impact on FLFP.[3]

Why did moving to a large city lead to increases in women's formal employment? In the early 20^{th} century, the U.S. Census Report noted that women were drawn to urban areas because of their distinct labor markets: "[i]n the urban communities there are many opportunities for the employment of women, in the rural comparatively few" (U.S. Census Bureau, 1913, p. 260). A similar explanation exists for the aggregate rise in FLFP over the past century, which is largely attributable to the structural shift towards the services sector that provided "respectable" jobs for women (Goldin, 1995; Olivetti, 2014; Rendall, 2018). We document an overlooked spatial dimension to this shift: Panel B of Figure 1 shows that the most populous places in Europe and the United States had already experienced a shift towards the service sector by the early 20^{th} century, which other places achieved only in the post–World War II era. In line with this, we show that the increased FLFP rates among migrants to Stockholm is primarily driven by an increased transition

[3]We additionally test for within-family sorting by controlling for pre-migration outcomes in adulthood, examining the role of distance from Stockholm and potential strategic family migration, and by applying the method of Oster (2019) to detect sorting based on unobservable characteristics.

into service jobs: the probability to be employed in a services job increases by about 40 percentage points among migrants, which can account for the vast majority of the increase in FLFP.

We then examine the types of jobs that female migrants transition into. First, while the bulk of migrants are concentrated in relatively low-skilled service work, there are increases in employment across the skill spectrum. Notably, the likelihood of working in higher-skilled occupations among Stockholm migrants increases with age, consistent with learning and a greater return to experience for migrants in large cities (Glaeser and Maré, 2001; Roca and Puga, 2017; Eckert et al., 2022). Second, migrants transition into jobs that on average pay better, which is also reflected in greater rates of intergenerational income mobility. Third, the economic gains among female migrants are not offset by an urban mortality penalty: female migrants in Stockholm live as long as their migrant sisters, which contrasts the case of male migrants who experience a steep reduction in lifespan in the capital.

At the turn of the 20^{th} century, female employment was deeply intertwined with marriage decisions. Women typically exited the labor force upon marriage, reflecting the substantial marriage penalties that existed historically (Goldin, 1990).[4] Marriage and fertility patterns of female migrants across different destinations reveal a mirror image of the results for FLFP: migrants to the most populous locations are considerably less likely to marry and have children. Again, the magnitudes are large with an estimated reduction in the probability to be married or having a child among Stockholm migrants of about 51 and 48 percentage points compared to their sisters migrating to other destinations, a difference that persists throughout the life cycle and beyond the prime reproductive years. Thus, the persistent increases in FLFP that we observe among migrants to large cities such as Stockholm are mirrored in similarly large and persistent declines in marriage and fertility.

To bolster a causal interpretation of our results, we propose an IV strategy that instruments for migration to Stockholm using two sources of variation: (i) the fact that women are more likely to migrate if they have an older sister; and (ii) that proximity from the childhood parish to Stockholm increases the likelihood of migrating there in adulthood. The instrument is defined as the

[4] As shown by Kleven et al. (2024), marriage penalties in terms of FLFP were substantial in both European countries such as Sweden and the United States around the turn of the 20^{th} century, which is similar to many lower-income countries today. In contrast, the role of fertility and child penalties were historically relatively less important in shaping women's employment patterns.

interaction of these variables, while controlling for their separate components directly, as well as fixed effects for individual birth order, birth year, number of siblings, and childhood parish. We provide a balance test for the instrument and show that it does not predict migration in general, but is highly predictive of migration to Stockholm. IV estimates are in line with our OLS estimates with family fixed effects: migration to Stockholm leads to a 43 percentage point increase in FLFP, with a commensurate decrease in marriage. We corroborate these findings using several specification and exclusion restriction tests. In addition, a placebo instrument has no impact on womens' outcomes, and the main instrument has no impact on males in our sample—a pattern that is difficult to reconcile with results being driven by unobserved confounders.

Our estimates across a variety of different empirical strategies indicates that moving to Stockholm had substantial effects on FLFP and women's marriage decisions. Crucially, the evidence suggests that only a quite particular form of within-family sorting could overturn these results. Specifically, such sorting would need to be orthogonal to observed individual characteristics in childhood and early adulthood, unrelated to both observable and unobservable family-level traits, and concentrated among women, since brothers and sisters migrating to the same destination experience markedly different outcomes. In other words, sorting would have to be driven by unobserved, time-varying, sex-specific individual factors and, given the magnitudes we document, it would need to be substantial. Our sensitivity analysis using the method of Oster (2019) indicates that this is unlikely: if anything, the bias-corrected estimates are somewhat larger.

Figure 1, Panel A, shows that the non-linear relationship between FLFP and local population in Sweden is similar in Britain and the United States, which raises the question of whether our findings generalize to other countries. To establish the external validity of our results, we draw on recent advancements in record-linkage of U.S. census data that combines genealogical information and machine learning (ML) algorithms to track millions of women between the 1880 and 1910 censuses (Buckles et al., 2023). Using a similar sibling design as above, we replicate all our main results for female migrants in the United States. Most importantly, we document large increases in FLFP and sharp reductions in marriage and fertility among migrants to the largest U.S. cities, including cities of a size similar to Stockholm. The fact that results are similar in Sweden and the United States—two countries that differed considerably in terms of culture, institutions, and urbanization—suggests that our main results are informative also for other historical contexts.

Our paper documents substantial spatial heterogeneity in FLFP within early 20^{th}–century Europe and the United States and establishes the key role that large cities played in accounting for women's early labor-market advances. The quantitative evidence that large cities substantially raised FLFP contributes to a mainly descriptive and qualitative literature on women's labor market outcomes in cities during industrialization (see, for example, Fuchs and Moch (1990), Hill (1994), Engel (1994), Deutsch (2000), Pooley and Turnbull (2005), and You (2020)). We also contribute new evidence on the crucial role of local structural change in shaping women's work. While prior work has emphasized the role of manufacturing in cities in contributing to FLFP during industrialization (Tilly and Scott, 1978; Goldin, 1980; Goldin and Sokoloff, 1982; Karlsson, 1996; Kim, 2005), we show that large cities had experienced an early structural shift toward services that was mirrored in comparatively high FLFP levels. By showing that the shift toward services in large cities was mirrored in high FLFP rates already around the turn of the 20^{th} century, the results are consistent with aggregate evidence that the development of the service sector 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).

A growing body of work examines the origins of the long-run increase in FLFP. First, one stream of work documents an important role for social and cultural influences (Fernández et al., 2004; Fernandez, 2007; Fernández, 2013; Alesina et al., 2013; Olivetti et al., 2020). While we show that the early development of the service sector is a key explanation behind the high levels of FLFP in large cities, it does not fully account for the entire difference compared to rural areas—cultural and social norms in large cities may thus partly play a role. Second, a key driver of women's rising FLFP in the early 20^{th} century was electrification and the spread of new household technologies, which was typically adopted earlier in large cities (Greenwood et al., 2005; Vidart, 2024). Yet, the fact that the gains we document in FLFP are concentrated among single women (less affected by new household appliances) suggests it is unlikely to explain our results. Third, a recent literature has documented the central role of urban labor markets in women's adaptation to automation and workplace technological changes (Ager et al., 2023; Feigenbaum and Gross, 2024; Rashid, 2025), which we corroborate by providing more general evidence that large cities offered substantial economic opportunity for women.

Our results also provides new evidence on how migration affected women's economic and demographic outcomes, which extends a vast literature studying male migrants including the returns to rural-to-urban migration (for example, Long, 2005; Abramitzky et al., 2012; Collins and Wanamaker, 2014; Ward, 2020; Zimran, 2024; Andersson and Molinder, 2025).[5] Analyzing the returns to migration for women is complicated by the fact that women can not be linked over time in historical census data using standard linkage techniques. While some have used marriage records or genealogical data to track subsets of women over time (Withrow, 2021; Feigenbaum and Gross, 2024; Eriksson et al., 2025), women are typically excluded from linked samples such as the pioneering Census Linking Project (Abramitzky et al., 2020). As described above, we use the historical Swedish censuses that provides a unique opportunity to link women using established automated methods to extend the analysis to women (Wisselgren et al., 2014a; Dribe et al., 2019; Andersson and Molinder, 2025), as well as the genealogical data from Buckles et al. (2023) to extend our analysis to another country to document the external validity of our findings.[6]

Taken together, the results on the role played by large cities in shaping women's economic and social outcomes in the early 20^{th} century challenge several stylized facts regarding the historical returns to migration based on studies of male migrants. First, while male migrants experience increases in LFP regardless of the migrant destination, we uncover strong non-linearities in the relationship between migration and LFP for women with substantial increases only in the largest cities. Second, female migrants' face no urban mortality penalty in the biggest cities, which contrasts the case of male migrants who saw substantial reductions in lifespan. Third, we show that the increases in FLFP due to migration to the emerging metropolises of the early 20^{th} century shaped marriage and fertility decisions in ways that contributed the rise of the "independent fe-

[5]A large body of work also examines the role of migration in shaping men's intergenerational mobility (Long and Ferrie, 2013; Pérez, 2019; Berger et al., 2023). Yet, most prior literature has studied women's intergenerational mobility by using marriage records or inferring their social status from their spouses (Dribe et al., 2019; Olivetti et al., 2024; Eriksson et al., 2025), or by relying on pseudo links (Olivetti and Paserman, 2015). We show that migration contributed to women's absolute and relative income mobility thus contributing to a recent literature that extends the study of intergenerational mobility beyond men (Althoff et al., 2025; Jácome et al., 2025).

[6]Although the U.S. data provide a valuable opportunity to replicate our main findings, we note that the Swedish census data has some distinct advantages compared to genealogical records. For Sweden, we can use established algorithms to link representative samples of men and women across censuses (Ferrie, 1996; Abramitzky et al., 2021), whereas the U.S. links rely on genealogical records. A key concern is that childless women in large cities are less likely to appear in such records, since they have no descendants to enter them into the family trees, potentially reducing representativeness. We address this by showing that our U.S. results remain robust to alternative linkage methods that does not directly rely on handmade links.

male worker"—a key phase in women's transition into the formal labor market (Goldin, 2006).

2 Historical context and stylized facts

2.1 Cities, services, and FLFP: stylized facts

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 census data of England & Wales (1911), Sweden (1910), and the United States (1910).[7] We divide counties and parishes into equal-sized groups within each country and calculate the mean FLFP rate among prime-aged (20–55) women in each group. To facilitate comparisons, we display the percentage point difference between FLFP in each group and the country average. Strikingly, there is a discontinuous increase in FLFP in the most populous places in all three countries.

A key factor that may account for the higher FLFP rates in large cities in early 20^{th}–century Britain, Sweden, and the United States is that they had experienced the shift toward services earlier than less populated areas. An influential literature argues that an important driver of women's entry into formal employment was the growth of the services sector.[8] Panel B of Figure 1 displays differences in the share of services jobs in counties and parishes where we again plot percentage point differences between the service share observed in a county or parish and the average service share across all locations within a country. Notably, the quantitative differences in FLFP rates and services jobs closely align: for example, in the largest parishes in Sweden both FLFP and the share of service jobs is about 20 percentage points higher than the average across parishes within the country.

Figure 2 then examines the association between FLFP and services more directly across rural

[7] See Appendix E.1 for details on data construction.

[8] A central finding in this literature is that more inclusive measures of FLFP, which include the informal and often unpaid work by women at family farms or family-owned enterprises, follows a U-shape with respect to economic development both across and within countries (Boserup, 1970; Goldin, 1990, 1995; Olivetti, 2014, 2013; Mammen and Paxson, 2000). While most women in the 19^{th} century were carrying out informal work within agricultural households, women's employment declined as production shifted from the household to the factory during industrialization. The subsequent shift towards the services sector led to an increase in women's entry into formal service jobs with less of a social stigma for female workers. Because we focus on women's formal employment and largely exclude informal and unpaid work, our FLFP measure is most relevant in characterizing the rising part of the U-shape.

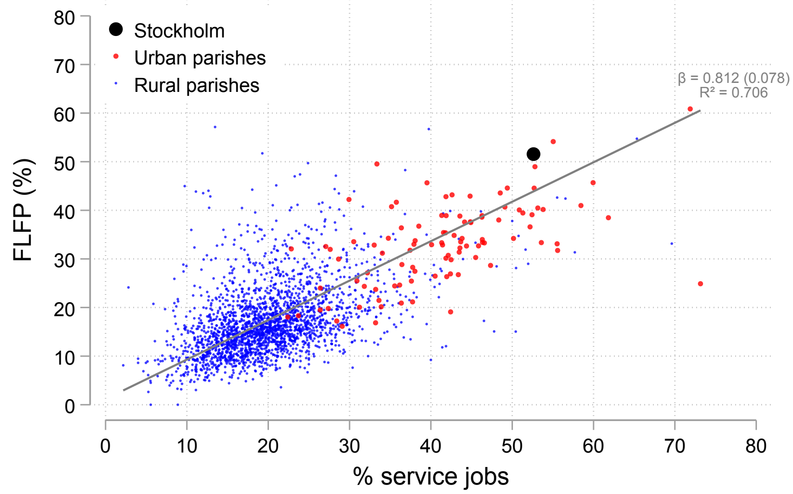


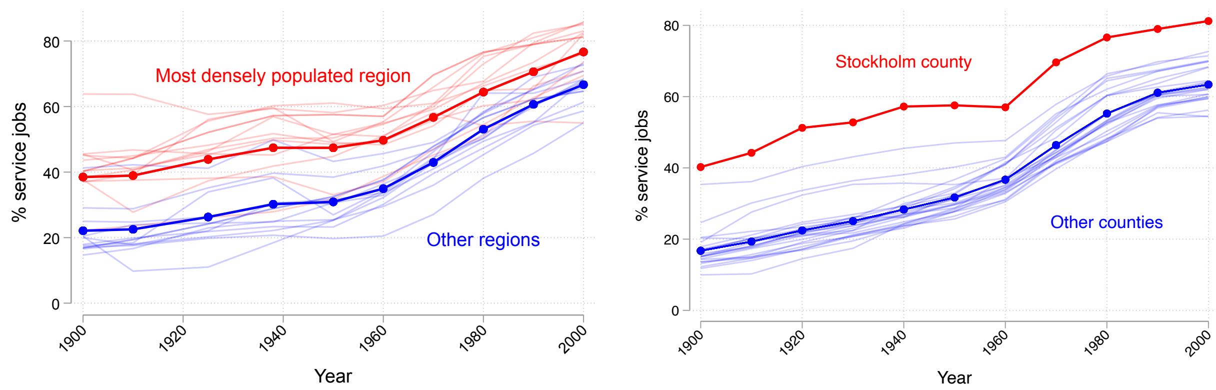
FIGURE 2: FLFP AND SERVICE JOBS ACROSS SWEDISH PARISHES, 1910.

Notes: This figure displays FLFP and the share of prime-aged (20–55) men and women that are employed in services (HISCO major groups 0–5) across 2,400 Swedish parishes in the 1910 census. For presentational purposes, we exclude a small number of military and congregational/religious parishes and aggregate the parishes that constitute the city of Stockholm. We separately report data for urban and rural parishes. The figure also displays a best-fit line and its estimated slope and robust standard error based on a parish-level regression weighted by parish populations.

and urban parishes in Sweden. Urban areas including Stockholm exhibited substantially higher FLFP rates compared to rural areas and there is a tight link between the share of service jobs in a parish and the employment rate of women. The reported parish-level regression indicates that a 10 percentage point increase in service jobs is associated with an 8 percentage point increase in the FLFP rate, while more than 70 percent of the variation in FLFP rates across parishes is accounted for by local differences in service jobs.

The early shift toward the service sector in more populous places is also evident across regions in Europe. Figure 3, 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.[9]

[9]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 nearly all European countries in our data.



(A) European Regions

(B) Swedish Regions

Figure 3: 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).

2.2 Historical context: Stockholm and urban migration

Around the mid-19th century, Sweden was a predominantly agricultural economy with about 90 percent of the population living in rural areas. Industrialization accelerated in the 1870s, which led to rapidly growing incomes and labor demand from urban areas. The next decades saw large numbers of migrants leaving the countryside for the city facilitated by the expanding railroad network (Berger and Enflo, 2017). Although most cities remained comparatively small, the capital Stockholm expanded into a major urban area due to massive immigration: around the mid-century it had less than 100,000 inhabitants, which by 1910 had more than tripled to about 342,000 inhabitants. More than half of Stockholm's population were migrants, a majority of whom were women.

Stockholm's growth was fueled by industries such as textiles, shoe making, and printing, creating a strong demand for semi-skilled and unskilled labor in manufacturing (Karlsson, 1996). 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. In fact, while the manufacturing sector was initially considered the main pull-factor for migration to Stockholm, it was replaced by the fast-growing service sector after the turn of the century (Johansson and Persson, 2004).

During this time, both men and women typically left their parental home at a young age to take up work and accumulate savings before forming a family (Dribe, 2000). This practice, known as life cycle service, was common across socio-economic classes, with migration of young individuals being the norm regardless of gender. This holds true in our sample, as 60 percent of women and 55 percent of men aged 30–46 had migrated from their parish of origin by 1910. 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 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.

Although many urban migrants were employed in relatively low-skill services jobs, they often offered advantages relative to agricultural work. For example, 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).

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 3 shows that Stockholm county 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.[10]

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 Table A.1. By contrast, 50–60 percent of women were married and had at least one child in rural areas and smaller cities. While the differences between Stockholm natives and in-migrants were modest, marriage rates were slightly higher among in-migrants. Similar patterns emerge for fertility and LFP. These patterns align with the argument that opportunities for independent wage labor contributed to Stockholm's low marriage rates, since women could establish financially stable lives on their own without relying on a male household head (Kyle, 1987).

3 Data and sample

Our main data comes from the full-count Swedish decennial censuses between 1880 and 1910 that we use to create a linked sample of women that we can follow from childhood into adulthood. 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 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, and family relationships between members of households. It also distinguishes between 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 into 12 broad groups (Leeuwen and Maas, 2011).[11] We use IPUMS's defi-

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

[11] The 12 HISCLASS groups are: 1) Higher managers; 2) Higher professionals; 3) Lower managers; 4) Lower professionals, clerical and sales personnel; 5) Lower clerical and sales personnel; 6) Foremen; 7) Medium-skilled

nition to measure FLFP: 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 (for example, ”worker's widow”), and is above 15 years old.[12] Using data on the family in 1880, we collect data on fathers' occupations, mother's labor force participation, as well as a number of characteristics at the household level: number of generations in the family, number of siblings, presence of extended family, and presence of non-relatives.

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.[13] 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.[14] In a second step, income scores are adjusted by county-level CPI to account for regional cost differences, which are further adjusted for urban-rural price differences across counties, based on Collin (2016). A Stockholm-specific cost index is applied to residents of Stockholm. 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 percent, respectively.

Linking procedure. In order to identify siblings and other family information, we focus on children observed in their childhood home in 1880 and link them to the 1910 census when they

workers; 8) Farmers and fishermen; 9) Lower-skilled workers; 10) Lower-skilled farm workers; 11) Unskilled workers; and 12) Unskilled farm workers.

[12]In Section 5.1, we examine the sensitivity of our results to a broader definition of FLFP including informal labor.

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

[14]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.

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 E.2.

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.[15] 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.[16] 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×birth year×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 score above a specified threshold for both the first and the last name, additionally requiring that there is no close runner-up.

In terms of linkage rates, we are able to identify 66 percent of individuals born 1864–1880 in the 1910 census back to the 1880 census.[17] 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 U.S. censuses 1910–1940, and Modalsli (2017) links 37% in Norway.[18] Our linkage rates are highly similar for men and women,

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

[16] We follow Wisselgren et al. (2014b) in handling different types of surnames (family and patronymic). See Appendix E.2 for details.

[17] The forward linkage rate is 43 percent.

[18] 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.

at 65 and 67 percent, respectively.

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. (2014b), 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 that we describe in more detail in Section 5.5 (Buckles et al., 2023).

Linked census samples. We focus on linked individuals (1880 to 1910) aged 0–16 in 1880 who reside with their father and mother in rural parishes at least five kilometers from a city. We denote this full sample of 274,711 women as the "full sample from rural origins". In our main analysis, where we compare siblings moving to different destinations, we restrict attention to individuals that have an identified sister living in the same family in 1880 and that have left their rural origin parish by 1910. These restrictions leave us with 70,512 women from 31,137 families. We denote this sample as the "migrant sisters sample". Summary statistics for different samples are presented in Appendix Table A.3. In terms of childhood characteristics measured in 1880, individuals with a same-sex sibling have similar characteristics as those without, although they are more likely to be of higher birth order.

While we are able to achieve relatively high 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 E.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.17).

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.

4 Empirical strategy

We start our analysis by estimating outcomes of migrants to destinations of different population size, using the full sample from rural origins, 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 family f. $Migrant_{if}^p$ is a binary variable taking the 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. $\mathbf{X}_{if}$ is a matrix of individual-level controls: fixed effects for birth year, birth order, being the eldest sister, and having a disability. $\mathbf{W}_f$ is a matrix of household-level controls measured in 1880: 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, indicators of whether the household consists an extended family (relatives only) or a composite household (family and non-relatives), as well as whether the household 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 relative other destinations. To do so, we estimate:

$$Y_{if} = \beta_1 Sthlm\ Migrant_{if} + \beta_2 Other\ Urban\ Migrant_{if} + \gamma \mathbf{X}_{if} + \delta \mathbf{W}_f + \varepsilon_{if} \tag{2}$$

where $Sthlm\ Migrant_{if}$ is an indicator taking the 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. Thus, β_1 captures the difference in outcomes between migrants to Stockholm and migrants to rural parishes.[19]

While the above models controls for a number of individual- and family-level pre-migration variables that might influence both the decision to migrate and labor-market outcomes, there may still be many unobserved factors that can lead to bias. Therefore, we limit the analysis to comparing outcomes between sisters who live in the same family in 1880. By restricting attention to within-family 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 models become:

$$Y_{if} = \sum_{p=1}^{100} \beta^p Migrant_{if}^p + \gamma \mathbf{X}_{if} + \phi_f + \varepsilon_{if} \tag{3}$$

and

$$Y_{if} = \beta_1 Sthlm\ Migrant_{if} + \beta_2 Other\ Urban\ Migrant_{if} + \gamma \mathbf{X}_{if} + \phi_f + \varepsilon_{if} \tag{4}$$

where ϕ_f is a fixed effect common for siblings from family f, which we refer to as a "family fixed effect". Since the family fixed effect absorbs any variation that does not vary within the family, the family-level controls are subsumed in these specifications. All standard errors are cluster-robust at the childhood family level.[20]

The within-family model absorbs any selection in terms of family characteristics and control for observable individual characteristics included in Figure A.2. However, migrants to Stockholm may also be inherently different from their siblings migrating to other destinations, for example in terms of ability and preferences for wage work or marriage. In Section 5.4, we examine the role of such sorting showing that it seemingly does not explain our main results.

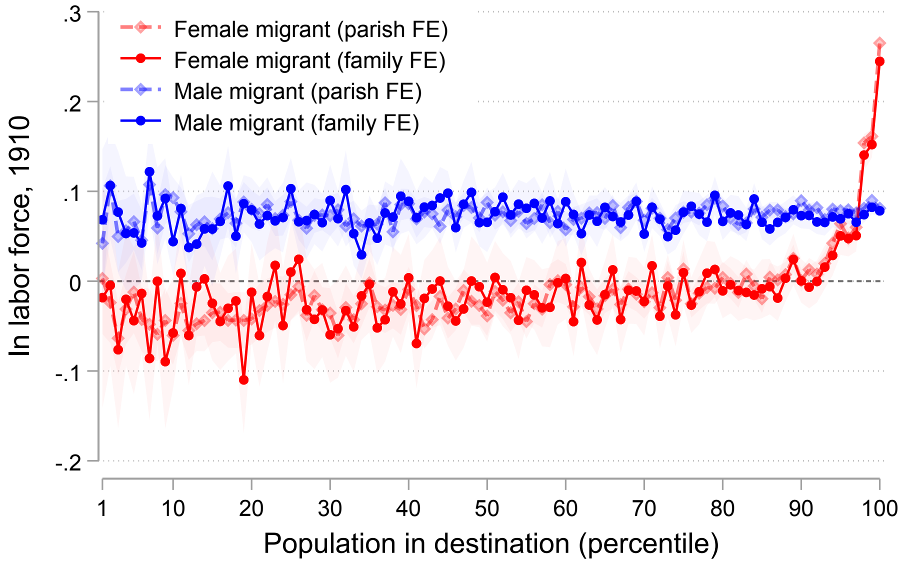


FIGURE 4: LFP BY POPULATION IN MIGRANT DESTINATION

Notes: This figure displays OLS estimates of Equations (1) and (3) using the full sample consisting of individuals born in rural parishes who have either migrated by 1910 or remain in the parish of origin (the omitted category). The outcome is an indicator taking the value one if an individual is in the labor force in 1910. The figure plots point estimates and 95 percent confidence intervals capturing the effect of moving to different parishes where the about 2,400 parish destinations are grouped into percentiles based on their population size. Dashed shaded lines denote a specification using origin parish fixed effects and family- and individual-level controls, while solid lines denote estimates that include family fixed effects with individual-level controls. See Section 4 for full list of control variables. Standard errors are clustered at the family level.

5 Results

5.1 Migration and FLFP

This section presents our main results documenting a non-linearity in the effects of migration on FLFP, with substantial increases particularly among those moving to the largest cities and Stockholm—the largest city—in particular. We show that the relative increase in LFP among migrants to large cities is solely a female phenomenon, as male migrants experience no additional gain from moving to a large city. We also show that the estimated increases in FLFP among migrants to large cities persist over the life cycle.

[19]Alternative specifications include a dummy for *ever* having migrated to Stockholm or another urban area by 1910, thus including temporary migrants. We also study two alternative sample definitions: (1) including stayers who remain in the childhood parish and (2) including only urban migrants. These changes yield results consistent with our main specification.

[20]Appendix Table A.18 displays our main results using alternative standard error calculations.

Main results. Figure 4 displays estimates of Equations (1) and (3) where we examine differences in FLFP in 1910 among migrants from rural origins to about 2,400 different parishes compared to stayers. As described above, we rank all destinations by their population and estimate the effect of moving to each percentile of the parish size distribution when excluding and including family fixed effects, respectively. The figure shows that female migrants moving to the most populous parishes are considerably more likely to be in the labor force compared to stayers or migrants moving to more sparsely populated areas. While FLFP is similar among stayers and migrants to parishes up to the 90^{th} population percentile, there is a marked non-linearity in the top with FLFP among migrants to the most populous destinations being almost 30 percentage points higher. Notably, estimates are similar when excluding or including family fixed effects indicating limited sorting based on family background.

Motivated by the non-linearity in FLFP with respect to population size among migrants, we turn to Equation (2) and estimate the effect of moving to Stockholm on FLFP by comparing the outcomes of rural-born migrants.[21] Table 1, column 1, shows that female migrants to Stockholm are 53.6 percentage points more likely to be in the labor force compared to female migrants to rural origins. The increase is substantial considering that the FLFP rate in the sample is 21 percent. In line with the evidence in Figure 4, migrants to other urban areas experience much smaller increases: 16.4 percentage points compared to migrants moving to rural areas.[22] To account for potential individual and family characteristics that may be correlated with both migration and FLFP, column 2 adds a rich set of individual- and family-level controls, as described in Section 4. These additions yield very similar estimates indicating modest sorting across destinations based on observable individual- and family-level characteristics.

We next use our sibling design and compare the FLFP of sisters who are observed in the same family in 1880 but that had moved to different destinations by 1910. Table 1, column 3

[21]Our main sample is restricted to migrants from rural origins to isolate the role of the city rather than migration itself. Appendix Table A.6 documents consistent results when using an extended sample including stayers, as well as a smaller sample using only urban migrants.

[22]We present additional estimates in Appendix Figure A.4 where we include separate indicators for moving to each city that had more than 20,000 inhabitants in 1910. The increase in FLFP among female migrants to other cities than Stockholm is broadly similar. Interestingly, the absence of an effect on FLFP among migrants to the city of Eskilstuna—by far Sweden's most industrial city by the early 20^{th} century (Berger and Jensen, 2023)—is consistent with our evidence below of a key role of the services sector in accounting for the higher FLFP rates among migrants in large cities.

TABLE 1: LABOR FORCE PARTICIPATION BY MIGRANT DESTINATION

Dependent variable:	In labor force (=1)							
Sample:	Female migrants		Migrant sisters				Migrant siblings	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Stockholm (=1)	0.536***	0.543***	0.533***	0.539***	0.501***	0.496***	-0.026	
	(0.005)	(0.005)	(0.007)	(0.008)	(0.011)	(0.011)	(0.062)	
Other urban area (=1)	0.164***	0.156***	0.171***	0.161***	0.136***	0.133***	-0.018	
	(0.003)	(0.003)	(0.004)	(0.004)	(0.006)	(0.006)	(0.026)	
Stockholm $\times$ Female (=1)							0.482***	0.474***
							(0.023)	(0.023)
Other urban area $\times$ Female (=1)							0.124***	0.121***
							(0.010)	(0.010)
Individual controls	No	Yes	No	Yes	No	Yes	Yes	Yes
Family 1880 controls	No	Yes	No	Yes	No	No	No	No
Family FE	No	No	No	No	Yes	Yes	Yes	No
Family-by-destination FE	No	No	No	No	No	No	No	Yes
Observations	164530	164530	70512	70512	70512	70512	53078	53078
Mean outcome	0.206	0.206	0.226	0.226	0.226	0.226	0.580	0.580

Notes: OLS regressions. The sample in columns 1–2 consists of rural-born women who were aged 0–16 in 1880 and have left their parish of origin by 1910, while the sample is restricted to women who have at least one sister in columns 3–6 (the migrant sisters sample). Columns 7–8 display results for a sample of migrants including both men and women with at least one sibling in the same migrant destination. *In labor force* is an indicator variable taking the value one if the individual is in the labor force in 1910. *Stockholm* is an indicator taking the value one if the individual lives in Stockholm city in 1910. *Other urban area* is an indicator taking the 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, being the eldest sister, and having a disability in columns 1 to 6, as well as sex in columns 7 and 8. *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 fixed effects for mother's employment, father's 1-digit HISCO occupation, and whether the household is multigenerational, extended family (relatives only), or composite family (with non-relatives). *Family fixed effects* is a fixed effect for originating from the same family in 1880. *Family-by-destination fixed effects* is a fixed effect for originating from the same family in 1880 and the same parish of residence in 1910. Standard errors, in parentheses, are clustered at the family level. *** - $p < 0.01$, ** - $p < 0.05$, * - $p < 0.1$.

first reproduces the results from column 2 using the migrant sisters sample, displaying nearly identical coefficients. Column 4 adds the full set of individual- and family-level controls. Again, adding these controls has a minor effect on the estimated increase in FLFP among migrants to Stockholm and other urban areas. Column 5 then introduces family fixed effects as in Equation (4). Absorbing all shared individual- and family-level characteristics that may influence both migration status and FLFP yields an estimated increase in FLFP among those migrating to Stockholm of 50.1 percentage points. Comparing this estimate to that without family fixed effects in column 3 indicates that migrants are somewhat positively sorted on unobservables in terms of across-family characteristics. The difference is small, however, with a reduction in the estimate of only 6.2 percent. Adding individual controls to the regression with family fixed effects in the last column, decreases the estimate only slightly, suggesting that sorting on our set of observable individual characteristics *within* families is modest.

Male vs. female migrants. Our findings highlight that the gains in FLFP were substantially larger for women moving to the largest cities such as Stockholm, yet these results stand in stark contrast to the relationship for male migrants. Figure 4 displays estimates of Equations (1) and (3) showing the estimated increase in LFP for male migrants across the parish size distribution. In contrast to female migrants, the relationship between migration and LFP is always positive for males but remains flat across the population distribution.

Focusing on Stockholm migrants, column 7 of Table 1 highlights the capital's differential role for male and female labor force participation. By interacting our indicator for Stockholm migration with a female indicator, we show that the substantial relative gains in LFP among migrants to large cities is entirely a female phenomenon. Although the coefficient on our Stockholm indicator is insignificant in this specification, the interaction with the female indicator suggests that women increased LFP with 48.2 percentage points relative to men.[23]

As noted above, a potential concern is that the results could be driven by within-family sorting across destination. To address this, we introduce a family-by-destination fixed effect in column 8, restricting the comparison to sisters and brothers within the same final destination. This adjustment has only a modest effect on our coefficient of interest, suggesting that within-family sorting is limited, provided it does not vary by sex. We examine this issue in greater detail in Section 5.4.

Women's formal and informal work. Our main focus is on women's transition into formal employment and we therefore use the census definition of FLFP: a measure that to a great extent captures full-time work in the paid market, outside the home (Stanfors, 2014). However, it is useful to examine the sensitivity of our results to adjusting FLFP rates for potential underreporting of women's informal work (Goldin, 1990). We first examine the sensitivity of our results by excluding those for whom FLFP was least likely to be precisely reported (Chiswick and Robinson,

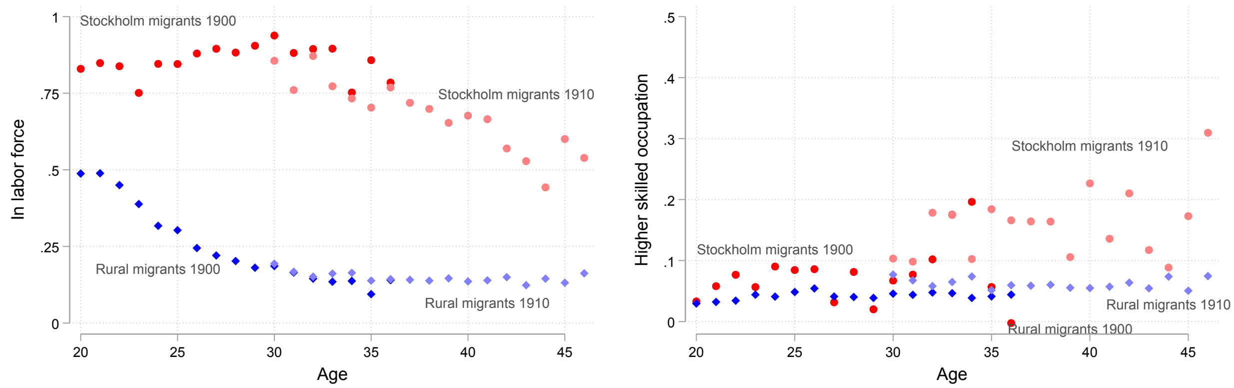
[23] In Appendix Table A.5, we show consistent results for a male only sample.

2021): women who in 1910 lived with a male household head (either her father or husband) who was engaged in agriculture. We then create a more inclusive measure of FLFP that counts women living with a male household head who is either engaged in farming or is a working proprietor in sales or services as in the labor force. Appendix Table A.4 shows a large increase in FLFP among Stockholm migrants even when adjusting the sample or when expanding the FLFP measures to include informal work. While we deliberately focus on women's transition into formal employment, it is reassuring that our results are similar when making conservative adjustments of our FLFP measure by including the informal work of mainly rural women.

Migration and FLFP over the life cycle. We next document that the differences in FLFP largely persist over the life cycle. Appendix Table A.10 studies persistence by examining FLFP rates in 1910 among those who migrated by 1900, thus allowing migrants to relocate in the intervening 10-year period. Even after 10 years, migrants to Stockholm are 34 percentage points more likely to be in the labor force. The estimates are similar when instead considering the effect of *ever* having migrated to Stockholm between 1890 and 1910. Taken together, the results in Appendix Table A.10 show that Stockholm migrants have persistently different labor-market outcomes than other migrants, even when allowing for later relocation decisions.[24]

To further explore how FLFP rates evolve over the life cycle of migrants, Figure 5 presents 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) that include the full set of individual controls and family fixed effects. Panel A indicates that while the FLFP rates of migrants to rural areas quickly decrease between 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 FLFP begin to decrease somewhat, but even by their mid-40s participation rates are substantially higher compared to their sisters migrating to rural areas. The FLFP rate of Stockholm migrants thus remains persistently higher across the life cycle.

[24]At the same time, the smaller estimates also indicate that temporary Stockholm migrants at least to some extent reverted towards the work decisions of migrants to other destinations. To closer study the outcomes of temporary migrants, Appendix Table A.11 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 in 1910 compared to other migrants, though estimates are considerably smaller.



(A) In labor force

(B) Higher skill occupation

Figure 5: FLFP and skills over the life cycle

Notes: This figure displays binned scatter plots using an extended migrant sisters sample which combines data on migrants in 1900 (when women in the sample are aged 20–36) and 1910 (when they are aged 30–46). The outcome is an indicator for being in the labor force (Panel A) and having a high- or medium-skilled occupation according to HISCLASS (Panel B) for female migrants to Stockholm and rural areas between ages 20 to 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 family fixed effects and individual controls (except birth year) following the method of Cattaneo et al. (2024). See Section 4 for full list of control variables.

5.2 Services, skills, and health

This section shows that the rise in FLFP among migrants to Stockholm is driven primarily by transitions into service-sector employment. Although the employment gains are concentrated in low-skill service jobs, we also find increases across the skill distribution, and migrants moved into positions that, on average, paid higher wages. Finally, the economic gains for migrant women are not offset by increased mortality associated with moving to a large city.

5.2.1 Migration and employment: the role of the service sector

We first show that the increases in FLFP among migrants to the most populous areas and Stockholm are mainly driven by an increased probability of work in the services sector. Figure 6 displays estimates from Equation (1) and (3) where the outcome is an indicator for working in services (HISCO major groups 0–5) or industry (7–9). As above, we report separate coefficients for migrants moving to different parishes ranked by their population, which capture the increase in the probability of being employed in a given sector compared to stayers. Increases in employment in services

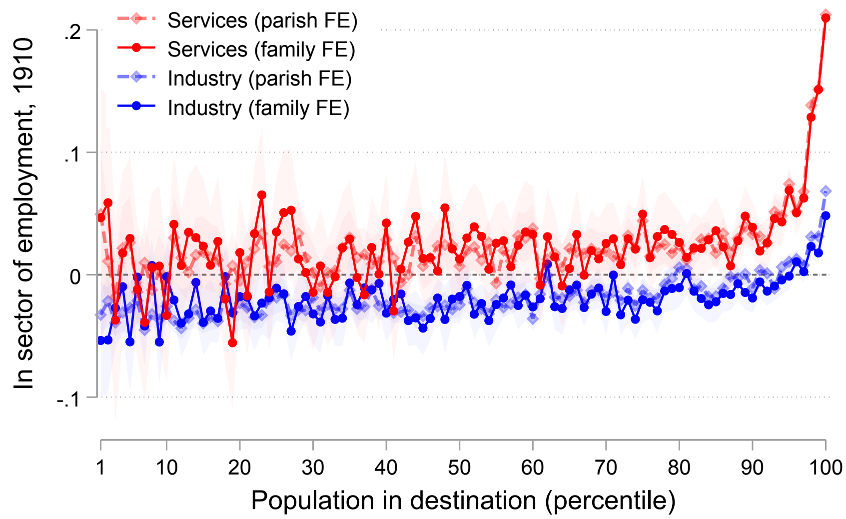


FIGURE 6: SECTOR OF EMPLOYMENT BY POPULATION IN MIGRANT DESTINATION

Notes: This figure displays OLS estimates of Equations (1) and (3) using the full sample consisting of individuals born in rural parishes who have either migrated by 1910 or remain in the parish of origin (the omitted category). The outcome is an indicator taking the value one if an individual is employed in the services or industrial sector in 1910, respectively. The figure plots point estimates and 95 percent confidence intervals capturing the effect of moving to different parishes where the about 2,400 parish destinations are grouped into percentiles based on their population size. Dashed shaded lines denote a specification using origin parish fixed effects and family- and individual-level controls, while solid lines denote estimates that include family fixed effects with individual-level controls. See Section 4 for full list of control variables. Standard errors are clustered at the family level.

exhibit the same non-linear pattern as the effect on FLFP displayed in Figure 4: migrants are not more likely to work in the service sector in destinations below the 90^{th} percentile, but are about 20 percentage points more likely to transition into service work in the most populated areas.

Table 2, column 1, presents estimates of Equation (4) showing that the probability of working within the service sector is 42.4 percentage point higher for women migrating to Stockholm compared to their sisters moving to rural areas, an almost threefold increase compared to the sample mean. Migrants to smaller urban areas also see an increased rate of employment in services, though the effect for Stockholm is about five times as large.

Although the service sector is key in explaining increases in FLFP, migrants also experienced more modest increases in industrial employment.[25] Table 2, column 2, indicates that the probability of employment in industry increases by 7.5 percentage points among migrants to Stockholm.

[25] Appendix Table A.19 breaks down the sector of employment further according to the HISCO major groups. Apart from agriculture, which unsurprisingly displays a negative relationship to Stockholm migration, we find positive estimates for all other occupational groups.

TABLE 2: SECTOR, SKILL, AND INCOME BY MIGRANT DESTINATION

Dependent variable:	Sector of employment		Occupational skill		Income score		Age at death	
	Services	Industry	High	Low	ln(income)	Pct. rank	Female	Male
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Stockholm (=1)	0.424***	0.075***	0.068***	0.426***	0.072***	7.333***	0.177	-2.315***
	(0.010)	(0.006)	(0.008)	(0.010)	(0.026)	(1.481)	(0.359)	(0.351)
Other urban area (=1)	0.086***	0.043***	0.018***	0.106***	-0.174***	-12.732***	0.225	-1.140***
	(0.005)	(0.003)	(0.004)	(0.005)	(0.020)	(1.314)	(0.204)	(0.241)
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Family FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	70512	70512	70512	70512	6514	6514	52881	44600
Mean outcome	0.161	0.053	0.078	0.142	7.004	53.540	73.303	71.900

Notes: OLS regressions using the migrant sisters sample in columns 1 to 7 and a similar migrant brothers sample in column 8. All outcomes are measured in 1910. *Services* and *Industry* are indicators taking the value 1 for the HISCO major groups 0–5 and 7–9, respectively. Similarly, *High* corresponds to high- and medium-skilled occupations (HISCLASS groups 1–4 and 6–8), while *Low* corresponds to low-skill and unskilled occupations (HISCLASS groups 5 and 9–12). *Stockholm* is an indicator taking the value 1 if the individual lives in Stockholm city in 1910, and zero if not. *Other urban area* is an indicator taking the value 1 if the individual lives in an urban area other than Stockholm in 1910, and zero if not. *Individual controls* include fixed effects for birth year, birth order, being the eldest sister, and having a disability. *Family fixed effects* is a fixed effect for originating from the same family in 1880. Standard errors, in parentheses, are clustered at the family level. *** - $p < 0.01$, ** - $p < 0.05$, * - $p < 0.1$.

However, in contrast to the discontinuous increase in service sector employment in the most populous areas, the increase in industrial work displays a more continuous increase throughout the population distribution in Figure 6, which is also evident in the relatively large coefficient for migrants to other urban areas in Table 2, column 2.

Together, these results document the key importance of the services sector in accounting for the greater FLFP among female migrants to large cities such as Stockholm. We further corroborate this interpretation in Appendix Table A.20 by documenting similar increases among female migrants to Stockholm relative to their migrant brothers in the same location, as well as in Appendix B where we show that a greater local services share in a destination is a central determinant of whether a female migrant transitions into formal employment, while the greater services share accounts for a substantial part of the effect on FLFP of moving to Stockholm.

5.2.2 Migration and employment: skills, income, and mortality

Skill-content of jobs. An important question is whether female migrants to Stockholm only transitioned into low-skilled service jobs, or if they also took on more qualified employment. Table 2, column 3 shows that Stockholm migrants are 6.8 percentage points more likely to work in higher- and medium-skilled occupations than their migrant sisters, which is a relatively large effect considering that only about 8 percent of the women in the sample work in such jobs. However, as shown in column 4, Stockholm migrants are about 42.6 percentage points more likely to work in low-skilled occupations suggesting that the overall increase in services employment is primarily driven by an increased entry into relatively low-skilled occupations. However, a pattern of employment gains across the skill spectrum is also evident when looking at more detailed occupations.[26]

We next examine how the occupational skills of migrants evolve over the life cycle. Panel B 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. Stockholm migrants exhibit signs of skill upgrading over time. In their 20s, migrants are about as likely to work in higher-skilled occupation in Stockholm as in rural areas, yet in their 30s and 40s Stockholm migrants are considerably more likely to do higher-skilled work. Turning instead to following the same individuals over time, column 4 of Appendix Table A.10 shows that individuals who move to Stockholm in 1900 are on average 7.6 percentage points more likely to work in higher-skilled occupations 10 years later; the estimate for those that *ever* moved to Stockholm is similar. Thus, migrants to Stockholm increase their likelihood to work in both low and higher skilled occupations. Moreover, they experience significant occupational skill-upgrading over time, a pattern not observed among migrants to rural areas.

Income and intergenerational mobility. We next show that while migrants to Stockholm primarily enter low-skill service occupations, they attain jobs that on average pay better compared to their migrant sisters moving elsewhere. Focusing on women with reported occupations, column 5 of Table 2 shows that female migrants have about 7 percent higher real income scores compared to their sisters migrating to rural locations. Column 6 shows that the gain in income corresponds to an increase of about seven percentile ranks in the national income distribution. One worry with using income scores at the individual level is the fact that it conditions on being employed, which

[26]Focusing on the ten most common occupations among migrant women in each skill group, Appendix Figure A.8 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.

is likely to be highly positively selected outside of Stockholm, where female employment is more rare. To account for this, Appendix Table A.21 presents estimates from a regression where the outcome is household income score per adult household member, which includes married women with no reported occupation. Household income increases by about 20 percent for Stockholm migrants, which is not driven by Stockholm migrants matching with higher-income spouses since the effect is even more pronounced when down-weighting the contribution of the spouse.[27]

A related question is whether the income gains of Stockholm migrants also was reflected in higher rates of intergenerational income mobility. We examine mobility differences in Appendix Figure A.12 where we plot the mean income ranks of daughters by their father's rank. Migrants to Stockholm attain higher income ranks conditional on their father's income and also exhibit lower rank-rank associations compared to stayers and migrants to other destinations, which indicates respectively a greater level of absolute mobility and a lower degree of intergenerational persistence. We corroborate a higher level of income mobility among Stockholm migrants in Appendix Table A.22 showing that women that moved to Stockholm attain higher rates of absolute mobility compared to their migrant sisters, also when focusing on sisters born to families at the bottom of the income distribution.

Mortality. Although we have showed that female migrants to Stockholm experience increases in FLFP and income, a remaining question is whether these economic gains are offset by an urban mortality penalty making the welfare gains of migration more ambiguous. Indeed, large cities still had substantially elevated mortality rates compared to rural areas due to overcrowding, contagious diseases, and poor sanitation in the early 20^{th} century (Haines, 2001; Cain and Hong, 2009; Ager et al., 2024). Yet when matching the individuals in our sample to the official death registers we find that female migrants to Stockholm—in sharp contrast to male migrants—did not experience reductions in average lifespans. Appendix Figure A.9 displays the age at death of the female and male migrants in our sample based on the population in their place of residence in 1910: the urban mortality penalty is an entirely male phenomenon, as women's life expectancy is very similar across the population size distribution. To corroborate these findings, Table 2, column 7 shows no significant difference in the age at death of migrants in Stockholm compared to their

[27] In fact, the intra-household ratio of female to male income is higher for Stockholm migrants, suggesting that they experience an increased economic independence relative to rural migrants.

migrant sisters. In contrast, column 8 shows that mortality among male migrants to Stockholm was considerably higher, as they died on average two years younger than their brothers that remained in rural areas.[28] The fact the female migrants transitioned into less hazardous service jobs could be one explanation for the absence of a mortality effect, yet including fixed effects for 5-digit HISCO occupations in column 2 of Appendix Table A.23 does not materially affect the estimates.[29]

Together, these results suggest that the economic gains made by female migrants to Stockholm were not offset through shorter lifespans. More broadly, these findings suggest that the economic benefits of moving were more unambiguous for women compared to men, given that the latter group traded off greater incomes against shorter lives in the city.

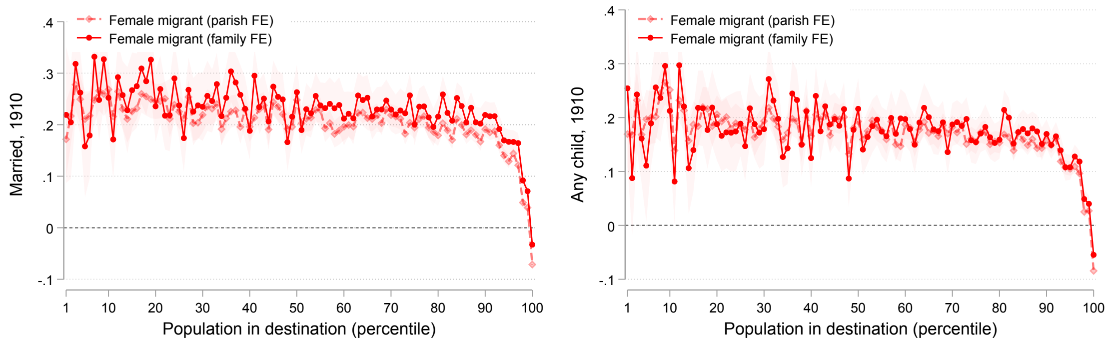
5.3 Migration, marriage, and fertility

Our findings in the previous sections have shown that migrants to Stockholm and other urban areas experience substantial increases in FLFP compared to migrants to rural areas. However, women in the early 20^{th} century typically faced a choice between having a family and being employed, as the vast majority of women exited the labor force upon marriage. We next document that the increases in FLFP among urban migrants were mirrored in large reductions in marriage and fertility.

Marriage and fertility. Figure 7 displays estimates based on Equations (1) and (3) where the outcome is an indicator for being married (Panel A) or having at least one child (Panel B) in 1910. As before, we report separate coefficients for migrant destinations grouped into percentiles based on their population size. Panels A and B show that female migrants moving to the most populous parishes are much *less* likely to marry and have a child than stayers. By contrast, migrants to destinations in the bottom-90 percent of the population distribution are substantially *more* likely to

[28] Appendix Figure A.10 displays the evolution of survival probabilities for female and male migrants over the life course. 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, while 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 percentage points by age 70. 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.

[29] Another potential explanation given that we document reductions in fertility among migrants below is that migration may have reduced the risk to to die in child birth or related complications. However, the estimates in Appendix Figure A.10 show that the mortality differential for female migrants is not apparently related to women's childbearing age. Even for the sample of women who are relatively young and may not have completed their fertility by 1910, Appendix Figure A.11 shows a similar pattern of nonsignificant differences in mortality.



(A) MARRIAGE

(B) ANY CHILD

FIGURE 7: FAMILY FORMATION BY POPULATION IN MIGRANT DESTINATION

Notes: This figure displays OLS estimates of Equations (1) and (3) using the full sample from rural origins, consisting of females born in rural parishes who have either migrated by 1910 or remained in the parish of origin (the omitted category). The outcome is an indicator taking the value on if an individual is married (Panel A) or has any child (Panel B) in 1910. The figure plots point estimates and 95 percent confidence intervals capturing the effect of moving to different parishes where the about 2,400 parish destinations are grouped into percentiles based on their population size. Dashed shaded lines denote a specification using origin parish fixed effects and family- and individual-level controls, while solid lines denote estimates that include family fixed effects with individual-level controls. See Section 4 for full list of control variables. Standard errors are clustered at the family level.

marry and have at least one child, respectively. Estimates are similar when including family fixed effects in both figures.[30]

Table 3 presents the estimated impact on family formation from moving to Stockholm and other urban areas in our sample of migrant sisters.[31] We first present estimates of Equation (2) in columns 1 and 2 where the outcome is an indicator for being married in 1910, while column 3 displays estimates of Equation (4) that includes family fixed effects. The latter estimate indicates that migrants to Stockholm are 51.1 percentage points less likely to be married in 1910 if they migrated to Stockholm, a substantial decrease relative to their sisters moving both to rural and other urban locations. As above, the relatively small reduction in absolute magnitude when comparing estimates obtained with and without family fixed effects, suggests that sorting based on family characteristics was limited in nature.

[30] Appendix Figure A.3 displays the corresponding estimates for the male sample revealing that male migrants were more likely to marry than stayers even in the most populous destinations.

[31] We present additional results in the appendix. First, Appendix Table A.7 presents results for alternative samples. Second, Appendix Table A.15 that the results are not driven by differences in co-habitation practices outside marriage or children born out of wedlock. Third, Appendix Table A.16 documents large reductions in marriage and fertility among Stockholm migrants also compared to their migrant brothers residing in the same destination.

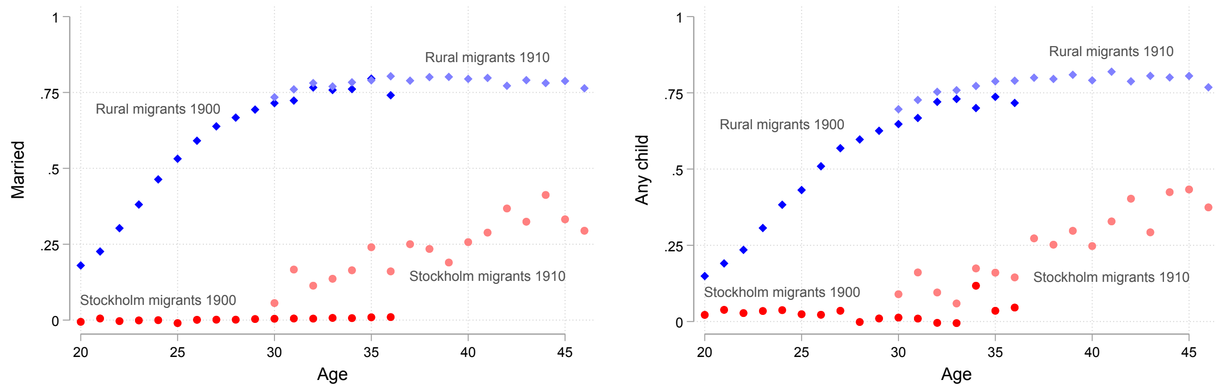
TABLE 3: FAMILY FORMATION BY MIGRANT DESTINATION

Dependent variable:	Married (=1)			Any child (=1)			In labor force (=1)	
							Married	Unmarried
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Stockholm (=1)	-0.546***	-0.545***	-0.511***	-0.525***	-0.518***	-0.476***	0.032***	0.184***
	(0.007)	(0.007)	(0.010)	(0.007)	(0.007)	(0.010)	(0.006)	(0.020)
Other urban area (=1)	-0.158***	-0.146***	-0.123***	-0.150***	-0.133***	-0.107***	0.018***	0.119***
	(0.004)	(0.004)	(0.006)	(0.004)	(0.004)	(0.006)	(0.002)	(0.017)
Individual controls	No	Yes	Yes	No	Yes	Yes	Yes	Yes
Family 1880 controls	No	Yes	No	No	Yes	No	Yes	No
Family FE	No	No	Yes	No	No	Yes	No	Yes
Observations	70512	70512	70512	70512	70512	70512	50170	10112
Mean outcome	0.712	0.712	0.712	0.709	0.709	0.709	0.018	0.744

Notes: OLS regressions. Columns 1 to 6 use the migrant sisters sample. Columns 7 and 8 restrict the sample to women that are married and unmarried in 1910, respectively. All outcomes are measured in 1910. *Stockholm* is an indicator taking the value one if the individual lives in Stockholm city in 1910. *Other urban area* is an indicator taking the 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, being the eldest sister, and having a disability. *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 fixed effects for mother's employment, father's 1-digit HISCO occupation, and whether the household is multigenerational, extended family (relatives only), or composite family (with non-relatives). *Family fixed effects* is a fixed effect for originating from the same family in 1880. Standard errors, in parentheses, are clustered at the family level. *** - $p < 0.01$, ** - $p < 0.05$, * - $p < 0.1$.

Columns 4–6 of Table 3 present estimates where the outcome is an indicator capturing whether a migrant has any children by 1910. The comparison of migrant sisters in column 6 shows that migrants to Stockholm are 47.6 percentage points less likely to have any children. Notably, the reductions in marriage and likelihood of having a child among Stockholm migrants compared to their sisters are both substantial considering that about 70 percent of migrants in the sample were married or had a child by 1910.

Marriage and FLFP. While most female migrants that took up formal employment in Stockholm at the same time remained single, the increases in employment are not solely driven by the lower marriage rates among migrants. Table 3, column 7, reports estimates where we limit the sample to sisters that were all married in 1910. The estimated impact on FLFP is smaller in absolute magnitude due to the lower FLFP among married women, but relative to the sample mean it represents an almost twofold increase. In column 8, 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, which indicates a much higher probabil-



(A) MARRIAGE (B) ANY CHILD

FIGURE 8: MARRIAGE AND FERTILITY OVER THE LIFE CYCLE

Notes: This figure displays binned scatter plots using an extended migrant sisters sample which combines data on migrants in 1900 (when women in the sample are aged 20–36) and 1910 (when they are aged 30–46). The outcome is an indicator for being married (Panel A) or having at least one child (Panel B) for female migrants to Stockholm and rural areas between ages 20 to 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 family fixed effects and individual controls (except birth year) following the method of Cattaneo et al. (2024). See Section 4 for full list of control variables.

ity to enter formal employment than their sisters moving to rural areas. Given that we condition on an outcome that is closely linked to being in the labor force, it is not surprising that this estimate is lower than the unconditional estimates in Table 1.

Next, we address issues of timing of marriage and employment. Is it the case that unmarried women move to Stockholm to work, while married women remain in rural areas? Column 1 of Appendix Table A.8 restricts attention to sisters who were unmarried and had not yet migrated in 1900. Within this group, odd-numbered columns indicate that for those that eventually migrate to Stockholm, we still find similar differences for marriage, fertility, and FLFP. Further restricting the sample to individuals who were unmarried and had already migrated in 1900, we again find large differences in 1910 outcomes depending on destination.

Marriage and fertility over the life cycle. We next document that the differences in marriage rates and child bearing among Stockholm migrants largely persist over the life cycle. Appendix Table A.10 examines marriage and fertility outcomes in 1910 among those that had migrated by 1900, irregardless of whether stayed or moved away in the intervening years. Migrants to Stockholm in

1900 are nearly 38 percentage points less likely to be married or have children even a decade later. The results are similar when estimating the effect of *ever* having migrated to Stockholm between 1890 and 1910.[32]

We then examine the evolution of marriage and fertility across the life cycle of migrants in Figure 8. Panel A shows that while marriage rates increase among rural migrants starting in the early 20s, their migrant sisters residing in Stockholm begin to exhibit modest increases in marriage first roughly ten years after their rural sisters. Panel B documents similar trajectories in terms of fertility measured as having any child. When the sisters in the sample have reached their mid-40s, having surpassed the typical childbearing age, differences in marriage and childbearing rates are substantial.[33] Thus, these results suggest that differences in family outcomes are persistent and cannot be explained simply by delayed family formation.

5.4 Accounting for within-family sorting across destinations

This section examines the potential role of within-family sorting across destinations in contributing to the economic and demographic effects we find among female migrants. First, we consider the importance of sorting based on observable or unobservable individual characteristics and find limited evidence that it drives our main estimates. Second, we develop an IV strategy that exploits plausibly exogenous variation within families in who moved to Stockholm to further alleviate concerns about potential sorting across destinations.

5.4.1 Testing for sorting within families

Our main estimates rely on a sibling design that absorbs all individual- and family-level characteristics shared by sisters, which may influence both their migration decisions and labor market outcomes. While this approach accounts for potential between-family sorting, it raises concerns about the possibility of sorting within families. As discussed above, the evidence in Table 1 suggests that within-family sorting is not related to our set of individual controls and that it would need to be sex-specific in order to bias our results. To further assess its importance, we here pro-

[32]When focusing on those women that had moved to Stockholm by 1900 but left in the next decade in Appendix Table A.11, we find that the estimated reduction in marriage and fertility is smaller in absolute magnitude.

[33]Appendix Figure A.5 shows the distribution of age at first child among women living in Stockholm in 1910 showing that virtually no women had their first child after age 40.

vide three additional tests. Together they suggest that within-family sorting—both in terms of observable and unobservable characteristics—is not a main driver of our results.

First, we examine whether controlling for observable differences in FLFP, marriage, and fertility between sisters *before* migration affects our estimates, which could indicate selection on underlying differences in ability or preferences regarding work and marriage. We focus on the subset of migrants who move after 1900, where we observe their pre-migration outcomes in the 1900 census when they are 20–36 years old. Appendix Table A.12 shows that including these controls accounts for very little of the estimated effect of migration to Stockholm on FLFP or family formation, which is consistent with modest sorting based on observables in these dimensions.[34]

Second, a standard Roy model predicts that the strength of migrant selection should be affected by the fixed costs of moving, which are partly shaped by the distance between the origin and destination. Yet Appendix Figure A.6 shows that the estimated impact on FLFP from moving to Stockholm does not differ by distance to Stockholm, which again indicates limited sorting within families. Next, young adults often chose migration destinations independently, but families may also have strategically have influenced sisters to move to particular destinations based on their individual characteristics. The extent of parental involvement likely varied with family context: farming households were more tied to land, larger sibships may have permitted greater individual discretion, the presence of sons could have reduced constraints on daughters' mobility, and emigration may indicate stronger coordination of migration choices. Yet, as shown in Appendix Figure A.7, the estimated effect of migrating to Stockholm is highly similar across all these subgroups, further suggesting limited within-family sorting across destinations.

Third, to examine the potential role of sorting based on unobservables, we use the method of Oster (2019) that compares estimates 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 family 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 *individual* variation within families. Appendix Table A.14 displays our results for FLFP, marriage, and fertility. All six models indicate that the bias-corrected estimate is nearly equal to or larger than the estimate

[34] Appendix Table A.13 presents additional robustness tests showing that our findings are similar when applying twin rather than family fixed effects to account for additional unobserved within-family differences.

based on observable controls. Thus, we find that there is somewhat *negative* selection within families, correspondingly pulling our estimates towards zero. Our baseline models find small differences between the controlled and uncontrolled regressions.[35] The bias-corrected for FLFP is 53.7 percentage points, close to our main estimate of 49.6. For marriage and childbearing, we find estimates that are again similar to the baseline estimates.[36]

5.4.2 An IV strategy

A remaining explanation is that the estimates also reflect the causal effect of migration to Stockholm. To shed light on this possibility, we propose an IV strategy for women's migration to Stockholm that is based on two sources of variation. First, as shown in columns 1 and 3 of Appendix Table C.3, women are more likely to migrate from the childhood parish or county if they have an older sister, as the eldest sister tended to have a greater household responsibilities and was more likely to take care of elderly parents. Second, as shown in Appendix Figure C.1, those growing up closer to Stockholm more often migrate there later in life, likely due to the lower cost of migration. Based on this, we construct an instrument defined as the *interaction* between having an older sister and the proximity from the childhood parish to Stockholm, while holding constant the direct effects of both variables.[37] The identifying variation is thus based on the idea that the marginal effect of having an older sister on migration to Stockholm is larger the closer one's childhood parish is to the capital. Being a younger sister acts as a push factor, while proximity to Stockholm affects the choice of destination. In other words, the strategy compares outcomes between younger and older sisters at different distances from Stockholm. Because identification relies on within-parish variation, we can include parish fixed effects that flexibly control for proximity to Stockholm and

[35] 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, $\delta = 1$.

[36] Using the 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 56.8 percentage points, which can be compared to the baseline estimate of 51.6.

[37] Controlling for the direct effects of having an older sister and the proximity to Stockholm is important for identification. This is because both variables are unlikely to be valid instruments for migration on their own. For example, if social norms push first-born daughters to stay in the parish of origin, using the indicator for having an older sister as an instrument would compare women who do not only have different propensities to migrate, but who are also differentially constrained by the social norm. Moreover, proximity to Stockholm is likely to be related to a number of personal and family characteristics, such as family wealth and social networks, that can have a direct impact on our outcomes of interest regardless of the migration status. Relying on the interaction term alone avoids this issue.

hold constant all local factors that may influence either migration or second-stage outcomes.

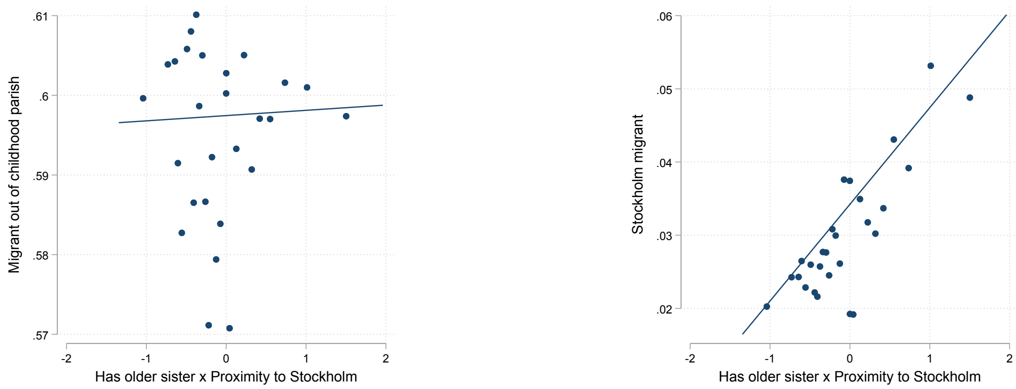
The first-stage equation, estimated on the full sample of women from rural origins, is:

$$\begin{aligned} Sthlm\ Migrant_{ip} =& \beta_1\ Has\ Older\ Sister_i \times Sthlm\ Proximity_p \\ &+ \beta_2\ Has\ Older\ Sister_i + \gamma \mathbf{X}_i + \phi_p + \varepsilon_{ip}, \end{aligned} \tag{5}$$

where $Sthlm\ Migrant_{ip}$ is an indicator for individual i from parish p living in Stockholm in 1910, $Has\ Older\ Sister_i$ indicates that the individual has an older sister, $Sthlm\ Proximity_p$ is defined as minus the log of distance from the childhood parish to Stockholm, and the single instrument is the interaction term $Has\ Older\ Sister_i \times Sthlm\ Proximity_p$. To control for potentially confounding family dynamics related to sibling gender, we include fixed effects for individual birth order, birth year, and number of siblings in $\mathbf{X}_i$. We also include the same set of individual and family controls as in Equation (1). All models include childhood parish fixed effects, ϕ_p. Standard errors are clustered at the level of the childhood parish.

Balance test. The gender of older siblings is, in principle, as good as randomly assigned conditional on birth order. A potential concern arises if parental gender preferences induce correlations between sibling gender and other background characteristics. However, because our identification strategy relies on spatial variation, such preferences would bias our estimates only if they vary systematically with proximity to Stockholm. Reassuringly, the flexible tests reported in Appendix Table C.1 show that, conditional on birth order, proximity to Stockholm does not predict the likelihood of having an older brother or sister. Appendix Table C.2 then shows a balance test of the instrument against the covariates from Equation (1), conditional on having an older sister and fixed effects for birth order, birth year, number of siblings, and childhood parish. The results in column 1 indicates no significant correlations at the 5-percent level and one marginally significant correlation with the childhood household including non-relatives.[38] In column 2, balance tests for a placebo instrument using older brothers shows a significant correlation for one out of fourteen variables.

[38]Below, we show that including the full set of family controls, including household type, slightly magnifies IV estimates.



(A) ANY MIGRATION

(B) MIGRATION TO STOCKHOLM

FIGURE 9: THE INSTRUMENT PREDICTS MIGRATION TO STOCKHOLM BUT NOT OVERALL MIGRATION

Notes: This figure shows the relationship between the instrument and an indicator for migrating out of the childhood parish (Panel A) and migrating to Stockholm (Panel B) by 1910. Graphs are produced using *binsreg* with 50 bins (Cattaneo et al., 2024). Values of the instrument above the 99th percentile are excluded for expositional reasons.

Exclusion restriction. The key identifying assumption is that the instrument satisfies the exclusion restriction: the instrument should only affect second-stage outcomes by raising the propensity to migrate to Stockholm, conditional on the set of controls. An important requirement for this to hold is that norms regarding migration of younger sisters should not depend on the proximity to Stockholm. For example, if younger sisters who grow up closer to Stockholm have more freedom to migrate than those growing up farther away, the instrument would capture variation in overall migration, rather than migration *destination* as intended. To test for this, Appendix Table C.3 shows the relationship between the instrument and indicators for any migration: whether one has moved out of the childhood parish or county by 1910. Columns 2 and 4 show that this association is close to zero and statistically insignificant. Panel A of Figure 9 illustrates this result in a binned scatterplot, showing no apparent relationship across the instrument's distribution.

Below, after presenting the main results, we consider several further potential threats to identification and the exclusion restriction.

IV estimates. Panel B of Figure 9 illustrates the first-stage relationship, with observations revealing a clear positive association across the full range of the instrument. Column 1 of Panel A in Table 4 confirms a strong and statistically significant first-stage at the 1 percent level. Quanti-

TABLE 4: IV RESULTS

Dependent variable:	Sthlm migrant	In labor force	Married	Any child
Panel A: Main results	OLS (1)	IV (2)	IV (3)	IV (4)
Prox to Sthlm × Has older sister	0.012*** (0.002)			
Stockholm (=1)		0.432** (0.174)	-0.491** (0.202)	-0.302 (0.210)
First-stage F-statistic		43.6	43.6	43.6
Panel B: Placebo instrument	OLS	OLS	OLS	OLS
Prox to Sthlm × Has older sister	0.012*** (0.002)	0.006** (0.002)	-0.006** (0.003)	-0.004 (0.003)
Prox to Sthlm × Has older brother	0.003* (0.002)	-0.003 (0.002)	0.001 (0.003)	-0.000 (0.003)
Observations	274711	274711	274711	274711
Mean outcome	0.036	0.177	0.693	0.698

Notes: OLS and IV regressions using the full sample from rural origins. Panel A displays first-stage (column 1) and second-stage (column 2–4) results. Panel B displays reduced-form results for the main and placebo instruments. The instrument is defined as the interaction between having an older sister and childhood parish proximity to Stockholm. The placebo instrument is constructed using older brothers instead of older sisters. All models include an indicator for having an older sister (Panels A and B), having an older brother (Panel B), as well as fixed effects for birth order, birth year, number of siblings, having a disability, and childhood parish. The following family controls are also included: father's income score percentile, family size, as well as fixed effects for mother's employment, father's 1-digit HISCO occupation, and whether the household is multigenerational, extended family (relatives only), or composite family (with non-relatives). Standard errors, in parentheses, are clustered at the childhood parish level. *** - $p < 0.01$, ** - $p < 0.05$, * - $p < 0.1$.

tatively, for those who have an older sister, growing up 200 kilometers closer to Stockholm raises the likelihood of migration by 1.3 percentage points, a sizable increase relative to the baseline migration rate of 3.6 percent.

Column 2 of Table 4 shows the IV estimate for our main outcome, indicating that migration to Stockholm increases FLFP by 43.2 percentage points. This can be compared to the estimate in Table 1 showing an increase of 49.6 percentage points.[39] The strong first stage is reflected in the F-statistic of 43.6. For family outcomes, column 3 and 4 show a corresponding decrease in the marriage probability of 49.1 percentage points and an insignificant decrease in childbearing by 30.2 percentage points. Together, the IV results for FLFP and marriage are in line with the OLS estimates obtained using the sibling design, although the relationship between migration and

[39]The corresponding OLS estimate using the same sample as in column 2 is 51.5 percentage points.

childbearing is smaller in magnitude and less precisely estimated.[40]

Placebo and exclusion restriction tests. Parents with strong gender preferences may influence the gender composition of their children. If so, having an older sister could correlate with unobserved parental characteristics that drive the results in Table 4. To assess this possibility, we implement a placebo test that uses an alternative instrument based on having an older *brother* rather than an older sister. If the original instrument is valid, the placebo should have little or no effect, as migration should primarily be driven by older sisters. By contrast, if unobserved confounders explain the IV results, the placebo instrument should yield effects in the opposite direction, since any characteristics correlated with having an older sister would typically be negatively correlated with having an older brother. Panel B of Table 4 reports the first-stage and reduced-form estimates for migration and our three main outcomes. Reassuringly, the older-brother placebo shows near-zero associations with the main outcomes and a small, marginally significant relationship with migration. Moreover, Column 2 of Appendix Table C.6 shows that the IV estimates slightly increase when including the placebo instrument as a control.

As a second placebo test, we apply the older-sister instrument to the male sample. The logic is straightforward: if the IV results for females reflect correlations between the instrument and unobserved parental characteristics, this should be reflected in males' outcomes as well. Appendix Table C.7, however, shows no significant effects of the instrument on men's migration or on the main outcomes.

Taken together, the placebo tests indicate that the instrument based on older sisters, rather than older brothers, has meaningful effects only for females. This pattern aligns with the hypothesized mechanism of increased migration to Stockholm and is difficult to reconcile with unobserved confounders driving results.

We next consider several further threats to the exclusion restriction assumption. First, it may be that having an older sister is fundamentally different depending on proximity to cities. While Figure 9 and Table C.3 showed that overall migration propensities are the same for younger sisters

[40] We provide additional results in Appendix Section C. In Table C.4, we use an indicator for *ever* having migrated to Stockholm as the endogenous variable, with results in line with those for ever-migration discussed in Section 5.1 and Table A.10. IV estimates for sector of work, skills, income and health are presented in Appendix Table C.5, with results that are similarly consistent with previous estimates.

regardless of proximity to Stockholm, we can also directly test for this potential confounder along the urbanization dimension by including an interaction between having an older sister and proximity to the nearest of the ten largest cities. Column 3 of Table C.6 shows that our results are robust to this inclusion.

Another potential issue is that having an older sister correlates with having a higher birth order. If younger siblings tend to have different outcomes by proximity to Stockholm, perhaps due to underlying economic environments, our results may be driven by such effects rather than by migration alone. Column 4 of Table C.6 therefore includes the interaction of birth order and proximity to Stockholm as a control. This specification is demanding from a data point of view, as it is very close to the variation that drives the instrument. This is reflected in the lower F-statistic and less precise estimates. Point estimates are nonetheless very close to the main IV results in column 1. Thus, gender of older siblings, rather than birth order, are driving the results.

While our baseline models include the full set of family controls, column 1 of Appendix Table C.8 shows that results are similar without them. Column 2 then adds the full set of family controls, including father's occupation fixed effects. The inclusion of these controls magnify our results somewhat, indicating that potential confounding variation picked up by the instrument tends to push estimates towards zero, if anything. Column 3 excludes parishes at a distance of above 510 kilometers from Stockholm, corresponding to the 90^{th} percentile. Lastly, column 4 excludes parishes located north of the city of Gävle, considered the beginning of the northern region of Norrland. These changes yields similar results, indicating that our results are not based on variation from highly rural or remote parishes.

5.5 External validity: evidence from U.S. migrants

We lastly replicate our results using linked data from the U.S. 1880–1910 censuses showing that all the main results we find among female migrants in Sweden are also evident among migrant women in 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 since we observe women's maiden names. An important question is therefore whether our results can be generalized to other historical contexts. While we descriptively show in the introduction and

Section 2.1 that FLFP rates were substantially higher in the most populous places in both England and Wales as well as the United States, it is an open question whether the estimated impacts of migration on FLFP and demographic outcomes are similar in these other contexts.

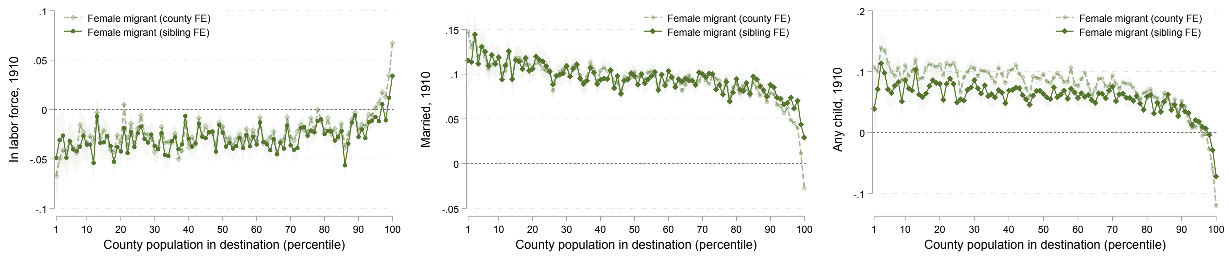
Data. To examine the returns to migration for women in the United States, we make use of links from the Census Tree project that recently has made substantial progress in linking women across the decennial U.S. censuses (Buckles et al., 2023). The Census Tree draws on the efforts by members of the largest genealogy community in the world (FamilySearch.org) that have created millions of handmade links between individuals recorded in the U.S. censuses. Crucially, because genealogists have private information about individuals (for example, maiden names) about half of the 317 million links created are for women. The Census Tree draws on these handmade links (the "Family Tree") as training data for a supervised ML algorithm that generates additional census-to-census links.[41] We document below that our results are not reliant on using any particular set of links.

To make the U.S. data more comparable with our Swedish linked data, we focus on native-born white women residing in rural areas in the U.S. in 1880 that can be linked between the 1880 and 1910 census. We further limit the sample to girls aged 0–16 in 1880, which thus have reached ages 30–46 when we observe their adult outcomes in the 1910 census. By combining these links with individual-level economic and demographic data from the U.S. censuses of 1880 and 1910 (IPUMS, 2020), we create a dataset of women and their sisters in the U.S. where we observe the same main economic and demographic outcomes as in our Swedish sample.[42]

Empirical strategy. Our empirical strategies are identical to the ones used above. First, we estimate Equations (1) and (3) and compare women in the U.S. that grew up in rural areas and that

[41] Additional links are also added to the Census Tree in a number of ways. First, additional links are generated based on potential links ("hints") to census records that are suggested by FamilySearch's own proprietary ML algorithm. Second, the Census Tree also includes links created through the Census Linking Project (Abramitzky et al., 2020) and the IPUMS Multigenerational Longitudinal Panel (Helgertz et al., 2023) though we do not rely on these links since the former does not link women and because the current public version of the Census Tree does not include links from the Multigenerational Longitudinal Panel between the 1880 and 1910 census. Lastly, the Census Tree adds "implied" links (for example, if two individuals in the 1880 and 1890 census are linked to the same individual in the 1910 census, then there exists an implied link between the two individuals in the 1880 and 1890 censuses).

[42] To faciliate comparisons with the Swedish census data, we crosswalk the occupations (`occ1950`) observed in the U.S. census to the HISCO system, which are then assigned to social status groups using the HISCLASS scheme as above.



(A) IN LABOR FORCE (B) MARRIED (C) ANY CHILD

FIGURE 10: EMPLOYMENT AND FAMILY FORMATION BY POPULATION IN U.S. MIGRANT DESTINATIONS

Notes: This figure displays OLS estimates of Equations (1) and (3) using the sample of U.S. women consisting of native-born white women born in rural location who have either migrated by 1910 or remain in their county of origin (the omitted category). The outcome is an indicator that takes the value one if a woman is in the labor force (Panel A), married (Panel B), or has a child (Panel C) in 1910, respectively. The figure plots point estimates and 95 percent confidence intervals capturing the effect of moving to different counties where all destination counties are grouped into percentiles based on their population size. Dashed shaded lines denote a specification using origin county fixed effects as well as individual-level controls (fixed effects for birthyear, birth order, and an indicator for being the eldest sister), while solid lines denote estimates that include individual-level controls and family fixed effects. Standard errors are clustered at the family level.

subsequently migrated by 1910 to stayers, which allows us to examine differences in the returns to migration across the size distribution of destinations. We include individual controls (fixed effects for birthyear, birth order, and an indicator for being the eldest sister) and either county of origin or family fixed effects, so that the estimates reflect differences relative to women in the county of origin or stayer sisters. Second, we estimate the equivalent to Equation (4) that uses a sibling design and examine the outcomes for sisters that moved to large and small cities to their migrant sisters moving to rural areas. See Section 4 for more details.

Main results. Figure 10 displays estimates of Equations (1) and (3) using the linked U.S. 1880–1910 census sample showing the effect of migration on FLFP by the percentile population rank in the destination county, relative to stayers in their county of origin. Migrants experience substantial increases in FLFP when moving to the most populous counties, while women moving to counties below the 90th percentile were less likely to transition into the formal labor market compared to their stayer sisters. Notably, the gains in FLFP due to migration are thus driven solely by destinations within the top of the population distribution in the U.S. as well in Sweden (see Figure 4).

TABLE 5: U.S. MIGRANTS ECONOMIC AND DEMOGRAPHIC OUTCOMES BY DESTINATION

Dependent variable:	Labor force participation			Occupational skill		Income score	Family formation	
	In labor force	In services	In industry	High skill	Low skill	ln(Income)	Married	Any child
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Migrant: large city (=1)	0.094***	0.079***	0.029***	0.052***	0.044***	0.120***	-0.103***	-0.167***
	(0.004)	(0.004)	(0.002)	(0.003)	(0.003)	(0.033)	(0.004)	(0.005)
Migrant: small city (=1)	0.051***	0.038***	0.026***	0.025***	0.025***	0.094***	-0.060***	-0.097***
	(0.002)	(0.002)	(0.001)	(0.002)	(0.001)	(0.021)	(0.002)	(0.003)
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sibling fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	556946	556946	556946	556946	556946	21444	556946	556946
Mean outcome	0.151	0.080	0.033	0.077	0.057	2.735	0.845	0.672

Notes: OLS regressions. All outcomes are measured in 1910. *Migrant: large city* is an indicator taking the value 1 if the individual lives in one of the 15 largest U.S. cities by population in 1910, and zero if not. *Migrant: small city* is an indicator taking the value 1 if the individual lives in a city other than the 15 largest in 1910, and zero if not. *Individual controls* include fixed effects for birth year and birth order, and an indicator for eldest sister. *Family 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 5 presents estimates similar to Equation (4) where we compare outcomes for migrants to "large" respectively "small" U.S. cities relative to migrants moving to rural destinations. To define large cities, we rank all U.S. cities in the 1910 census based on their population and select the 15 largest cities. We choose this cutoff because it approximately lines up with the population size of Stockholm in 1910, which facilitates comparisons between countries. Column 1 shows that U.S. migrants to cities that were at least as large as Stockholm see a 9.4 percentage point increase in FLFP compared to rural migrants, which can be compared to about 5.2 percentage point increase for migrants moving to a smaller city. The relative gains in FLFP are thus larger in U.S. cities at least as large as Stockholm.[43]

An increase in FLFP among U.S. female migrants can partly be ascribed to a greater transition into formal service jobs, similar to the case among Swedish migrants. Table 5 columns 2 and 3 shows that a migrant is significantly more likely to be observed in a service job, also relative to the increased probability to be employed in manufacturing.[44] The gains in service employment is further not concentrated mainly among low-skill jobs: columns 5 to 6 of Table 5 shows that

[43]To show that the estimated gains in FLFP among migrants are evident among cities of a similar size as Stockholm, Appendix Figure D.1 displays estimates of a modified version of Equation (4) where we include 16 different indicators for each of the 15 largest U.S. cities and all other (smaller) cities showing that U.S. migrants to cities that had a population similar to Stockholm (for example, Cincinnati or New Orleans) saw substantial gains in FLFP.

[44]Again, these results are evident in U.S. cities of a similar size as Stockholm as increases in service employment among female migrants is evident in virtually all of the 15 largest cities as shown in Appendix Figure D.2.

migrants to larger cities were relatively more likely to be observed in high-, middle-, or low-skill jobs, rather than unskilled work. Similarly, column 7 uses data on occupational income scores showing that the subset of female migrants that held an occupation in 1910 were employed in occupations that on average paid more compared to their sisters.[45]

A key finding above was that the gains in FLFP among Swedish migrants were deeply intertwined with reductions in marriage and fertility—reflecting the substantial marriage penalties that existed historically in both Sweden and the United States (Kleven et al., 2024). Panels B and C of Figure 10 show that marriage and childbearing outcomes follow the same pattern as in Sweden: female migrants were typically *more* likely to marry and have children compared to their stayer sisters, while those that moved to the largest places were much *less* likely to either marry or have a child.[46] The final two columns of Table 5 confirms this interpretation showing that U.S. migrants to the largest cities were substantially less likely to both marry and have children compared to migrants to smaller cities, or the countryside. Although magnitudes naturally differ, the general pattern of effects is remarkably similar across the United States and Sweden.

Additional estimates and robustness. We present three sets of robustness checks of our results for U.S. migrants. First, Appendix Table D.1 shows that the results are robust to excluding farm households (as defined in 1910) that could inflate our estimates of the impacts of cities on FLFP given the underenumeration of informal work at family farms (Goldin, 1990; Chiswick and Robinson, 2021). Second, FLFP and migrant flows differed by geography, nativity, and race (Goldin and Sokoloff, 1982; Goldin, 1990; Boustan and Collins, 2014): the main results are broadly similar in the South and non-South (Appendix Table D.2), among both natives and immigrants (Appendix Table D.3), as well as among Black and White women (Appendix Table D.4). Third, Appendix Table D.5 documents that the main results are robust to using alternative linkage methods.[47] Al-

[45] A similar result is obtained if one instead uses the LIDO score developed by Saavedra and Twinam (2020) that accounts for differences in income across genders.

[46] Although the likelihood of marriage is slightly higher even in the most populous counties when including family fixed effects in Panel B, the estimates in Table 5 shows that migrant women to large cities were much less likely to marry compared to their migrant sisters.

[47] Appendix Table D.5, Panel A, reports our baseline estimates using the full Census Tree, which we then restrict in a number of dimensions. First, in Panel B we limit the sample to links that are part of the Family Tree, which have been generated by hand by genealogists that have been shown to be highly reliable (Buckles et al., 2023). Second, Panel C limits the sample to the set of links that have been independently identified using three (or more) methods that are presumably of higher quality. Third, we lastly exclude links that have not been generated by genealogists: Panels D to F excludes links that have been (*i*) generated through the ML algorithm; (*ii*) identified by the proprietary ML algorithm

though the choice of linkage method matters for the estimated magnitudes, the main results are evident irregardless of the specific method chosen.

6 Conclusion

Our paper documents substantial variation in FLFP within early 20^{th}–century Europe and the United States showing that women in large cities were much more likely to be employed in the formal labor market. To establish the role of large cities in shaping women's employment, we use linked census data to show that American and Swedish women who moved to large cities were substantially more likely to enter formal employment and to remain unmarried than their sisters who migrated elsewhere. That the same patterns emerge in two countries with starkly different culture, institutions, and urbanization patterns indicates that the transformative role of large cities in expanding women's economic opportunities was a general phenomenon of early 20^{th}–century Europe and the United States.

An important driver of the high levels of FLFP in large cities was an early structural transformation toward the service sector. We show that large cities already by the early 20^{th} century had experienced a shift toward services, often up to half a century before similar shifts are observed in less populated regions. The early structural change crucially opened new opportunities for women in a variety of service occupations: the gains in FLFP that we observe among migrants are primarily explained by an increased transition to service jobs, underscoring the importance of the local economic structure in shaping women's opportunities.

Our study provides the first systematic quantitative evidence that large cities were instrumental for women's labor-market advances in the early 20^{th} century. The growing number of what Goldin (2006) termed "independent female workers" in large cities arguably laid the foundation for the later expansion of female employment and played a key role in the women's movement that flourished in large cities around the turn of the century. In this sense, the early rise of working women in American and European cities not only foreshadowed but may have directly contributed to the broader transformations in gender roles and labor markets that characterized the 20^{th} century.

used by FamilySearch as potential links; and (*iii*) the set of implied links that are generated when individuals are linked across multiple censuses.

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]," 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.

Ager, Philipp, James J Feigenbaum, Casper W Hansen, and Hui Ren Tan, "How the other half died: Immigration and mortality in US cities," *Review of Economic Studies*, 2024, *91* (1), 1–44.

_ , Marc Goñi, and Kjell G. Salvanes, "Gender-Biased Technological Change: Milking Machines and the Exodus of Women from Farming," IZA Discussion Paper 16347, IZA Institute of Labor Economics July 2023.

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

Alesina, Alberto, Paola Giuliano, and Nathan Nunn, "On the origins of gender roles: Women and the plough," *The quarterly journal of economics*, 2013, *128* (2), 469–530.

Althoff, Lukas, Harriet Brookes Gray, and Hugo Reichardt, "America's Rise in Human Capital Mobility," CESifo Working Paper No 11729 2025.

Andersson, Jonatan and Jakob Molinder, "Did cities increase skills during industrialization? Evidence from rural-urban migration," *Journal of Urban Economics*, 2025, *148*, 103772.

Berger, Thor and Kerstin Enflo, "Locomotives of local growth: The short-and long-term impact of railroads in Sweden," *Journal of Urban Economics*, 2017, *98*, 124–138.

— **and Peter Sandholt Jensen**, “Guilds and Growth: Evidence from the Free City,” CEPR Discussion Paper 18589, CEPR Press, Paris & London 2023.

— **, 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.

Boustan, Leah Platt and William J Collins, “The origin and persistence of Black-White differences in women’s labor force participation,” in “Human capital in history: The American record,” University of Chicago Press, 2014, pp. 205–240.

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

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, John N. Friedman, Nathaniel Hendren, Maggie R. Jones, and Sonya R. Porter, "The Opportunity Atlas: Mapping the Childhood Roots of Social Mobility," *American Economic Review*, forthcoming.

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.

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

Dribe, Martin, *Leaving Home in a Peasant Society. Economic Fluctuations, Household Dynamics and Youth Migration in Southern Sweden, 1829-1866.*, Vol. 13, Lund University, 2000.

_ , Björn Eriksson, and Francesco Scalone, "Migration, marriage and social mobility: Women in Sweden 1880–1900," *Explorations in Economic History*, 2019, *71*, 93–111.

Eckert, Fabian, Mads Hejlesen, and Conor Walsh, "The return to big-city experience: Evidence from refugees in Denmark," *Journal of Urban Economics*, 2022, *130*, 103454.

Ejermo, Olof, Kerstin Enflo, Björn Eriksson, and Erik Prawitz, "Home, Sweet Home: Returns to Returning in the Age of Mass Migration," *American Economic Journal: Applied Economics*, forthcoming.

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.

Eriksson, Katherine, Gregory Niemesh, Myera Rashid, and Jacqueline Craig, "Marriage and the Intergenerational Mobility of Women: Evidence from Marriage Certificates 1850–1920," 2025. Mimeo.

Feigenbaum, James and Daniel P Gross, "Answering the call of automation: How the labor market adjusted to mechanizing telephone operation," *The Quarterly Journal of Economics*, 2024, *139* (3), 1879–1939.

Fernandez, Raquel, "Women, work, and culture," *Journal of the European Economic Association*, 2007, *5* (2-3), 305–332.

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.

Greenwood, Jeremy, Ananth Seshadri, and Mehmet Yorukoglu, "Engines of liberation," *The Review of economic studies*, 2005, *72* (1), 109–133.

Haines, Michael R, "The urban mortality transition in the United States, 1800-1940," in "Annales de démographie historique," Vol. 101 Belin 2001, pp. 33–64.

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

Helgertz, Jonas, Steven Ruggles, John Robert Warren, Catherine A. Fitch, J. David Hacker, Matt A. Nelson, Joseph P. Price, Evan Roberts, and Matthew Sobek, "IPUMS Multigenerational Longitudinal Panel: Version 1.1 [dataset]," 2023.

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.

Jácome, Elisa, Ilyana Kuziemko, and Suresh Naidu, "Mobility for all: Representative intergenerational mobility estimates over the twentieth century," *Journal of Political Economy*, 2025, *133* (1), 306–354.

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

Karlsson, Lynn, "Women workers in figures: a picture of Swedish industry 1863–1912," Technical Report 40, Uppsala University, Department of Economic History, Uppsala 1996.

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.

Kleven, Henrik, Camille Landais, and Gabriel Leite-Mariante, "The child penalty atlas," *Review of Economic Studies*, 2024, p. rdae104.

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," *American Economic Review*, 2013, *103* (4), 1109–1137.

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.

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.

_ **, M Daniele Paserman, Laura Salisbury, and E Anna Weber**, "The emergence of the socioeconomic gradient in women's marriage outcomes," *Review of Economics and Statistics*, 2024, pp. 1–29.

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

Pérez, Santiago, "Intergenerational occupational mobility across three continents," *The Journal of Economic History*, 2019, *79* (2), 383–416.

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

Rashid, Myera, "Keys to Upward Mobility: Typewriter Adoption and Women's Economic Outcomes," 2025. Mimeo.

Rendall, Michelle, "Female market work, tax regimes, and the rise of the service sector," *Review of Economic Dynamics*, 2018, *28*, 269–289.

Roca, Jorge De La and Diego Puga, "Learning by working in big cities," *The Review of Economic Studies*, 2017, *84* (1), 106–142.

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.

Saavedra, Martin and Tate Twinam, "A machine learning approach to improving occupational income scores," *Explorations in Economic History*, 2020, *75*, 101304.

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.

U.S. Census Bureau, *Thirteenth Census of the United States Taken in the Year 1910: Volume 1. Population, General Report and Analysis*, Washington, D.C.: U.S. Government Printing Office, 1913.

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.

Vidart, Daniela, "Human capital, female employment, and electricity: Evidence from the early 20th-century United States," *Review of Economic Studies*, 2024, *91* (1), 560–594.

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.

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.

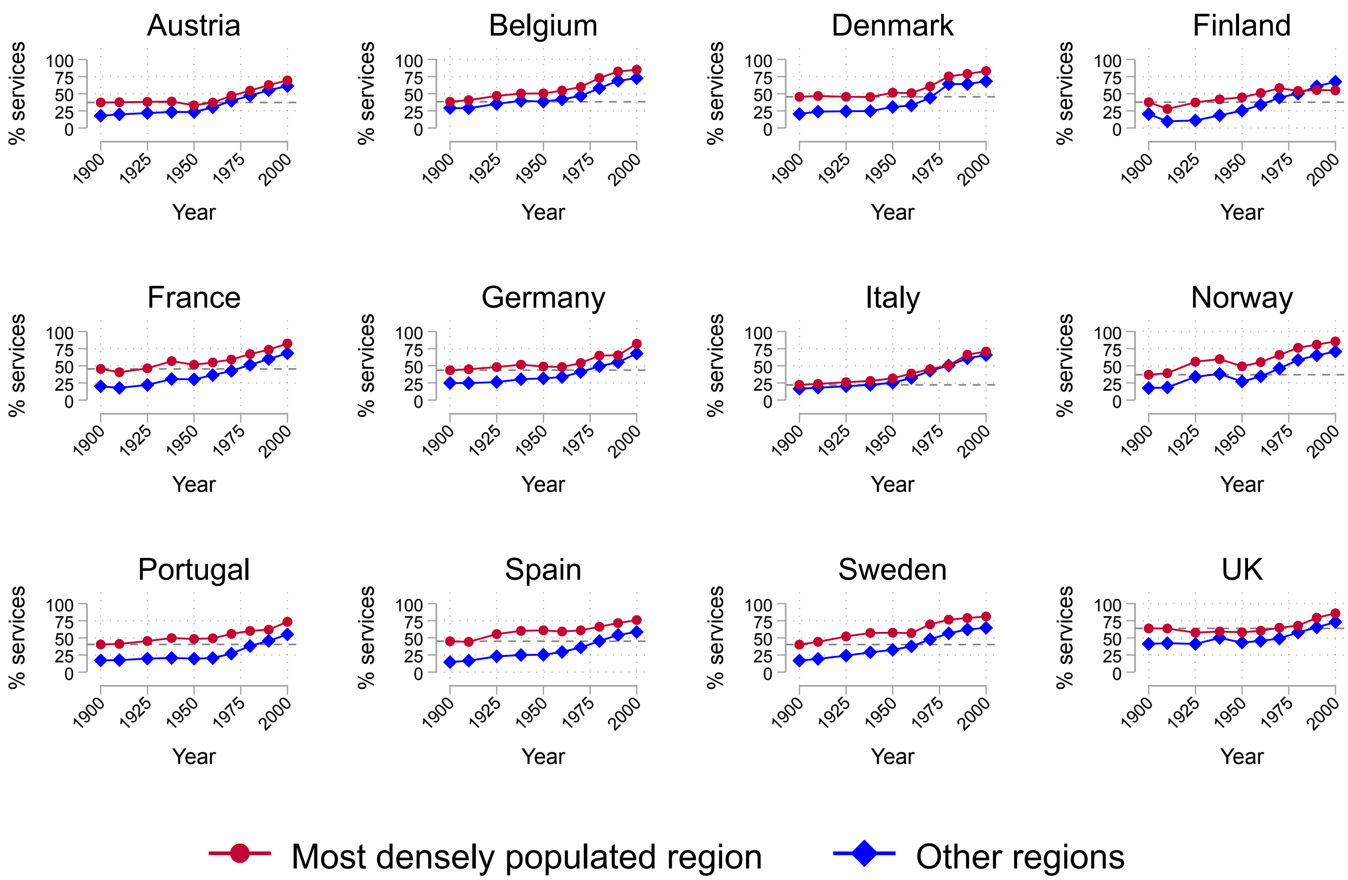
Wright, Carroll D., *Fourth Annual Report of the Commissioner of Labor, 1888: Working Women in Large Cities*, Washington, D.C.: U.S. Government Printing Office, 1889.

You, Xuesheng, “Women’s labour force participation in nineteenth-century England and Wales: evidence from the 1881 census enumerators’ books,” *The Economic History Review*, 2020, *73* (1), 106–133.

Zimran, Ariell, “Internal Migration in the United States: Rates, Selection, and Destination Choice, 1850–1940,” *The Journal of Economic History*, 2024, *84* (3), 727–766.

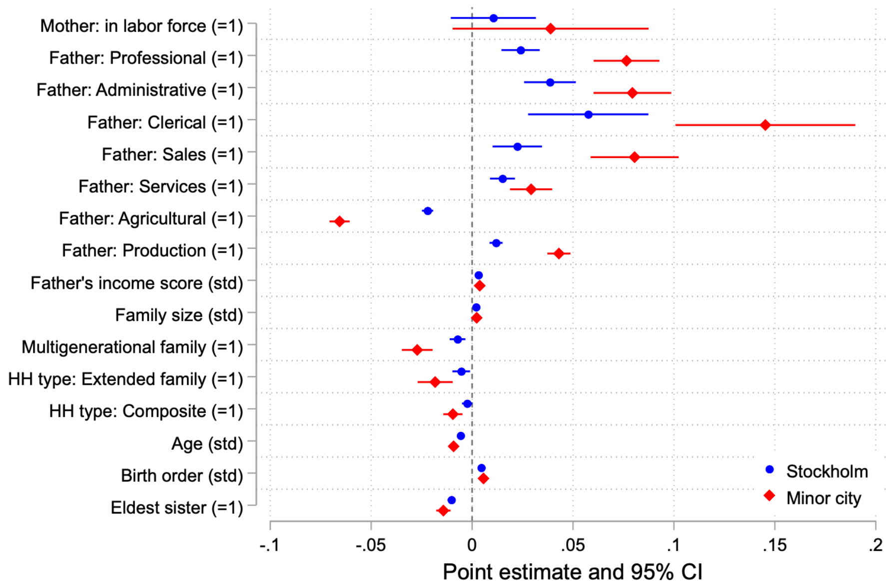
Online Appendix—not for publication

A Robustness and Additional Material

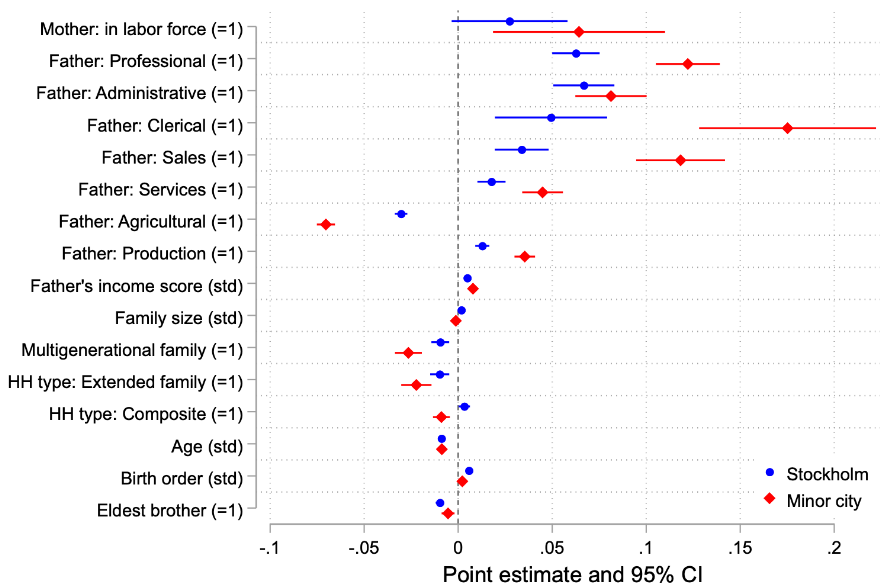


Notes: This figure shows that the most densely populated region in nearly all European countries had a higher share of employment in the service sector compared to other regions. Data on the employment share of services is drawn from Rosés and Wolf (2018). For each country, we report the employment share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



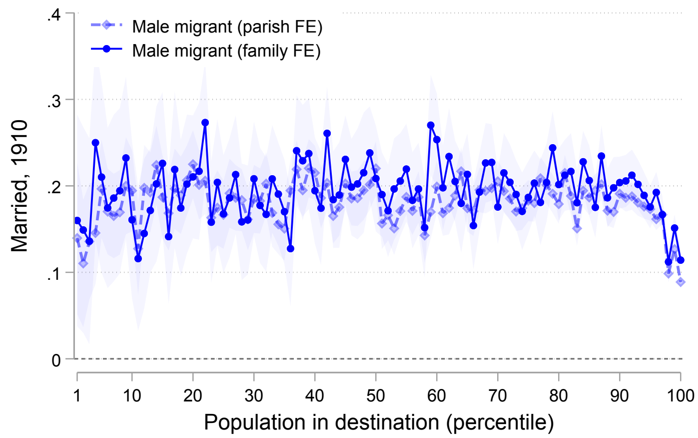
(A) WOMEN



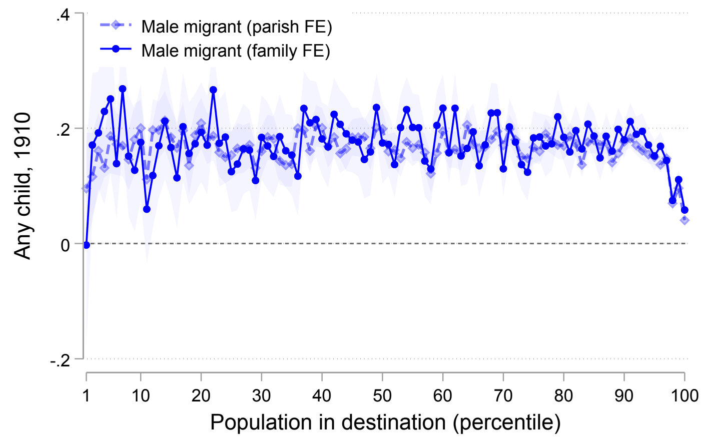
(B) MEN

FIGURE A.2: SORTING ACROSS MIGRANT DESTINATIONS

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 capturing the effect of moving to different parishes where the about 2,400 parish destinations are grouped into percentiles based on their population size. Dashed shaded lines denote a specification using origin parish fixed effects and family- and individual-level controls, while solid lines denote estimates that include family fixed effects with individual-level controls. See Section 4 for full list of control variables. Standard errors are clustered at the family level.

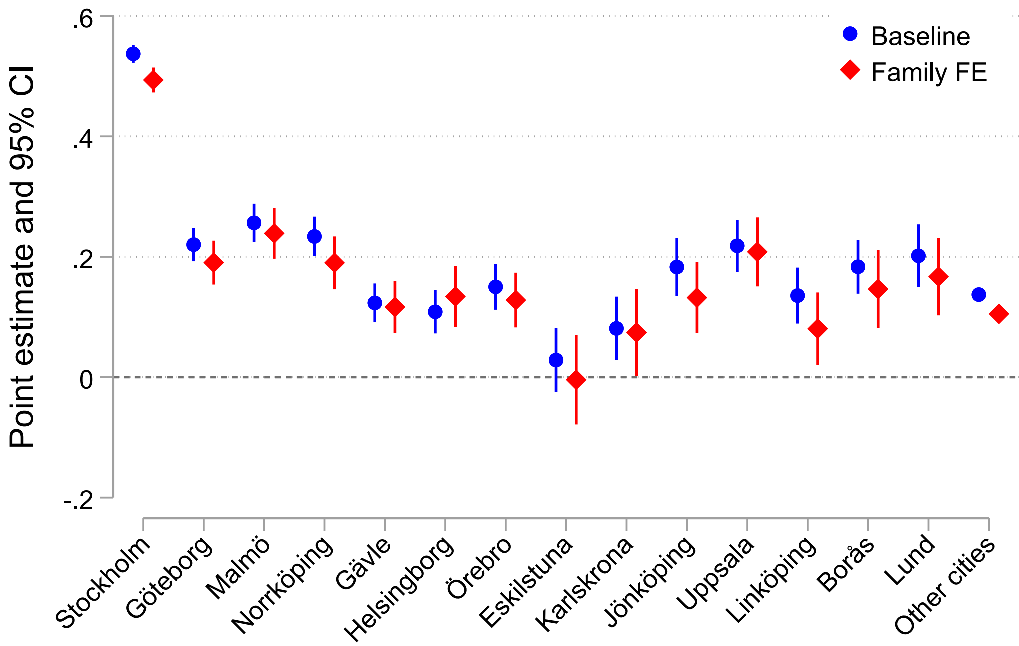


FIGURE A.4: FEMALE EMPLOYMENT BY MIGRANT DESTINATION

Notes: This figure displays OLS coefficients from regressions similar to those in Table 1 in the sample of migrant sisters from rural origins. The outcome is an indicator variable taking the value one if an individual is in the labor force in 1910. The figure plots point estimates and 95 percent confidence intervals that capture the returns to migrating to each Swedish city with a population above 20,000 in 1910 and the remaining smaller cities. We present baseline estimates including family/individual-level controls and parish of origin fixed effects, as well as estimates that include family fixed effects and individual-level controls. See Section 4 for full list of control variables. Standard errors are clustered at the family level.

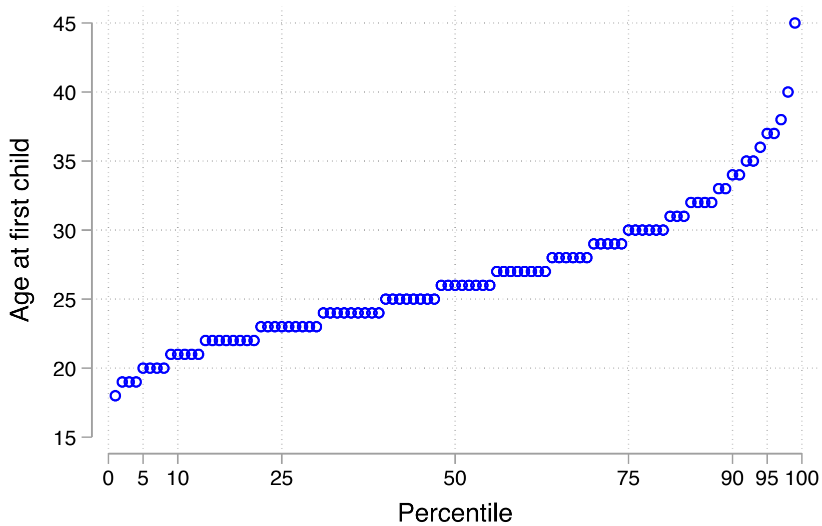


FIGURE A.5: AGE AT FIRST CHILD IN STOCKHOLM

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.

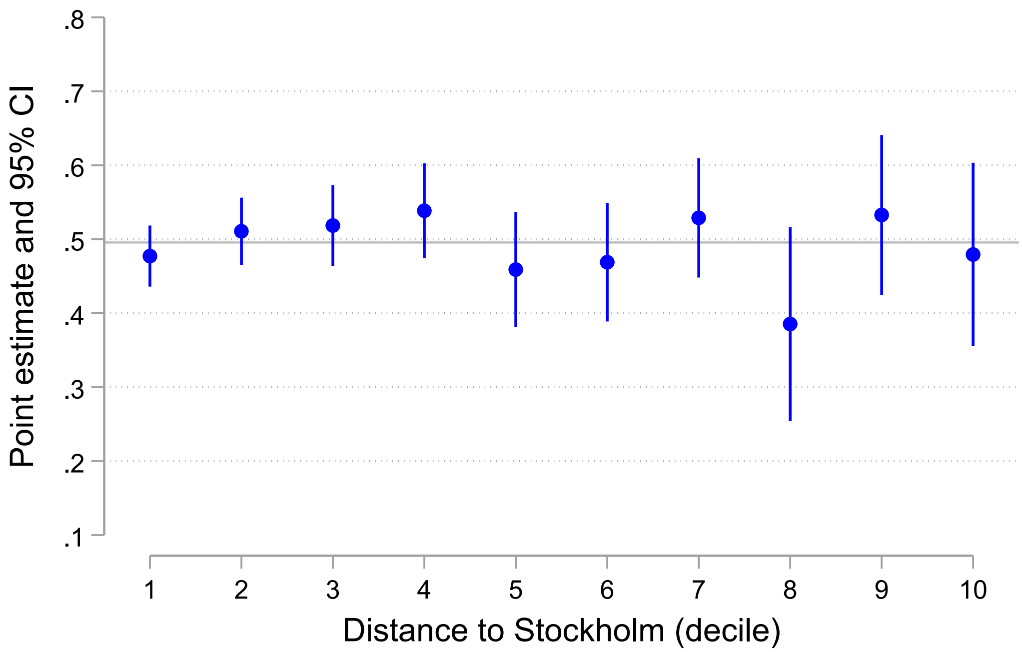
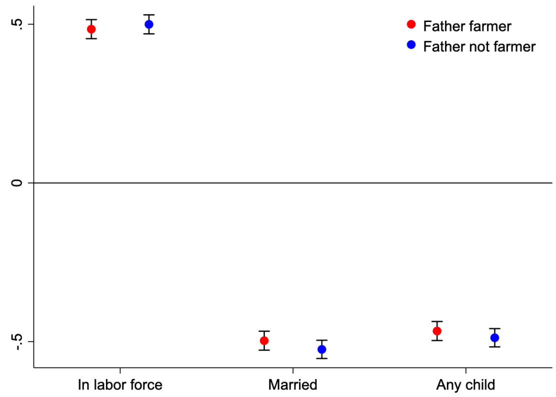
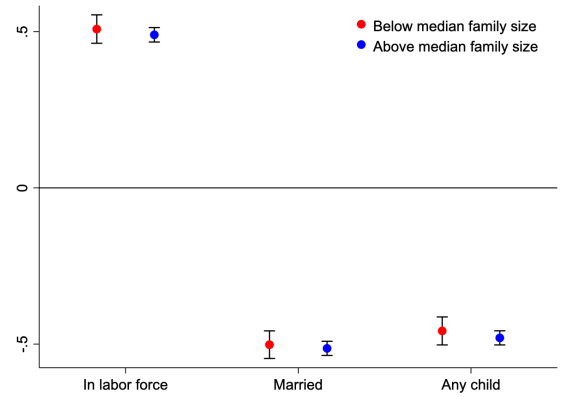


FIGURE A.6: FLFP AND MOVING TO STOCKHOLM FROM NEAR AND FAR

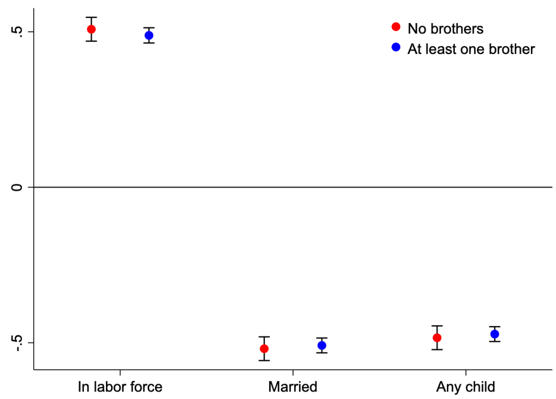
Notes: This figure displays point estimates and 95 percent confidence intervals from estimating Equation (4) where the outcome is an indicator taking the value one if a woman is in the labor force in 1910. We split the sample into deciles based on the distance from each parish to Stockholm and report estimates from ten separate regressions. For comparison, we also report the baseline estimate from Column 6 of Table 1 as a grey horizontal line. Standard errors are clustered at the family level.



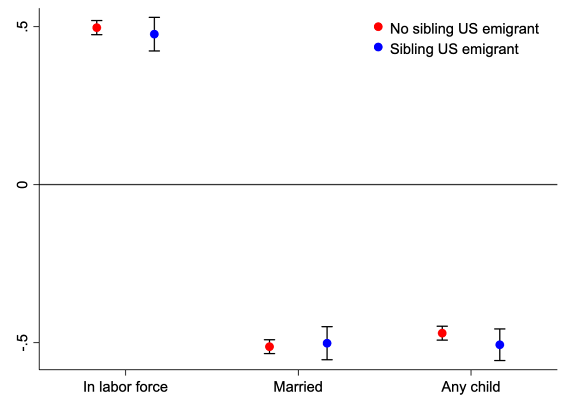
(A) FATHER FARMER IN 1880



(B) BELOW/ABOVE MEDIAN FAMILY SIZE



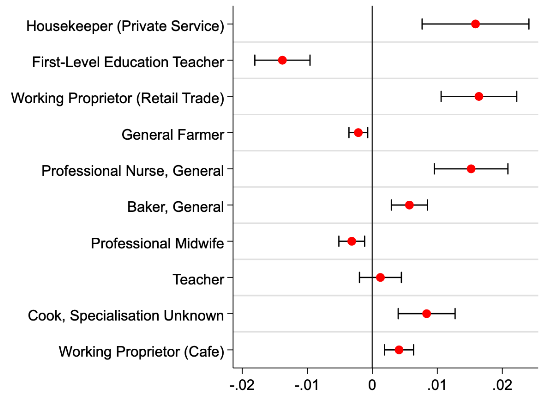
(C) ANY BROTHERS



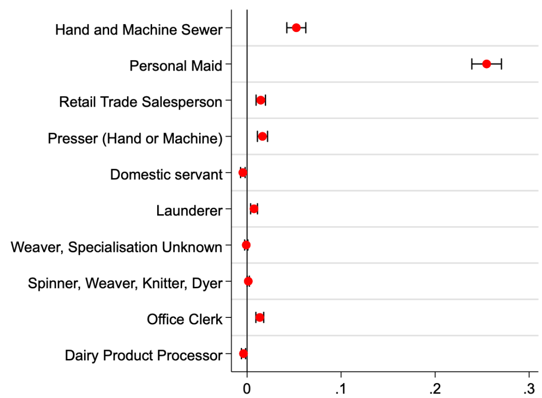
(D) SIBLINGS EMIGRATE 1880-1910

FIGURE A.7: SPLITTING SAMPLES

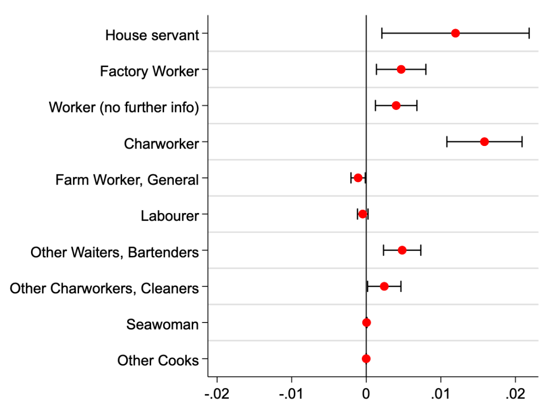
Notes: This figure displays regression coefficients when splitting the main sample along four dimensions. In Panel A, coefficients in red indicate households in which the father was a farmer, whereas blue coefficients show the opposite scenario. In Panel B, red coefficients are estimated on the subsample 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 on the subsample of families that have no sons in 1880, whereas blue coefficients are for families having at least one son. In panel D, red coefficients are estimated on the subsample of individuals that have a sibling observed as emigrated to North America, whereas blue coefficients are for individuals without any emigrant sibling. Specifications correspond to that of Column 6 of Table 1. Standard errors are clustered at the family level.



(A) HIGHER/MEDIUM SKILLED



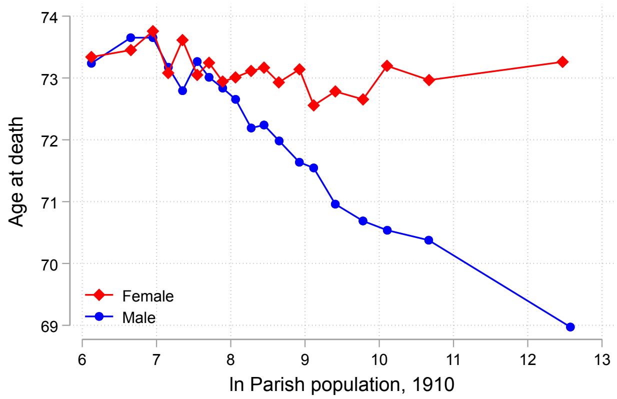
(B) LOWER SKILLED



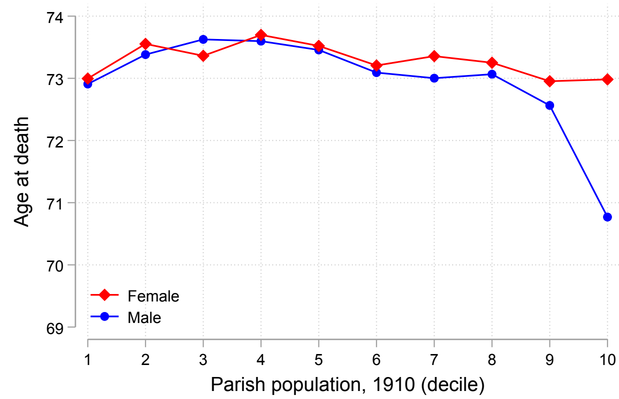
(C) UNSKILLED

FIGURE A.8: 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.



(A) LN POPULATION, 1910



(B) POPULATION DECILE, 1910

FIGURE A.9: AGE AT DEATH ACROSS MIGRANT DESTINATIONS

Notes: This figure displays binned scatter plots of the age at death for migrant men and women in our sample based on the population in their parish of residence in 1910 when migrants are aged 30 to 46.

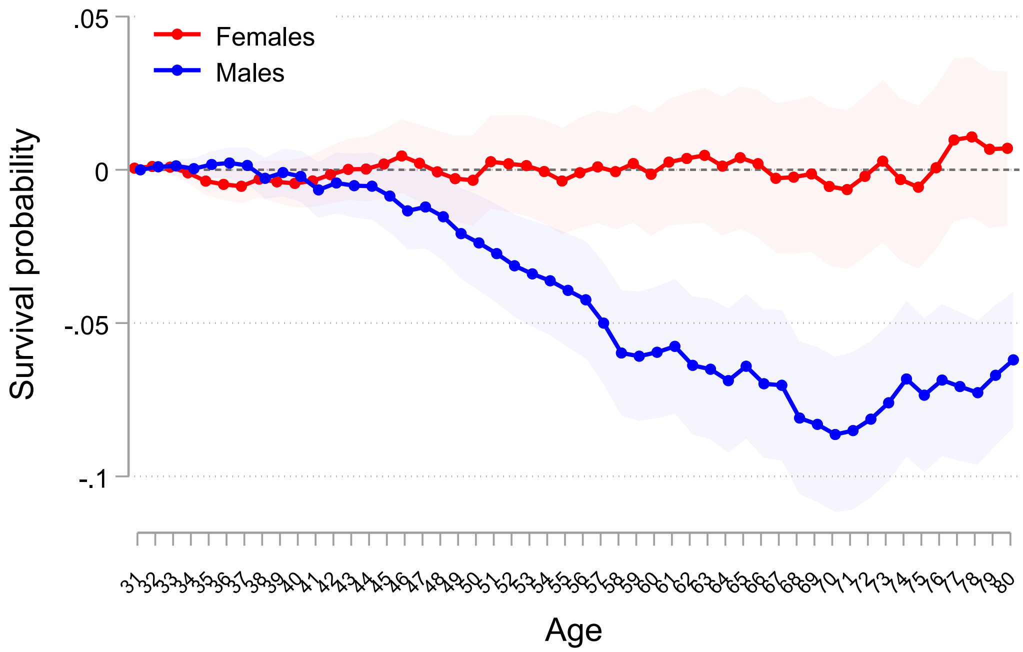


FIGURE A.10: SURVIVAL PROBABILITY IN STOCKHOLM COMPARED TO RURAL AREAS

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.

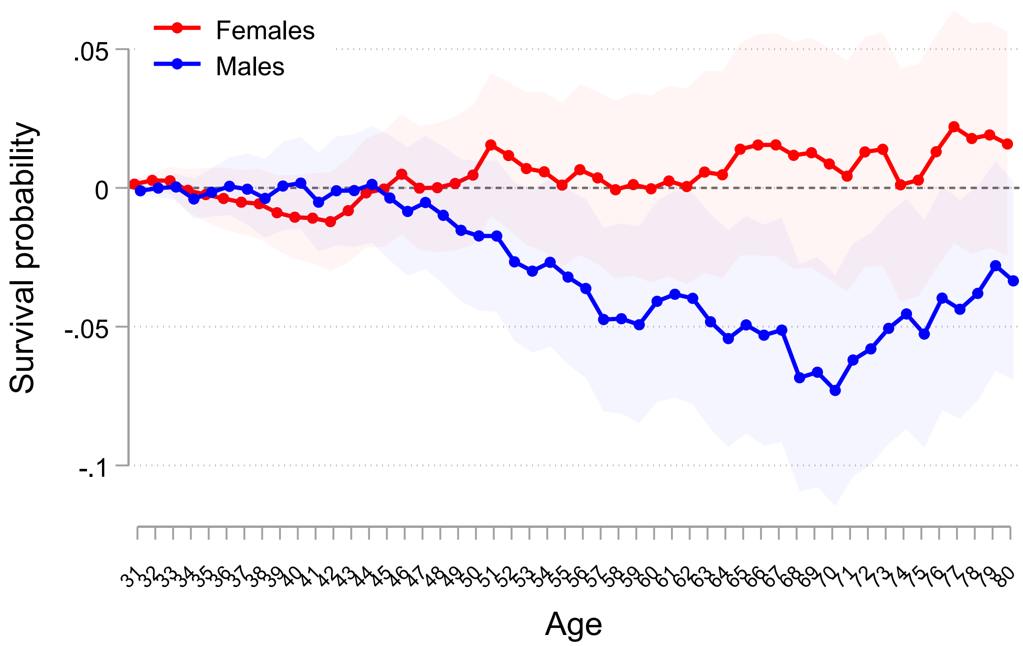
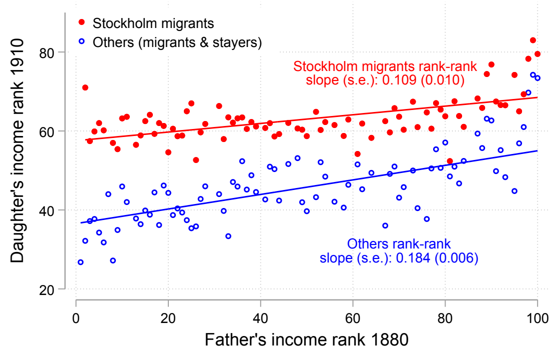
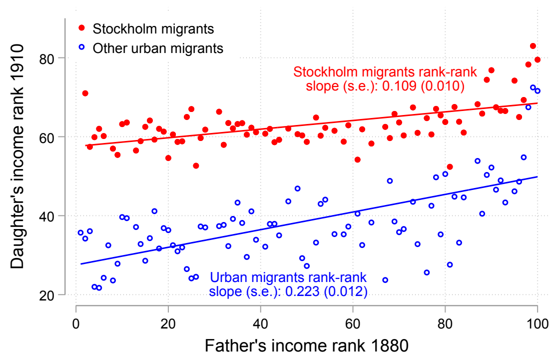


FIGURE A.11: SURVIVAL PROBABILITY IN STOCKHOLM COMPARED TO RURAL AREAS (AGE 30–37 IN 1910)

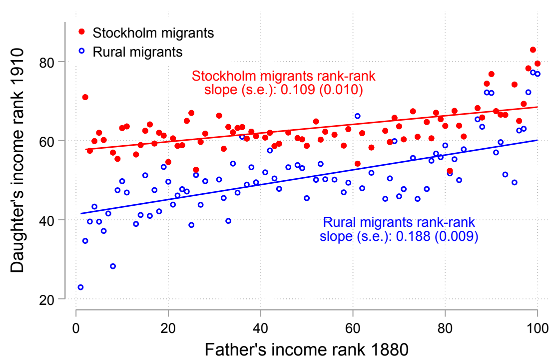
Notes: This figure displays regressions coefficients for likelihood of surviving past ages 31 to 80 in Stockholm relative to rural areas. The sample consists of men and women who were between 30 and 37 years old in 1910. Each coefficient is from a separate regression with siblings fixed effects and individual 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) STOCKHOLM MIGRANTS VS. OTHERS



(B) STOCKHOLM MIGRANTS VS. URBAN MIGRANTS



(C) STOCKHOLM MIGRANTS VS. RURAL MIGRANTS

FIGURE A.12: INTERGENERATIONAL INCOME MOBILITY

Notes: This figure presents a binned scatter plot of daughters' income ranks against their fathers' income ranks, where ranks are based on the mean income of the father's occupation in 1880 and the daughter's occupation in 1910. Panel A shows the mean income rank for female migrants to Stockholm (red circles) compared with all other female migrants and stayers (hollow blue circles) in our sample of women from rural origins. Panels B and C restrict the comparison groups to urban and rural migrants, respectively. We also report rank–rank slopes from regressions based on the underlying individual-level data, with standard errors clustered at the family level.

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 DIFFERENT SAMPLES

	Full	Siblings	Migrant	Migrant siblings sample			
				All destinations	Stockholm	Other urban	Rural parish
Panel A: Women	mean	mean	mean	mean	mean	mean	mean
Age 1880	7.213	7.197	7.085	7.064	6.504	6.878	7.180
Mother in labor force 1880	0.002	0.002	0.002	0.002	0.002	0.003	0.002
Father's income score 1880	7.485	7.495	7.484	7.499	7.524	7.522	7.489
Birthorder	2.748	3.094	2.745	3.116	3.243	3.156	3.091
Eldest sister	0.480	0.316	0.476	0.310	0.265	0.294	0.319
Disabled (=1)	0.001	0.001	0.001	0.001	0.000	0.000	0.001
In labor force 1910	0.177	0.186	0.206	0.226	0.683	0.322	0.151
Married 1910	0.693	0.688	0.736	0.712	0.239	0.627	0.785
Any child 1910	0.698	0.692	0.729	0.709	0.254	0.629	0.779
Observations	274711	143416	164530	70512	4565	16905	49042

Panel B: Men	mean	mean	mean	mean	mean	mean	mean
Age 1880	7.085	7.093	6.939	6.918	6.271	6.726	7.071
Mother in labor force 1880	0.002	0.002	0.003	0.003	0.003	0.003	0.002
Father's income score 1880	7.480	7.489	7.481	7.494	7.520	7.537	7.476
Birthorder	2.765	3.108	2.727	3.079	3.205	3.115	3.049
Eldest brother	0.458	0.303	0.460	0.302	0.271	0.297	0.309
Disabled (=1)	0.002	0.001	0.001	0.001	0.001	0.000	0.001
In labor force 1910	0.928	0.933	0.965	0.966	0.974	0.969	0.963
Married 1910	0.697	0.700	0.765	0.757	0.600	0.736	0.786
Any child 1910	0.644	0.648	0.701	0.694	0.494	0.662	0.733
Observations	262725	135620	142390	59851	5902	12956	40993

Notes: Summary statistics for the full sample from rural origins in column 1, a subsample of individuals in the full sample but with at least one same-sex sibling in column 2, all migrants from rural origins in column 3, and the migrant siblings (sisters/brothers) sample in column 4. The latter is 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: FLFP BY MIGRANT DESTINATION: ADJUSTED FLFP

Dependent variable:	In labor force (=1)		
		HHH not farmer	Conservative definition
	(1)	(2)	(3)
Stockholm (=1)	0.496***	0.449***	0.227***
	(0.011)	(0.012)	(0.012)
Other urban area (=1)	0.133***	0.091***	-0.093***
	(0.006)	(0.007)	(0.007)
Individual controls	Yes	Yes	Yes
Family FE	Yes	Yes	Yes
Observations	70512	43795	70512
Mean outcome	0.226	0.310	0.495

Notes: OLS regressions. *In labor force* is an indicator variable taking the value one if the individual is in the labor force in 1910. *Stockholm* is an indicator taking the value one if the individual lives in Stockholm city in 1910. *Other urban area* is an indicator taking the value one if the individual lives in an urban area other than Stockholm in 1910. The sample in column 1 consists of rural-born women who were aged 0–16 in 1880 and have left their parish of origin by 1910. Column 2 excludes women living with a household head (either father or husband) who is a farmer. *Conservative definition* defines women living with a household head (either father or husband) who is a farmer or a working proprietor as being in the labor force. *Individual controls* include fixed effects for birth year, birth order, being the eldest sister, and having a disability. *Family fixed effects* is a fixed effect for originating from the same family in 1880. Standard errors, in parentheses, are clustered at the family level. *** - $p < 0.01$, ** - $p < 0.05$, * - $p < 0.1$.

TABLE A.5: EMPLOYMENT, MARRIAGE, AND CHILDREN FOR MEN 1910

A. Dependent variable:	In labor force (=1)				
	(1)	(2)	(3)	(4)	(5)
Stockholm (=1)	0.009***	0.010***	0.010***	0.010***	0.010**
	(0.002)	(0.002)	(0.002)	(0.003)	(0.004)
Other urban area (=1)		0.005***	0.006***	0.006***	-0.000
		(0.001)	(0.001)	(0.002)	(0.003)
Observations	142390	142390	142388	59851	59851
Mean outcome	0.965	0.965	0.965	0.966	0.966

B. Dependent variable:	Married (=1)				
	(1)	(2)	(3)	(4)	(5)
Stockholm (=1)	-0.193***	-0.205***	-0.199***	-0.179***	-0.176***
	(0.005)	(0.005)	(0.005)	(0.007)	(0.010)
Other urban area (=1)		-0.055***	-0.045***	-0.039***	-0.041***
		(0.003)	(0.003)	(0.005)	(0.006)
Observations	142390	142390	142388	59851	59851
Mean outcome	0.765	0.765	0.765	0.757	0.757

C. Dependent variable:	Any child (=1)				
	(1)	(2)	(3)	(4)	(5)
Stockholm (=1)	-0.241***	-0.259***	-0.246***	-0.226***	-0.219***
	(0.005)	(0.005)	(0.005)	(0.007)	(0.010)
Other urban area (=1)		-0.075***	-0.063***	-0.058***	-0.060***
		(0.003)	(0.003)	(0.005)	(0.007)
Individual controls	No	No	Yes	Yes	Yes
Family 1880 controls	No	No	Yes	Yes	No
Family FE	No	No	No	No	Yes
Observations	142390	142390	142388	59851	59851
Mean outcome	0.701	0.701	0.701	0.694	0.694

Notes: OLS regressions. This table shows the effect of migration on marriage, children, and LFP in 1910. *Individual controls* include fixed effects for birth year, birth order, being the eldest sister, and having a disability. *Family 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.6: MIGRANTS' EMPLOYMENT USING ALTERNATIVE SAMPLES

Dependent variable:	In labor force (=1)				
Panel A: Extended sisters sample	(1)	(2)	(3)	(4)	(5)
Stockholm (=1)	0.548***	0.545***	0.548***	0.520***	0.515***
	(0.007)	(0.007)	(0.007)	(0.009)	(0.009)
Other urban area (=1)	0.175***	0.174***	0.162***	0.139***	0.137***
	(0.004)	(0.004)	(0.004)	(0.005)	(0.005)
Observations	143416	143416	143416	143416	143416
Mean outcome	0.186	0.186	0.186	0.186	0.186
Panel B: Migrant sisters	(1)	(2)	(3)	(4)	(5)
Stockholm (=1)	0.533***	0.529***	0.539***	0.501***	0.496***
	(0.007)	(0.007)	(0.008)	(0.011)	(0.011)
Other urban area (=1)	0.171***	0.169***	0.161***	0.136***	0.133***
	(0.004)	(0.004)	(0.004)	(0.006)	(0.006)
Observations	70512	70512	70512	70512	70512
Mean outcome	0.226	0.226	0.226	0.226	0.226
Panel C: Urban migrant sisters	(1)	(2)	(3)	(4)	(5)
Stockholm (=1)	0.356***	0.355***	0.389***	0.364***	0.361***
	(0.011)	(0.011)	(0.013)	(0.020)	(0.020)
Individual controls	No	Yes	Yes	No	Yes
Family 1880 controls	No	No	Yes	No	No
Family FE	No	No	No	Yes	Yes
Observations	12451	12451	12451	12451	12451
Mean outcome	0.422	0.422	0.422	0.422	0.422

Notes: OLS regressions. Panel A displays results for an extended sample of siblings where we include also stayers. Panel B replicates results for the main sample of migrant sisters. Panel C restricts the sample to only include migrants to urban areas. All outcomes are measured in 1910. All outcomes are measured in 1910. *Individual controls* include fixed effects for birth year, birth order, being the eldest sister, and having a disability. *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, 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.7: MIGRANTS' MARRIAGE AND FERTILITY USING ALTERNATIVE SAMPLES

Dependent variable:	Married (=1)					Any child (=1)				
Panel A: Extended sisters sample	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Stockholm (=1)	-0.481***	-0.474***	-0.479***	-0.455***	-0.444***	-0.470***	-0.462***	-0.461***	-0.440***	-0.426***
	(0.006)	(0.006)	(0.006)	(0.009)	(0.009)	(0.006)	(0.006)	(0.006)	(0.009)	(0.009)
Other urban area (=1)	-0.081***	-0.078***	-0.065***	-0.051***	-0.047***	-0.088***	-0.085***	-0.068***	-0.055***	-0.050***
	(0.004)	(0.004)	(0.004)	(0.005)	(0.005)	(0.004)	(0.004)	(0.004)	(0.005)	(0.005)
Observations	143416	143416	143416	143416	143416	143416	143416	143416	143416	143416
Mean outcome	0.688	0.688	0.688	0.688	0.688	0.692	0.692	0.692	0.692	0.692
Panel B: Migrant sisters	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Stockholm (=1)	-0.546***	-0.541***	-0.545***	-0.519***	-0.511***	-0.525***	-0.518***	-0.518***	-0.486***	-0.476***
	(0.007)	(0.007)	(0.007)	(0.010)	(0.010)	(0.007)	(0.007)	(0.007)	(0.010)	(0.010)
Other urban area (=1)	-0.158***	-0.156***	-0.146***	-0.126***	-0.123***	-0.150***	-0.147***	-0.133***	-0.111***	-0.107***
	(0.004)	(0.004)	(0.004)	(0.006)	(0.006)	(0.004)	(0.004)	(0.004)	(0.006)	(0.006)
Observations	70512	70512	70512	70512	70512	70512	70512	70512	70512	70512
Mean outcome	0.712	0.712	0.712	0.712	0.712	0.709	0.709	0.709	0.709	0.709
Panel C: Urban migrant sisters	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Stockholm (=1)	-0.375***	-0.374***	-0.401***	-0.396***	-0.391***	-0.376***	-0.373***	-0.390***	-0.371***	-0.366***
	(0.010)	(0.010)	(0.012)	(0.019)	(0.019)	(0.010)	(0.010)	(0.012)	(0.019)	(0.019)
Individual controls	No	Yes	Yes	No	Yes	No	Yes	Yes	No	Yes
Household controls	No	No	Yes	No	No	No	No	Yes	No	No
Family FE	No	No	No	Yes	Yes	No	No	No	Yes	Yes
Observations	12451	12451	12451	12451	12451	12451	12451	12451	12451	12451
Mean outcome	0.512	0.512	0.512	0.512	0.512	0.522	0.522	0.522	0.522	0.522

Notes: OLS regressions. Panel A displays results for an extended sample of siblings where we include also stayers. Panel B replicates results for the main sample of migrant sisters. 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, birth order, being the eldest sister, and having a disability. *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, 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). *Family 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.8: UNMARRIED MIGRANTS

Dependent variable:	Married (=1)		Any child (=1)		In labor force (=1)	
	Unmarried 1900 sample					
	pre-migration (1)	post-migration (2)	pre-migration (3)	post-migration (4)	pre-migration (5)	post-migration (6)
Stockholm (=1)	-0.468***	-0.324***	-0.472***	-0.310***	0.559***	0.363***
	(0.058)	(0.039)	(0.057)	(0.039)	(0.058)	(0.044)
Other urban area (=1)	-0.131***	-0.107***	-0.118***	-0.108***	0.147***	0.155***
	(0.036)	(0.029)	(0.036)	(0.029)	(0.035)	(0.031)
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes
Family FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2862	6669	2862	6669	2862	6669
Mean outcome	0.604	0.409	0.564	0.423	0.299	0.426

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. *Stockholm* is an indicator taking the value one if the individual lives in Stockholm city in 1910. *Other urban area* is an indicator taking the 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, being the eldest sister, and having a disability. *Family fixed effects* is a fixed effect for originating from the same family in 1880. Standard errors, in parentheses, are clustered at the family level. *** - $p < 0.01$, ** - $p < 0.05$, * - $p < 0.1$.

TABLE A.9: 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)
Stockholm (=1)	-0.024	-0.210***	-0.340***	-0.208***	0.519***	0.250***
	(0.200)	(0.045)	(0.072)	(0.070)	(0.075)	(0.088)
Other urban area (=1)	-0.003	-0.092**	-0.107**	-0.064	0.119***	0.224***
	(0.151)	(0.039)	(0.046)	(0.057)	(0.046)	(0.066)
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes
Family FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	451	3643	2857	1449	2857	1449
Mean outcome	0.503	0.344	0.495	0.277	0.332	0.582

Notes: OLS regressions. This table shows the effect of migration on marriage, child bearing, and FLFP 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, birth order, being the eldest sister, and having a disability. *Family 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.10: 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)	0.339***		-0.370***		-0.373***		0.084***	
	(0.016)		(0.016)		(0.016)		(0.012)	
Migrant in 1900: other urban area (=1)	0.091***		-0.089***		-0.088***		0.015***	
	(0.008)		(0.008)		(0.008)		(0.005)	
Ever migrated: Stockholm (=1)		0.359***		-0.091***		-0.371***		0.060***
		(0.009)		(0.015)		(0.009)		(0.006)
Ever migrated: other urban area (=1)		0.100***		-0.096***		-0.083***		0.019***
		(0.005)		(0.005)		(0.005)		(0.003)
Migrant in 1900: Stockholm (=1)			0.000					
			(.)					
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Family FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	39262	78519	39262	78519	39262	78519	39262	78519
Mean outcome	0.217	0.221	0.712	0.710	0.715	0.709	0.073	0.076

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 the value one if the individual lives in Stockholm city in 1900. *Migrant in 1900: other urban area* is an indicator taking the value one if the individual lives in an urban area other than Stockholm in 1900. *Ever migrated: Stockholm* is an indicator taking the 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 the 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, being the eldest sister, and having a disability. *Family fixed effects* is a fixed effect for originating from the same family in 1880. Standard errors, in parentheses, are clustered at the family level. *** - $p < 0.01$, ** - $p < 0.05$, * - $p < 0.1$.

TABLE A.11: 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***	-0.176***	-0.212***	0.062***
	(0.022)	(0.023)	(0.023)	(0.016)
Migrant in 1900: other urban area (=1)	0.077***	-0.078***	-0.078***	0.012**
	(0.008)	(0.009)	(0.008)	(0.005)
Individual controls	Yes	Yes	Yes	Yes
Family FE	Yes	Yes	Yes	Yes
Observations	36067	36067	36067	36067
Mean outcome	0.186	0.744	0.745	0.069

Notes: OLS regressions. All outcomes are measured in 1910.
textitIndividual controls include fixed effects for birth year, birth order, being the eldest sister, and having a disability. *Family fixed effects* is a fixed effect for originating from the same family in 1880. Standard errors, in parentheses, are clustered at the 1880 household level. *** - $p < 0.01$, ** - $p < 0.05$, * - $p < 0.1$.

TABLE A.12: 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)	0.543***	0.545***	0.531***	-0.472***	-0.453***	-0.454***	-0.483***	-0.456***	-0.456***
	(0.054)	(0.054)	(0.055)	(0.055)	(0.055)	(0.055)	(0.052)	(0.053)	(0.053)
Migrant after 1900: other urban area (=1)	0.140***	0.139***	0.126***	-0.133***	-0.113***	-0.114***	-0.115***	-0.093***	-0.092***
	(0.028)	(0.028)	(0.028)	(0.029)	(0.028)	(0.028)	(0.031)	(0.030)	(0.030)
In labor force 1900		0.124***	0.069**			-0.016			0.012
		(0.031)	(0.032)			(0.031)			(0.031)
Married 1900			-0.189***		0.254***	0.274***			0.062
			(0.041)		(0.023)	(0.041)			(0.051)
Any child 1900			0.011			-0.029		0.298***	0.247***
			(0.041)			(0.041)		(0.025)	(0.051)
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Family FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	4003	4003	4003	4003	4003	4003	4003	4003	4003
Mean outcome	0.240	0.240	0.240	0.685	0.685	0.685	0.651	0.651	0.651

Notes: OLS regressions. All outcomes are measured in 1910. *Individual controls* include fixed effects for birth year, birth order, being the eldest sister, and having a disability. *Family fixed effects* is a fixed effect for originating from the same family in 1880. Standard errors, in parentheses, 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 family 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.13 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.13: TWIN FIXED EFFECTS

Dependent variable:	In labor force	Married	Any children
	(1)	(2)	(3)
Stockholm (=1)	0.751***	-0.732***	-0.551***
	(0.120)	(0.120)	(0.130)
Other urban area (=1)	0.096	-0.055	-0.046
	(0.082)	(0.088)	(0.088)
Individual controls	Yes	Yes	Yes
Twin FE	Yes	Yes	Yes
Observations	443	443	443
Mean outcome	0.253	0.709	0.698

Notes: OLS regressions in sample of twins. All outcomes are measured in 1910. *Individual controls* include fixed effects for birth year, birth order, being the eldest sister, and having a disability. *Twin fixed effects* is a fixed effect for same-sex twin siblings. Standard errors, in parentheses, are clustered at the 1880 household level. *** - $p < 0.01$, ** - $p < 0.05$, * - $p < 0.1$.

TABLE A.14: INFLUENCE OF UNOBSERVED WITHIN-FAMILY 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)	0.537	0.568	-0.548	-0.458	-0.498	-0.453
Uncontrolled beta	0.463	0.516	-0.483	-0.443	-0.455	-0.463
Controlled beta	0.496	0.516	-0.511	-0.453	-0.476	-0.456

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 the value one for individuals who live in Stockholm in 1910, and zero otherwise. All models include family fixed effects. Estimates from the barebones model with only family 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. *** - $p < 0.01$, ** - $p < 0.05$, * - $p < 0.1$.

Cohabitation and children born out of wedlock

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 3 may be misleading. However, we show in Appendix Table A.15 that the increase in cohabitation for female migrants is small in magnitude.[48]

Moreover, 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.15 shows that female migrants are less likely to have any children even when restricting attention to those born out of wedlock.

TABLE A.15: COHABITATION AND CHILDREN BORN OUT OF WEDLOCK

Dependent variable:	Cohabit	Cohabit or married	Any illegitime child	Any child
	(1)	(2)	(3)	(4)
Stockholm (=1)	0.004	-0.508***	-0.065***	-0.476***
	(0.003)	(0.010)	(0.004)	(0.010)
Other urban area (=1)	-0.000	-0.123***	0.005	-0.107***
	(0.001)	(0.006)	(0.003)	(0.006)
Individual controls	Yes	Yes	Yes	Yes
Family FE	Yes	Yes	Yes	Yes
Observations	70512	70512	70512	70512
Mean outcome	0.007	0.719	0.070	0.709

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, birth order, being the eldest sister, and having a disability. Standard errors are given in parentheses and are clustered at the 1880 family level. *** - $p < 0.01$, ** - $p < 0.05$, * - $p < 0.1$.

[48] 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.

TABLE A.16: FAMILY FORMATION BY MIGRANT DESTINATION

Dependent variable:	Married, 1910 (=1)	Any child, 1910 (=1)
	(1)	(2)
Stockholm × Female (=1)	-0.249***	-0.178***
	(0.027)	(0.028)
Other urban area × Female (=1)	-0.078***	-0.060***
	(0.012)	(0.012)
Individual controls	Yes	Yes
Family-by-destination FE	Yes	Yes
Observations	53078	53078
Mean outcome	0.656	0.640

Notes: OLS regressions using the sample of migrant sisters and brothers. *Stockholm × Female* is an indicator taking the value 1 if a woman lives in Stockholm city in 1910, and zero if not. *Other urban area × Female* is an indicator taking the value 1 if a woman lives in an urban area other than Stockholm in 1910, and zero if not. *Individual controls* include fixed effects for sex, birth year, birth order, being the eldest sister, and having a disability. *Family-by-destination fixed effects* is a fixed effect for originating from the same family in 1880 and the same parish of residence in 1910. Standard errors, in parentheses, are clustered at the family level. *** - $p < 0.01$, ** - $p < 0.05$, * - $p < 0.1$.

TABLE A.17: OLS REGRESSIONS: WEIGHTING REGRESSIONS WITH THE INVERSE PROBABILITY OF BEING LINKED ACROSS CENSUSES

Dependent variable:	In labor force	Married	Any children
	(1)	(2)	(3)
Stockholm (=1)	0.502***	-0.519***	-0.486***
	(0.011)	(0.010)	(0.010)
Other urban area (=1)	0.134***	-0.123***	-0.108***
	(0.006)	(0.006)	(0.006)
Individual controls	Yes	Yes	Yes
Family FE	Yes	Yes	Yes
Observations	69397	69397	69397
Mean outcome	0.224	0.715	0.713

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, birth order, being the eldest sister, and having a disability. Standard errors, in parentheses, are clustered at the 1880 family level. *** - $p < 0.01$, ** - $p < 0.05$, * - $p < 0.1$.

TABLE A.18: ALTERNATIVE STANDARD ERRORS

Dependent variable:	In labor force (=1)			Married (=1)			Any child (=1)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Stockholm (=1)	0.543	0.539	0.496	-0.552	-0.545	-0.511	-0.527	-0.518	-0.476
Robust	(0.007)***	(0.007)***	(0.011)***	(0.007)***	(0.007)***	(0.010)***	(0.007)***	(0.007)***	(0.010)***
Cluster(Sibling)	{0.008}***	{0.008}***	{0.011}***	{0.007}***	{0.007}***	{0.010}***	{0.007}***	{0.007}***	{0.010}***
Cluster(Origin Municip.)	[0.008]***	[0.008]***	[0.010]***	[0.007]***	[0.007]***	[0.010]***	[0.007]***	[0.007]***	[0.011]***
Other urban area (=1)	0.164	0.161	0.133	-0.151	-0.146	-0.123	-0.139	-0.133	-0.107
Robust	(0.004)***	(0.004)***	(0.006)***	(0.004)***	(0.004)***	(0.006)***	(0.004)***	(0.004)***	(0.006)***
Cluster(Sibling)	{0.004}***	{0.004}***	{0.006}***	{0.004}***	{0.004}***	{0.006}***	{0.004}***	{0.004}***	{0.006}***
Cluster(Origin Municip.)	[0.005]***	[0.005]***	[0.006]***	[0.005]***	[0.005]***	[0.006]***	[0.005]***	[0.005]***	[0.006]***

Notes: OLS regressions. All outcomes are measured in 1910. *Stockholm* is an indicator taking the value one if the individual lives in Stockholm city in 1910, and zero if not. *Other urban area* is an indicator taking the value one if the individual lives in an urban area other than Stockholm in 1910, and zero if not. Standard errors for each coefficient are shown in parentheses using different calculations as indicated in italics. *** - $p < 0.01$, ** - $p < 0.05$, * - $p < 0.1$.

TABLE A.19: MIGRANTS' MAJOR OCCUPATIONAL (HISCO) GROUPS

Dependant variable:	Professional	Administrative	Clerical	Sales	Service	Agricultural	Production
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Stockholm (=1)	0.011**	0.018***	0.024***	0.032***	0.338***	-0.005***	0.075***
	(0.005)	(0.005)	(0.003)	(0.004)	(0.010)	(0.001)	(0.006)
Other urban area (=1)	-0.004*	0.000	0.006***	0.016***	0.068***	-0.005***	0.043***
	(0.002)	(0.002)	(0.001)	(0.002)	(0.004)	(0.001)	(0.003)
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Family FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	70512	70512	70512	70512	70512	70512	70512
Mean outcome	0.027	0.025	0.007	0.015	0.086	0.006	0.053

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, birth order, being the eldest sister, and having a disability. Standard errors, in parentheses, are clustered at the 1880 family level. *** - $p < 0.01$, ** - $p < 0.05$, * - $p < 0.1$.

TABLE A.20: SECTOR, SKILL, AND INCOME BY MIGRANT DESTINATION

Dependent variable:	Sector of employment		Occupational skill		Income score		Age at death
	Services	Industry	High	Low	ln(income)	Pct. rank	Female
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Stockholm × Female (=1)	0.141***	-0.089***	0.144***	0.314***	0.074**	2.507	5.367***
	(0.029)	(0.025)	(0.028)	(0.030)	(0.033)	(1.628)	(0.962)
Other urban area × Female (=1)	-0.049***	-0.191***	-0.005	0.117***	0.005	0.427	1.333***
	(0.010)	(0.011)	(0.011)	(0.013)	(0.021)	(1.439)	(0.469)
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Family-by-destination FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	53078	53078	53078	53078	21021	21021	36748
Mean outcome	0.145	0.283	0.243	0.329	7.354	51.401	72.531

Notes: OLS regressions using the sample of migrant sisters and brothers. All outcomes are measured in 1910. *Services* and *Industry* are indicators taking the value 1 for the HISCO major groups 0–5 and 7–9, respectively. Similarly, *High* corresponds to high- and medium-skilled occupations (HISCLASS groups 1–4 and 6–8), while *Low* corresponds to low-skill and unskilled occupations (HISCLASS groups 5 and 9–12). *Stockholm* × *Female* is an indicator taking the value 1 if a woman lives in Stockholm city in 1910, and zero if not. *Other urban area* × *Female* is an indicator taking the value 1 if a woman lives in an urban area other than Stockholm in 1910, and zero if not. *Individual controls* include fixed effects for sex, birth year, birth order, being the eldest sister, and having a disability. *Family-by-destination fixed effects* is a fixed effect for originating from the same family in 1880 and the same parish of residence in 1910. Standard errors, in parentheses, are clustered at the family level. *** - $p < 0.01$, ** - $p < 0.05$, * - $p < 0.1$.

TABLE A.21: MIGRANTS' HOUSEHOLD INCOME WITH DIFFERENT WEIGHTS FOR SPOUSE INCOME

Dependant variable:	Household income score					Intra-household
	100	80	60	40	20	ratio
	(1)	(2)	(3)	(4)	(5)	(6)
Stockholm (=1)	0.201***	0.322***	0.479***	0.699***	1.077***	0.514***
	(0.012)	(0.012)	(0.013)	(0.015)	(0.021)	(0.010)
Other urban area (=1)	-0.073***	-0.045***	-0.008	0.044***	0.133***	0.127***
	(0.006)	(0.006)	(0.007)	(0.008)	(0.011)	(0.006)
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes
Family FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	58915	58915	58915	58915	58915	70512
Mean outcome	6.843	6.672	6.453	6.144	5.617	0.294

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, birth order, being the eldest sister, and having a disability. *Family fixed effects* is a fixed effect for originating from the same family in 1880. Standard errors, in parentheses, are clustered at the 1880 family level. *** - $p < 0.01$, ** - $p < 0.05$, * - $p < 0.1$.

TABLE A.22: INTERGENERATIONAL MOBILITY OF FEMALE MIGRANTS TO STOCKHOLM

	Absolute mobility						Relative mobility
Dependent variable:	$\mathbb{1}\{\text{Rank}^{\text{daughter}} > \text{Rank}^{\text{father}}\}$		$\text{Rank}^{\text{daughter}} \mid \text{Rank}^{\text{father}} < 50$		P(Q5 | Q1)		$\text{Rank}^{\text{daughter}}$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Stockholm (=1)	0.140***	0.089***	16.129***	17.382***	0.082	0.063	20.279***
	(0.018)	(0.024)	(1.995)	(2.344)	(0.061)	(0.057)	(2.295)
Income $\text{rank}^{\text{father}}$							0.138***
							(0.031)
Migrant: Stockholm (=1) × Income $\text{rank}^{\text{father}}$							-0.123***
							(0.035)
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Family 1880 controls	Yes	No	Yes	No	Yes	No	Yes
Family FE	No	Yes	No	Yes	No	Yes	No
Observations	6160	6160	2393	2393	502	502	6160
Mean outcome	0.434	0.434	49.992	49.992	0.129	0.129	54.142

Notes: OLS regressions. The sample consists of the migrant sisters sample and excludes women with missing income scores. The outcome in columns 1 and 2 is an indicator taking the value one if a woman attains a higher income rank than her father. The outcome in columns 3 and 4 is a woman's income rank, where the sample is restricted to sisters with fathers in the bottom half of the income distribution. The outcome in columns 5 and 6 is the probability that a woman born to a father in the bottom quintile of the income distribution reaches the top income quintile. *Stockholm* is an indicator taking the value one if the individual lives in Stockholm city in 1910. *Individual controls* include fixed effects for birth year, birth order, being the eldest sister, and having a disability. *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 fixed effects for mother's employment, father's 1-digit HISCO occupation, and whether the household is multigenerational, extended family (relatives only), or composite family (with non-relatives). *Family fixed effects* is a fixed effect for originating from the same family in 1880. Standard errors, in parentheses, are clustered at the family level. *** - $p < 0.01$, ** - $p < 0.05$, * - $p < 0.1$.

TABLE A.23: MIGRANTS' HEALTH OUTCOMES

Dependent variable:	Age at death			
	Women		Men	
	(1)	(2)	(3)	(4)
Stockholm (=1)	0.177	-0.179	-2.315***	-1.931***
	(0.359)	(0.400)	(0.351)	(0.372)
Other urban area (=1)	0.225	0.158	-1.140***	-0.703***
	(0.204)	(0.208)	(0.241)	(0.258)
Individual controls	Yes	Yes	Yes	Yes
Family FE	Yes	Yes	Yes	Yes
Occupation FE	No	Yes	No	Yes
Observations	52881	52781	44600	44462
Mean outcome	73.303	73.305	71.900	71.902

Notes: OLS regressions. *Individual controls* include fixed effects for birth year, birth order, being the eldest sister, and having a disability. *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, unrelated members, as well as an indicator for farming households. *Family fixed effects* is a fixed effect for same sex siblings. *** - $p < 0.01$, ** - $p < 0.05$, * - $p < 0.1$.

B Migration and FLFP: the role of the services sector

A key finding above is that female migrants to large cities experienced increases in FLFP primarily due to the transition into services jobs. For example, the estimate in Table 1 indicates that among migrants to Stockholm the probability to be in formal employment increased by 49.7 percentage points, while the estimate reported in Table 2 indicates an increased probability to work in services of about 40.3 percentage points suggesting that the vast majority of the increase in FLFP is driven by services employment. Additional suggestive evidence on a key role of the services sector in accounting for increases in FLFP is evident in Figure 1 and 2 showing that differences in FLFP across counties and parishes in Britain, Sweden, and the United States is highly correlated with local differences in the supply of services jobs.

To provide additional evidence on the key role of services, we exploit 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 B.1, 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 assess the extent to which the relationship between urban migration and employment is mediated by the service sector, we include both migration dummies and measures of the sectoral composition. Compared to our main estimate in Table 1 with family fixed effects, column 2 shows that the indicator for migration to Stockholm is reduced by almost 20 percentage points when controlling for sectoral employment shares at migrants' destinations (corresponding to about a 40 percent decrease). Similarly, the estimate for migration to other urban areas is diminished by about 10 percentage points.

Although the service sector appears key in accounting for our results, another potential explanation to our results may be the fact that large cities such as Stockholm typically were majority female. If a female-biased sex ratio reduces marriage rates among women, it may also explain why more women are working. To test for the influence of unbalanced sex ratios, we calculate sex ratios for each destination area in 1910 and include it as a control in column 3. As seen from the table, the local sex ratio has a positive association with paid labor.[49] Compared to our main

[49] Moreover, Appendix Table B.2 and B.3 show that the sex ratio has an expected negative correlation with family formation.

TABLE B.1: ACCOUNTING FOR THE INCREASE IN FLFP

Dependent variable:	In labor force (=1)					
	(1)	(2)	(3)	(4)	(5)	(6)
Stockholm (=1)	0.424***	0.234***	0.402***	0.229***	0.420***	0.240***
	(0.029)	(0.034)	(0.029)	(0.034)	(0.030)	(0.035)
Other urban area (=1)	0.136***	0.040***	0.111***	0.032***	0.133***	0.047***
	(0.006)	(0.011)	(0.008)	(0.011)	(0.007)	(0.011)
Destination service share		0.261***		0.223***		0.273***
		(0.037)		(0.039)		(0.040)
Destination industry share		-0.008		-0.007		-0.006
		(0.019)		(0.019)		(0.020)
Destination workers (log)		0.024***		0.024***		0.022***
		(0.004)		(0.004)		(0.004)
Adult sex ratio			0.145***	0.094***		
			(0.027)	(0.028)		
Share women signed petition					0.014	-0.069**
					(0.032)	(0.033)
Female turnout					0.059**	0.012
					(0.027)	(0.028)
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes
Family FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	53830	53830	53830	53830	51895	51895
Mean outcome	0.190	0.190	0.190	0.190	0.191	0.191

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

estimate in Table B.1, the sex ratio inclusion lowers the estimated impact of Stockholm migration by approximately 2 percentage points. Thus, while high female sex ratios can account for some of the effects of migration to Stockholm and can be seen as a contributing factor, its importance appears less pronounced as compared to the sectoral composition of destination locations.

Yet another potential explanation to our results may be the presence of differential norms in cities related to female labor. To test for the role of gender norms, we propose two broad measures: the support for female suffrage and female turnout in the 1922 prohibition referendum. While the former is an indicator for the belief that women can assume the responsibility of voting, the latter captures actual political participation among women after universal suffrage was introduced.[50] Interestingly, controlling for the share of women who signed the petition or female turnout at destination make essentially no impact on our estimates. Column 5 of Table B.1 shows that higher female turnout is positively associated with being in the labor force, but the relationship becomes insignificant once we control for sectoral composition in column 6. In contrast, we find no independent significant correlation for women signing the petition in column 5, while column 6 indicates a negative correlation conditional on sectoral composition.

Together, the evidence in this section suggests that the availability of service sector jobs appears to be a major determinant determinant of the increases in FLFP observed among female migrants. Nevertheless, a substantial difference in employment and family formation remains, most notably in Stockholm. This suggests the presence of a city-specific component. While our broad measures of progressive attitudes related to gender norms appear not to predict a greater fraction of women in paid work, we cannot fully exclude their importance.

[50]To measure the support for female suffrage, we digitize signatures to a nationwide petition for female suffrage that was distributed in 1913–14, which was signed by a total of 351,000 women, approximately 18 percent of all adult women at the time. To measure female turnout in the 1922 referendum, we use previously digitized data from the Swedish National Data Service (SND).

TABLE B.2: ACCOUNTING FOR THE DECREASE IN FAMILY FORMATION, MARRIAGE

Dependent variable:	Married (=1)					
	(1)	(2)	(3)	(4)	(5)	(6)
Stockholm (=1)	-0.462***	-0.290***	-0.432***	-0.283***	-0.453***	-0.282***
	(0.028)	(0.034)	(0.029)	(0.034)	(0.029)	(0.035)
Other urban area (=1)	-0.131***	-0.037***	-0.098***	-0.025**	-0.125***	-0.037***
	(0.007)	(0.011)	(0.008)	(0.012)	(0.008)	(0.012)
Destination service share		-0.324***		-0.270***		-0.346***
		(0.039)		(0.041)		(0.042)
Destination industry share		0.018		0.017		0.021
		(0.020)		(0.020)		(0.022)
Destination workers (log)		-0.017***		-0.017***		-0.016***
		(0.004)		(0.004)		(0.004)
Adult sex ratio			-0.195***	-0.134***		
			(0.029)	(0.030)		
Share women signed petition					-0.038	0.048
					(0.034)	(0.036)
Female turnout					-0.052*	-0.013
					(0.030)	(0.031)
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes
Family FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	53830	53830	53830	53830	51895	51895
Mean outcome	0.746	0.746	0.746	0.746	0.745	0.745

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

TABLE B.3: ACCOUNTING FOR THE DECREASE IN FAMILY FORMATION, CHILDREN

Dependent variable:	Any child (=1)					
	(1)	(2)	(3)	(4)	(5)	(6)
Stockholm (=1)	-0.405***	-0.257***	-0.379***	-0.252***	-0.395***	-0.245***
	(0.028)	(0.035)	(0.029)	(0.035)	(0.029)	(0.036)
Other urban area (=1)	-0.110***	-0.021*	-0.081***	-0.011	-0.097***	-0.014
	(0.007)	(0.012)	(0.008)	(0.012)	(0.008)	(0.012)
Destination service share		-0.345***		-0.303***		-0.352***
		(0.040)		(0.042)		(0.043)
Destination industry share		-0.002		-0.003		0.006
		(0.020)		(0.020)		(0.022)
Destination workers (log)		-0.010***		-0.010***		-0.011***
		(0.004)		(0.004)		(0.004)
Adult sex ratio			-0.171***	-0.105***		
			(0.029)	(0.030)		
Share women signed petition					-0.081**	0.003
					(0.034)	(0.036)
Female turnout					-0.072**	-0.032
					(0.030)	(0.031)
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes
Family FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	53830	53830	53830	53830	51895	51895
Mean outcome	0.741	0.741	0.741	0.741	0.741	0.741

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

C Instrumental variables: additional results

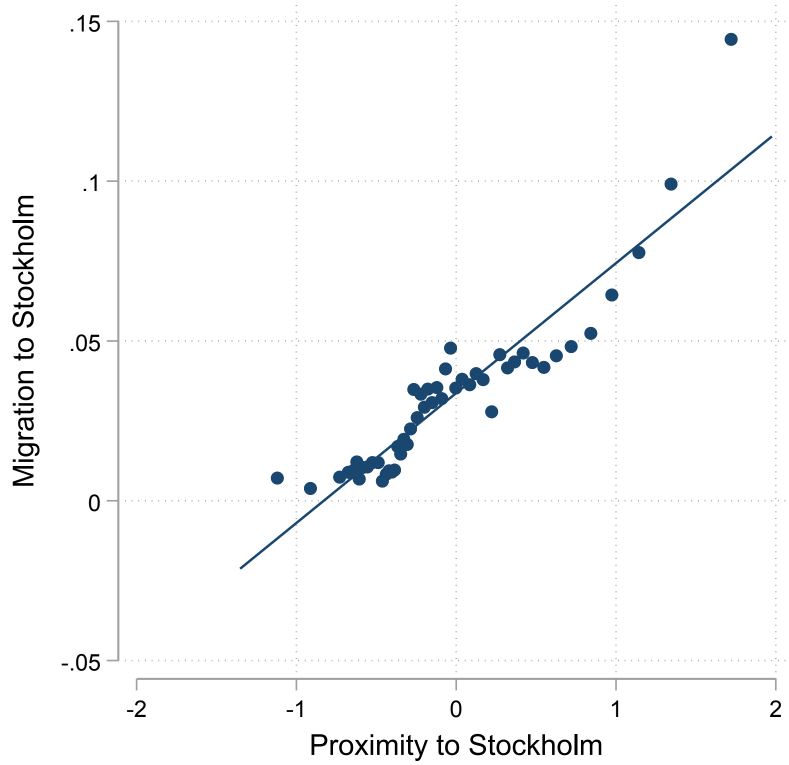


FIGURE C.1: CHILDHOOD PROXIMITY TO STOCKHOLM PREDICTS MIGRATION TO STOCKHOLM

Notes: This figure shows the relationship childhood proximity to Stockholm and the likelihood of migration to Stockholm by 1910. The sample consists of rural women who are 0 to 16 years old in 1880. Graphs are produced using *binsreg* with 50 bins (Cattaneo et al., 2024). Fixed effects for birth order, birth year, and number of siblings are absorbed. Parishes located above the 99^{th} percentile of proximity to Stockholm are excluded for expositional reasons.

TABLE C.1: PROXIMITY TO STOCKHOLM DOES NOT PREDICT GENDER OF OLDER SIBLINGS

Dependent variable:	Has older sister		Has older brother	
	(1)	(2)	(3)	(4)
Prox to Sthlm × Birth order	0.0005		-0.0001	
	(0.0004)		(0.0004)	
Prox to Sthlm × Birth order = 2		-0.0009		0.0008
		(0.0029)		(0.0029)
Prox to Sthlm × Birth order = 3		0.0013		0.0017
		(0.0028)		(0.0028)
Prox to Sthlm × Birth order = 4		0.0017		0.0002
		(0.0028)		(0.0026)
Prox to Sthlm × Birth order = 5		0.0022		0.0006
		(0.0028)		(0.0025)
Prox to Sthlm × Birth order = 6		0.0036		0.0006
		(0.0034)		(0.0024)
Prox to Sthlm × Birth order = 7		-0.0038		-0.0013
		(0.0036)		(0.0024)
Prox to Sthlm × Birth order = 8		0.0044		-0.0058*
		(0.0040)		(0.0034)
Observations	274711	270474	274711	270474
Mean outcome	0.520	0.514	0.544	0.538

Notes: OLS regressions. This table shows that, conditional on birth order, proximity to Stockholm does not predict the gender of older siblings. All models include fixed effects for birth order and childhood parish. In columns 2 and 4, individuals with sibships above the 99^{th} percentile (9 children or more) are excluded. Standard errors clustered at the childhood parish level.*** - $p < 0.01$, ** - $p < 0.05$, * - $p < 0.1$.

TABLE C.2: IV BALANCE TEST

Outcome	Instrument: Has Older Sister × Proximity to Stockholm (1)	Placebo instrument: Has Older Brother × Proximity to Stockholm (2)	Observations
Disabled (=1)	0.0001	0.0002	274711
	(0.0002)	(0.0002)	
Father's income score rank	0.1525	0.2473	254943
	(0.1490)	(0.1920)	
Father: Professional (=1)	-0.0004	0.0003	254949
	(0.0009)	(0.0011)	
Father: Administrative (=1)	0.0000	0.0010	254949
	(0.0009)	(0.0011)	
Father: Clerical (=1)	-0.0000	-0.0000	254949
	(0.0004)	(0.0007)	
Father: Sales (=1)	-0.0005	-0.0003	254949
	(0.0005)	(0.0008)	
Father: Services (=1)	-0.0014	0.0005	254949
	(0.0012)	(0.0014)	
Father: Agricultural (=1)	0.0037	0.0048	254949
	(0.0026)	(0.0032)	
Father: Production (=1)	-0.0014	-0.0062 **	254949
	(0.0023)	(0.0029)	
Mother: in labor force, 1880 (=1)	0.0000	-0.0003	274711
	(0.0003)	(0.0003)	
Family size	0.0020	-0.0006	274711
	(0.0023)	(0.0032)	
Multigenerational family, 1880 (=1)	0.0001	-0.0002	274711
	(0.0012)	(0.0016)	
HH type: Extended family, 1880 (=1)	0.0009	0.0010	272512
	(0.0010)	(0.0013)	
HH type: Composite, 1880 (=1)	0.0043 *	0.0016	272512
	(0.0023)	(0.0031)	

Notes: OLS regressions. Each cell represents a separate regression. This table shows the correlation between the instrument—the interaction between having an older sister and childhood parish proximity to Stockholm—on a set of pre-determined covariates from Equation (1). It also shows correlations for the placebo instrument constructed using older brothers instead of older sisters. Each model controls for an indicator for having an older sister (column 1) or brother (column 2) as well as fixed effects for birth order, birth year, number of siblings, and childhood parish. Standard errors clustered at the childhood parish level.*** - $p < 0.01$, ** - $p < 0.05$, * - $p < 0.1$.

TABLE C.3: THE INSTRUMENT DOES NOT PREDICT OVERALL PROPENSITY TO MIGRATE

Dependent variable:	Any migration			
	Outside childhood parish		Outside childhood county	
	(1)	(2)	(3)	(4)
Has older sister	0.012***	0.012***	0.008***	0.008***
	(0.002)	(0.002)	(0.002)	(0.002)
Prox to Sthlm × Has older sister		-0.002		-0.000
		(0.002)		(0.002)
Observations	274711	274711	274711	274711
Mean outcome	0.599	0.599	0.220	0.220

Notes: OLS regressions. This table shows the relationship between the instrument—the interaction between having an older sister and childhood parish proximity to Stockholm—and the overall propensity to migrate out of the childhood parish or county by 1910. All models control for fixed effects for birth order, birth year, number of siblings, and childhood parish. Standard errors clustered at the childhood parish level.*** - $p < 0.01$, ** - $p < 0.05$, * - $p < 0.1$.

TABLE C.4: IV: USING EVER MIGRATED AS THE ENDOGENOUS VARIABLE

Dependent variable:	Sthlm migrant	In labor force	Married	Any child
	OLS	IV	IV	IV
	(1)	(2)	(3)	(4)
Prox to Sthlm $\times$ Has older sister	0.016***			
	(0.002)			
Ever migrated: Stockholm (=1)		0.337**	-0.383**	-0.232
		(0.136)	(0.157)	(0.164)
Observations	274711	274711	274711	274711
First-stage F-statistic		68.7	68.7	68.7
Mean outcome	0.051	0.177	0.693	0.698

Notes: OLS regressions and IV regressions. Outcomes are measured in 1910. This table shows results when using a dummy for ever having migrated to Stockholm as the endogenous variable. The instrument is the interaction between having an older sister and childhood parish proximity to Stockholm. All models include an indicator for having an older sister as well as fixed effects for birth order, birth year, number of siblings, having a disability, childhood parish, as well as family controls listed in Table 4. Standard errors clustered at the childhood parish level. *** - $p < 0.01$, ** - $p < 0.05$, * - $p < 0.1$.

TABLE C.5: IV RESULTS ON SECTOR, SKILL, AND INCOME

Dependent variable:	Sector of employment		Occupational skill		Occ. income score		Age at death
	Services	Industry	High	Low	ln(Income)	Pct. rank	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Stockholm (=1)	0.261*	-0.021	0.061	0.357**	0.076	9.668	-0.492
	(0.156)	(0.100)	(0.120)	(0.146)	(0.390)	(22.525)	(7.123)
Observations	274711	274711	274711	274711	46785	46785	214170
First-stage F-statistic	43.6	43.6	43.6	43.6	6.0	6.0	38.5
Mean outcome	0.114	0.042	0.060	0.112	6.930	48.750	72.901

Notes: IV regressions. Outcomes are measured in 1910. The instrument is the interaction between having an older sister and childhood parish proximity to Stockholm. All models include an indicator for having an older sister as well as fixed effects for birth order, birth year, number of siblings, having a disability, childhood parish, as well as family controls listed in Table 4. Standard errors clustered at the childhood parish level. *** - $p < 0.01$, ** - $p < 0.05$, * - $p < 0.1$.

TABLE C.6: EXCLUSION RESTRICTION ROBUSTNESS TESTS

Panel A. Dependent variable:	Stockholm migrant			
	(1)	(2)	(3)	(4)
Prox to Sthlm × Has older sister	0.012***	0.012***	0.013***	0.010***
	(0.002)	(0.002)	(0.002)	(0.002)
Mean outcome	0.036	0.036	0.036	0.036
Panel B. Dependent variable:	**In labor force**			
Stockholm (=1)	0.432**	0.500***	0.357*	0.529*
	(0.174)	(0.191)	(0.199)	(0.283)
Mean outcome	0.177	0.177	0.177	0.177
Panel C. Dependent variable:	**Married**			
Stockholm (=1)	-0.491**	-0.535**	-0.618***	-0.463
	(0.202)	(0.225)	(0.223)	(0.330)
Mean outcome	0.693	0.693	0.693	0.693
Panel D. Dependent variable:	**Any child**			
Stockholm (=1)	-0.298	-0.308	-0.285	-0.360
	(0.210)	(0.228)	(0.228)	(0.326)
Has older brother × Prox to Sthlm	No	Yes	No	No
Has older sister × Prox to city	No	No	Yes	No
Birth order × Prox to Sthlm	No	No	No	Yes
Observations	274711	274711	274711	274711
First-stage F-statistic	43.6	38.0	43.7	16.3
Mean outcome	0.698	0.698	0.698	0.698

Notes: OLS (Panel A) and IV (Panels B-D) regressions. Outcomes are measured in 1910. The instrument is the interaction between having an older sister and childhood parish proximity to Stockholm. All models include an indicator for having an older sister as well as fixed effects for birth order, birth year, number of siblings, having a disability, childhood parish, as well as family controls listed in Table 4. Standard errors clustered at the childhood parish level. *** - $p < 0.01$, ** - $p < 0.05$, * - $p < 0.1$.

TABLE C.7: PLACEBO TESTS USING THE INSTRUMENT ON THE MALE SAMPLE

Dependent variable:	Sthlm migrant	In labor force	Married	Any child
	(1)	(2)	(3)	(4)
Prox to Sthlm × Has older sister	0.002	-0.001	0.004	0.004
	(0.002)	(0.001)	(0.003)	(0.003)
Prox to Sthlm × Has older brother	0.008***	0.000	0.001	0.001
	(0.002)	(0.001)	(0.002)	(0.003)
Observations	262725	262725	262725	262725
Mean outcome	0.049	0.928	0.697	0.644

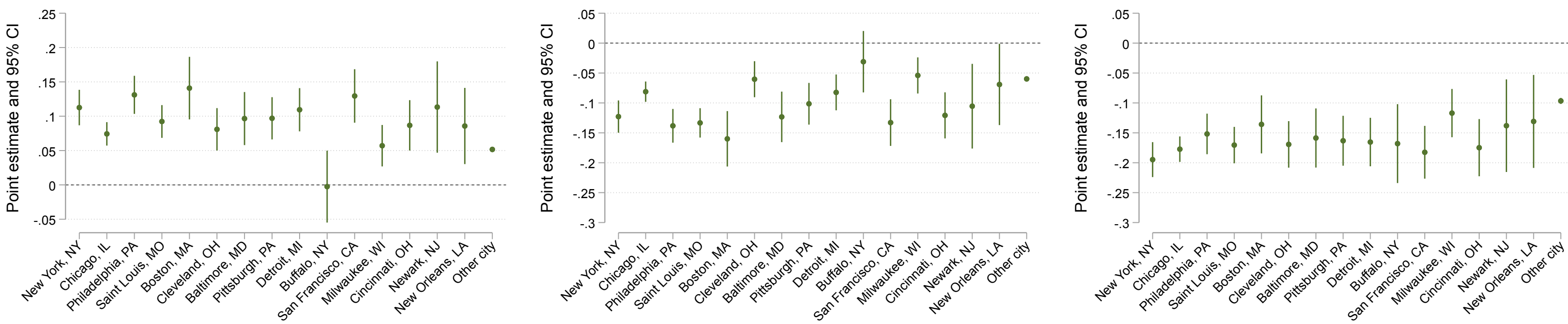
Notes: OLS regressions. Outcomes are measured in 1910. This table shows the result of a placebo test applying the instrument on the male sample. The instrument is the interaction between having an older sister and childhood parish proximity to Stockholm. All models include an indicator for having an older sister as well as fixed effects for birth order, birth year, number of siblings, having a disability, childhood parish, as well as family controls listed in Table 4. Standard errors clustered at the childhood parish level. *** - $p < 0.01$, ** - $p < 0.05$, * - $p < 0.1$.

TABLE C.8: IV SPECIFICATION TESTS

Panel A. Dependent variable:	Stockholm migrant			
	(1)	(2)	(3)	(4)
Prox to Sthlm × Has older sister	0.012***	0.012***	0.012***	0.013***
	(0.002)	(0.002)	(0.002)	(0.002)
Mean outcome	0.036	0.036	0.038	0.040
Panel B. Dependent variable:	**In labor force**			
Stockholm (=1)	0.425**	0.432**	0.410**	0.355*
	(0.176)	(0.174)	(0.191)	(0.187)
Mean outcome	0.177	0.177	0.178	0.182
Panel C. Dependent variable:	**Married**			
Stockholm (=1)	-0.507**	-0.491**	-0.440**	-0.450**
	(0.203)	(0.202)	(0.216)	(0.211)
Mean outcome	0.693	0.693	0.689	0.683
Panel D. Dependent variable:	**Any child**			
Stockholm (=1)	-0.310	-0.298	-0.187	-0.155
	(0.211)	(0.210)	(0.226)	(0.221)
Family controls	No	Yes	Yes	Yes
Exclude parishes far from Stockholm	No	No	Yes	No
Exclude northern parishes	No	No	No	Yes
Observations	274711	274711	254340	224798
First-stage F-statistic	42.7	43.6	36.5	37.5
Mean outcome	0.698	0.698	0.693	0.683

Notes: OLS (Panel A) and IV (Panels B–D) regressions. All outcomes are measured in 1910. All models include an indicator for having an older sister, and fixed effects for birth order, birth year, number of siblings, having a disability, and childhood parish. *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 fixed effects for mother's employment, father's 1-digit HISCO occupation, and whether the household is multigenerational, extended family (relatives only), or composite family (with non-relatives). Column 3 excludes parishes further than 510 kilometers from Stockholm, corresponding to the 90^{th} percentile, while column 4 excludes parishes located north of the city of Gävle. Standard errors are given in parentheses and are clustered at the childhood parish level. *** - $p < 0.01$, ** - $p < 0.05$, * - $p < 0.1$.

D External validity: evidence from U.S. migrants



(A) IN LABOR FORCE

(B) MARRIED

(C) ANY CHILD

FIGURE D.1: U.S. MIGRANTS OUTCOMES IN THE 15 LARGEST CITIES: FLFP, MARRIAGE, AND FERTILITY

Notes: This figure displays OLS estimates and 95 percent confidence intervals from a regression similar to Equation (4) where we include separate indicators for moving to each of the 15 largest U.S. cities by population in 1910 and all other (smaller) cities. The outcome is an indicator for being in the labor force (Panel A), being married (Panel B), or have any child (Panel C) in 1910. We include family fixed effects and the baseline set of individual controls: fixed effects for birth year and birth order, and an indicator for eldest sister. Standard errors are clustered at the family level.

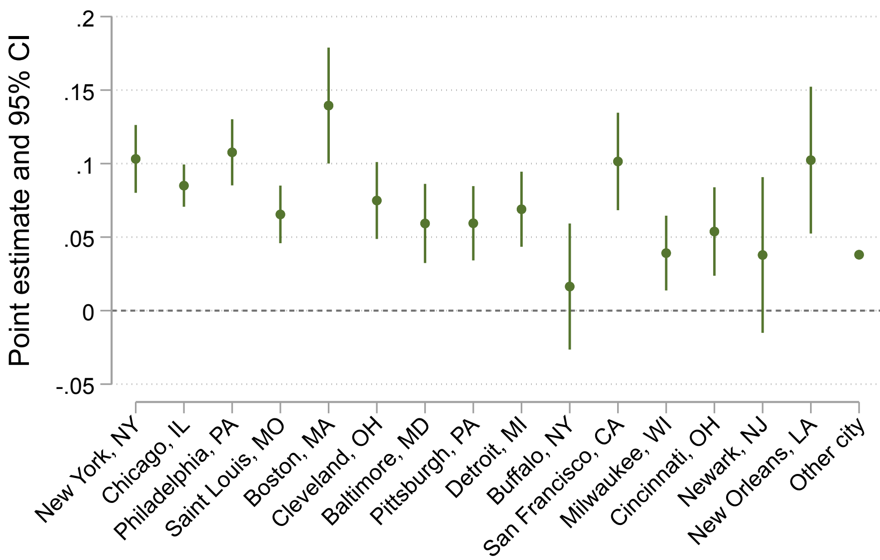


FIGURE D.2: U.S. MIGRANTS OUTCOMES IN THE 15 LARGEST CITIES: SERVICES

Notes: This figure displays OLS estimates and 95 percent confidence intervals from a regression similar to Equation (4) where we include separate indicators for moving to each of the 15 largest U.S. cities by population in 1910 and all other (smaller) cities. The outcome is an indicator for being employed in the services sector in 1910. We include family fixed effects and the baseline set of individual controls: fixed effects for birth year and birth order, and an indicator for eldest sister. Standard errors are clustered at the family level.

TABLE D.1: U.S. MIGRANTS OUTCOMES BY DESTINATION: EXCLUDING FARMING HOUSEHOLDS

Dependent variable:	In labor force		Married		Any child	
	All	Non-farm	All	Non-farm	All	Non-farm
	(1)	(2)	(3)	(4)	(5)	(6)
Migrant: large city (=1)	0.094***	0.066***	-0.103***	-0.083***	-0.167***	-0.138***
	(0.004)	(0.005)	(0.004)	(0.005)	(0.005)	(0.006)
Migrant: small city (=1)	0.051***	0.023***	-0.060***	-0.037***	-0.097***	-0.065***
	(0.002)	(0.003)	(0.002)	(0.002)	(0.003)	(0.003)
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes
Sibling fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	556946	271324	556946	271324	556946	271324
Mean outcome	0.151	0.193	0.845	0.788	0.672	0.574

Notes: OLS regressions. All outcomes are measured in 1910. *Migrant: large city* is an indicator taking the value 1 if the individual lives in one of the 15 largest U.S. cities by population in 1910, and zero if not. *Migrant: small city* is an indicator taking the value 1 if the individual lives in a city other than the 15 largest in 1910, and zero if not. *Individual controls* include fixed effects for birth year and birth order, and an indicator for eldest sister. *Family 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 D.2: U.S. MIGRANTS OUTCOMES BY DESTINATION: NON-SOUTH VS. SOUTH MIGRANTS

Dependent variable:	In labor force		Married		Any child	
	Non-South	South	Non-South	South	Non-South	South
	(1)	(2)	(3)	(4)	(5)	(6)
Migrant: large city (=1)	0.095***	0.094***	-0.103***	-0.103***	-0.165***	-0.181***
	(0.005)	(0.012)	(0.005)	(0.013)	(0.006)	(0.015)
Migrant: small city (=1)	0.055***	0.038***	-0.058***	-0.064***	-0.096***	-0.100***
	(0.002)	(0.004)	(0.002)	(0.004)	(0.003)	(0.005)
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes
Sibling fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	386826	170119	386826	170119	386826	170119
Mean outcome	0.153	0.145	0.840	0.859	0.633	0.762

Notes: OLS regressions. All outcomes are measured in 1910. *Migrant: large city* is an indicator taking the value 1 if the individual lives in one of the 15 largest U.S. cities by population in 1910, and zero if not. *Migrant: small city* is an indicator taking the value 1 if the individual lives in a city other than the 15 largest in 1910, and zero if not. *Individual controls* include fixed effects for birth year and birth order, and an indicator for eldest sister. *Family 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 D.3: U.S. MIGRANTS OUTCOMES BY DESTINATION: NATIVE VS. IMMIGRANT MIGRANTS

Dependent variable:	In labor force		Married		Any child	
	Native	Immigrant	Native	Immigrant	Native	Immigrant
	(1)	(2)	(3)	(4)	(5)	(6)
Migrant: large city (=1)	0.094***	0.172***	-0.103***	-0.204***	-0.167***	-0.239***
	(0.004)	(0.038)	(0.004)	(0.040)	(0.005)	(0.042)
Migrant: small city (=1)	0.051***	0.054**	-0.060***	-0.097***	-0.097***	-0.094***
	(0.002)	(0.021)	(0.002)	(0.020)	(0.003)	(0.024)
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes
Sibling fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	556946	5501	556946	5501	556946	5501
Mean outcome	0.151	0.176	0.845	0.815	0.672	0.706

Notes: OLS regressions. All outcomes are measured in 1910. *Migrant: large city* is an indicator taking the value 1 if the individual lives in one of the 15 largest U.S. cities by population in 1910, and zero if not. *Migrant: small city* is an indicator taking the value 1 if the individual lives in a city other than the 15 largest in 1910, and zero if not. *Individual controls* include fixed effects for birth year and birth order, and an indicator for eldest sister. *Family 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 D.4: U.S. MIGRANTS OUTCOMES BY DESTINATION: BLACK VS. WHITE MIGRANTS

Dependent variable:	In labor force		Married		Any child	
	White	Black	White	Black	White	Black
	(1)	(2)	(3)	(4)	(5)	(6)
Migrant: large city (=1)	0.094***	0.133***	-0.103***	-0.143***	-0.167***	-0.263***
	(0.004)	(0.044)	(0.004)	(0.047)	(0.005)	(0.048)
Migrant: small city (=1)	0.051***	0.140***	-0.060***	-0.135***	-0.097***	-0.203***
	(0.002)	(0.027)	(0.002)	(0.027)	(0.003)	(0.029)
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes
Sibling fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	556946	4567	556946	4567	556946	4567
Mean outcome	0.151	0.561	0.845	0.589	0.672	0.578

Notes: OLS regressions. All outcomes are measured in 1910. *Migrant: large city* is an indicator taking the value 1 if the individual lives in one of the 15 largest U.S. cities by population in 1910, and zero if not. *Migrant: small city* is an indicator taking the value 1 if the individual lives in a city other than the 15 largest in 1910, and zero if not. *Individual controls* include fixed effects for birth year and birth order, and an indicator for eldest sister. *Family 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 D.5: U.S. MIGRANTS OUTCOMES: ALTERNATIVE LINKS

Dependent variable:	In labor force (=1)	Married (=1)	Any child (=1)
	Panel A. All Census Tree links		
	(1)	(2)	(3)
Migrant: large city (=1)	0.094***	-0.103***	-0.167***
	(0.004)	(0.004)	(0.005)
Migrant: small city (=1)	0.051***	-0.060***	-0.097***
	(0.002)	(0.002)	(0.003)
Observations	556946	556946	556946
Mean outcome	0.151	0.845	0.672
	Panel B. Only Family Tree links		
	(1)	(2)	(3)
Migrant: large city (=1)	0.058***	-0.055***	-0.141***
	(0.005)	(0.005)	(0.007)
Migrant: small city (=1)	0.032***	-0.039***	-0.079***
	(0.002)	(0.002)	(0.003)
Observations	388691	388691	388691
Mean outcome	0.127	0.883	0.723
	Panel C. Only links /w 3+ methods		
	(1)	(2)	(3)
Migrant: large city (=1)	0.074***	-0.069***	-0.139***
	(0.007)	(0.007)	(0.009)
Migrant: small city (=1)	0.033***	-0.038***	-0.079***
	(0.003)	(0.003)	(0.004)
Observations	255035	255035	255035
Mean outcome	0.130	0.875	0.727
	Panel D. Census Tree w/o ML links		
	(1)	(2)	(3)
Migrant: large city (=1)	0.070***	-0.074***	-0.155***
	(0.004)	(0.004)	(0.006)
Migrant: small city (=1)	0.039***	-0.047***	-0.090***
	(0.002)	(0.002)	(0.003)
Observations	532691	532691	532691
Mean outcome	0.136	0.870	0.692
	Panel E. Census Tree w/o hints		
	(1)	(2)	(3)
Migrant: large city (=1)	0.090***	-0.099***	-0.166***
	(0.005)	(0.005)	(0.006)
Migrant: small city (=1)	0.050***	-0.059***	-0.096***
	(0.002)	(0.002)	(0.003)
Observations	500343	500343	500343
Mean outcome	0.146	0.854	0.685
	Panel F. Census Tree w/o implied links		
	(1)	(2)	(3)
Migrant: large city (=1)	0.077***	-0.082***	-0.157***
	(0.005)	(0.005)	(0.007)
Migrant: small city (=1)	0.039***	-0.047***	-0.086***
	(0.002)	(0.002)	(0.003)
Individual controls	Yes	Yes	Yes
Sibling fixed effects	Yes	Yes	Yes
Observations	404466	404466	404466
Mean outcome	0.133	0.874	0.714

Notes: OLS regressions. All outcomes are measured in 1910. *Migrant: large city* is an indicator taking the value 1 if the individual lives in one of the 15 largest U.S. cities by population in 1910, and zero if not. *Migrant: small city* is an indicator taking the value 1 if the individual lives in a city other than the 15 largest in 1910, and zero if not. *Individual controls* include fixed effects for birth year and birth order, and an indicator for eldest sister. *Family fixed effects* is a fixed effect for same sex siblings. Standard errors, in parentheses, are clustered at the 1880 household level. *** - $p < 0.01$, ** - $p < 0.05$, * - $p < 0.1$.

E Data appendix

E.1 Cross-country census data: stylized facts

Our analysis is motivated by the fact that FLFP and service jobs exhibit substantial variation within countries around the early 20^{th} century. To examine differences in FLFP within countries, we use individual-level data drawn from the 100% 1911 census for England and Wales, the 100% 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. We exclude parishes with very small populations where FLFP rates are imprecisely measured. 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).

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.

E.2 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.

E.2.1 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.

E.2.2 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., 2014b). 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. (2014b) 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, that is, 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 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.

E.2.3 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.[51] 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

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

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 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, that is, 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×birth year×place of birth cell that satisfies a Jaro-Winkler score of above a specified threshold for *both* the first and the last name and require that there is no close runner-up. To balance between an unrepresentative small sample and the risk of false positives, we have followed previous linkage methodologies using the Swedish censuses by setting the threshold at 0.85 (for example, see Berger et al. (2023) and Ejermo et al. (forthcoming)). To disallow matched candidates when there is a close contestant, we set the cut-off at a distance of 0.05 units between the highest ranked candidate and the runner-up. Following this procedure in step 2, 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.

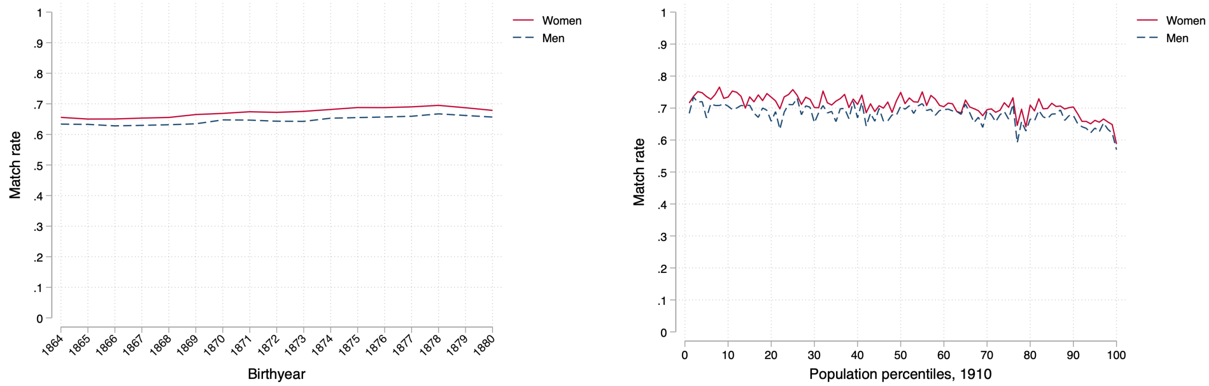
E.2.4 Linked sample and linkage rates

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.[52] 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 E.1), although more populated locations generally display somewhat lower name rates. The latter is expected due the fact that there are more potential 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 E.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.17).[53]

[52] The forward linkage rate is 43 percent.

[53] 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).



(A) Birthyear

(B) Location

Figure E.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.

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

	All mean	Linked mean
Age	7.328	7.218
Birth order	2.764	2.748
Eldest sister (=1)	0.477	0.480
Father's age	42.816	42.753
Mother's age	39.560	39.432
Mother in labor force (=1)	0.002	0.002
Father in labor force (=1)	0.921	0.928
Father white collar (=1)	0.138	0.140
Family members in household	6.253	6.207
Multigenerational family (=1)	0.060	0.061
HH type: Extended family (=1)	0.050	0.048
HH type: Composite (=1)	0.224	0.250
Observations	558959	272512

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