# Cities and the Rise of Working Women *

Thor Berger† Mounir Karadja‡ Erik Prawitz§ Martin Önnerfors¶

August 25, 2023

**Abstract**

This paper documents the unique role large cities played in women's economic and social advances in the early 20th century. We show that women in large cities in Europe and the United States were substantially more likely to be in the labor force, half a century before the aggregate rise in female labor force participation. To establish the role of large cities for women's labor-market advances, we turn to the case of Sweden where women can be linked over time in census data. Women moving to Sweden's largest city, the capital Stockholm, were about 50 percentage points more likely to enter the labor force and less likely to marry and have children compared to their migrant sisters. An early structural shift towards services and social interactions between working women partly explain the early labor-market advances of women in large cities.

*We thank Martín Fernández Sánchez, Raoul van Maarseveen, and JF Maystadt for helpful discussion 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@ifn.se

¶Institute for Housing and Urban Research, Uppsala University. E-mail: martin.onnerfors@nek.uu.se

Cities have throughout history been seedbeds of economic and social change. During American and European industrialization, millions of migrants left rural hinterlands in search of opportunity in the city. A less appreciated fact is that women were typically overrepresented among urban migrants. Women may have left the drudgery of rural life to pursue economic opportunity, better marriage market prospects, or social independence in the city. Yet we lack systematic evidence on how urban migration affected women since they can seldom be linked over time in historical census data.[1]

This paper shows that the rise of the modern metropolis played a central role in women's economic and social advances in early-20th century Europe and the United States. Using census data from Britain, Sweden, and the United States, we first document a discontinuity in female labor force participation (FLFP) with substantially higher FLFP rates in the most populous areas of each country.[2] While men found economic opportunity in rural and urban areas alike, only the largest cities provided labor-market opportunities for women. Notably, the higher FLFP rates in large cities predate the aggregate increase in FLFP that is largely a post-World War II phenomenon (Goldin, 1995; Olivetti, 2013).

To analyze how the rise of the modern metropolis affected women's economic and social outcomes, we focus on the case of Sweden where women can be linked in historical census data, which is rarely possible in other countries. Crucially, the Swedish censuses recorded women's maiden names allowing us to create a large and representative linked sample of women using automated record-linkage methods (Abramitzky et al., 2021). Using Swedish census data 1880—1910, we create a longitudinal dataset where we can follow women (and men) from their childhood households to their eventual migration to other rural and urban locations. Notably, our linkage algorithms yield equally high match rates for both men and women.

We first show that women who moved to the largest cities of Sweden saw substantial increases in labor-force participation. Notably, this sharply contrasts female migrants to less populous destinations that saw no changes in their employment status, or male migrants that experienced equally

[1]Young women can typically not be linked over time in historical census data using standard techniques since most women changed their surname upon marriage. Therefore, women are typically completely excluded from linked samples using historical data such as the pioneering Census Linking Project (Abramitzky et al., 2020). Some recent papers have relied on marriage records (Craig et al., 2019; Withrow, 2021) or genealogical data (Feigenbaum and Gross, 2020) to follow women over time in census data. While such methods are feasible to study subsets of women (e.g., telephone operators), obtaining large and representative linked samples of women is typically not possible given data constraints. A creative solution is devised by Olivetti and Paserman (2015) who creates "pseudo links" to include women in their study of intergenerational mobility, which is possible since their method does not require the actual linking of individuals over time. In contrast, Sweden's historical census data allows us to link both married and unmarried women using traditional census record-linking techniques that yields a large and representative sample of the underlying (female) population.

[2]By FLFP we mean reported, market-oriented work, rather than the often unpaid labor conducted on family farms that is often unreported in historical data. We address well-known issues relating to the underenumeration of (informal) female work in historical census data below.

high returns to migration regardless of the destination.

To corroborate these results, our main analysis zooms in on the largest city – Stockholm – around the turn of the 20th century. Stockholm is a paradigmatic case of an emerging metropolis in a rapidly industrializing country: between 1860 and 1910, the city tripled in size as Sweden was transformed from a stagnant agricultural economy to a modern industrial and urban nation. Throughout this period, more than half of Stockholm's population consisted of migrants, which were disproportionately female. To identify its role in shaping female work, we compare female migrants to Stockholm with their sisters migrating to other destinations. This has two main benefits. First, it allows us to control for all characteristics that are shared between siblings and that may influence both migration decisions and labor-market outcomes. Second, only comparing *migrant* sisters enables us to tease out the role of the largest cities rather than migration itself.

We then show that women who migrate to Stockholm are more than 50 percentage points more likely to be in the labor force compared to their sisters migrating to a rural location. While migrants to other urban areas were also more likely to enter the labor force, the effect is almost four times larger among those moving to Stockholm that underlines the unique role of the largest cities in driving FLFP rates. The empirical strategy allows us to rule out selection across families, but we cannot fully rule out the presence of within-family selection into migrant destinations. However, three complementary approaches all indicate that such selection has a very limited impact on our estimates.[3]

Women's labor-force participation is still today deeply intertwined with marriage and fertility decisions. In the early-20th century — the era of the "independent female worker" (Goldin, 2006) — women typically exited the labor force upon marriage. Indeed, the large increases in FLFP among female migrants to Stockholm is mirrored in sharp reductions in marriage and fertility rates, while migrants to other destinations (that saw no effects on FLFP) were *more* likely to marry and have children. Thus, while women often moved to improve their marriage-market outcomes, the rise of working women in large cities such as Stockholm is mirrored in a large fraction of women remaining single and childless.[4]

Why did large cities provide economic opportunity for women? A central explanation for the aggregate rise in FLFP during the late-20th century is the sectoral reallocation towards the service

[3]First, we show that estimates are very similar when excluding sibling fixed effects and individual controls, thus suggesting limited selection overall. Second, for the subset of migrants that move later in life, we control for lagged pre-migration outcomes to directly take into account within-sibling selection as well as reverse causation. These controls have minor impacts on our estimates. Third, we find little evidence that selection on unobservables drives more than a small share of our estimated effects using the approach developed by Oster (2019).

[4]We further document that the reduction in marriage and fertility rates does not reflect an increase in cohabitation (what was known as "Stockholm marriages" to contemporaries), nor an increase in out-of-wedlock births. Thus, while married female migrants were also significantly more likely to work in Stockholm, the overall increase in FLFP is mainly driven by women remained single and childless.

sector, which provided "respectable" jobs for women (e.g., Olivetti, 2013). We document a spatial dimension to this shift: large cities in Europe and the United States had experienced a shift towards the service sector more than half a century before other places within the country.[5] As a case in point, the share of employment in services in early-20th century Stockholm would be matched by other Swedish regions only in the late 1960s.

An early shift towards the service sector in large cities is mirrored in the fact that women moving to large cities such as Stockholm were disproportionately more likely to transition into jobs in the services sector. To corroborate the role of service jobs, we compare the FLFP between sisters migrating to more or less service-based urban economies that reveals a tight positive association between service sector size at destination and FLFP. While female migrants were more likely to take up service jobs across the skill spectrum, the bulk of migrants were concentrated in relatively low-skilled service jobs. However, the increase in FLFP rates does not seem to have come at the price of lower income and unhealthy working conditions in the city. Measuring economic returns using occupational income scores reveals that female migrants to Stockholm saw substantial increases in real incomes and we find that female migrants to Stockholm lived as long as their sisters migrating elsewhere.[6]

Although the early shift toward services is a key determinant of the labor-market advances of women, the modern metropolis may have provided other benefits. A particular advantage — emphasized already by contemporaries such as Alfred Marshall — is that big cities may facilitate social interactions. The role of such social interactions may be important for migrants if they facilitate the diffusion of information regarding employment opportunities, or social norms between working women (Fogli and Veldkamp, 2011). Interestingly, historians have argued that the lack of social networks hindered female migrants in Stockholm to integrate in the labor market and pioneering feminist novels such as Elin Wägner's *Men and Other Misfortunes* (1908) similarly underline the central role of female networks among working women the city.

To isolate the role of social interactions on migrants' outcomes, we link migrants identified in the census to unique administrative data (the *Roteman* archives) providing granular spatial information on the universe of Stockholm's inhabitants in each year that they reside in the city. The empirical strategy compares female migrants that arrive in the same year and reside in the same neighborhood, but reside in buildings with different levels of FLFP. The key identifying assumption is that more industrious migrants did not select into buildings with higher levels of FLFP, which we provide empirical support for by showing that there is no correlation between a wide range

[5] We corroborate this finding using regional employment data for 12 European countries 1900–2000 showing that the capital region (typically containing the largest metropolitan area) experienced a shift towards the service sector several decades before other regions.

[6] A non-negative effect on health outcomes among female migrants sharply contrasts the case of male migrants to Stockholm, who experienced a steep health penalty.

of migrant characteristics and the FLFP in a migrant's building of residence, once we control for arrival year-by-block fixed effects. In other words, while migrants sorted *across* neighborhoods, they did not select into particular buildings *within* those neighborhoods.[7]

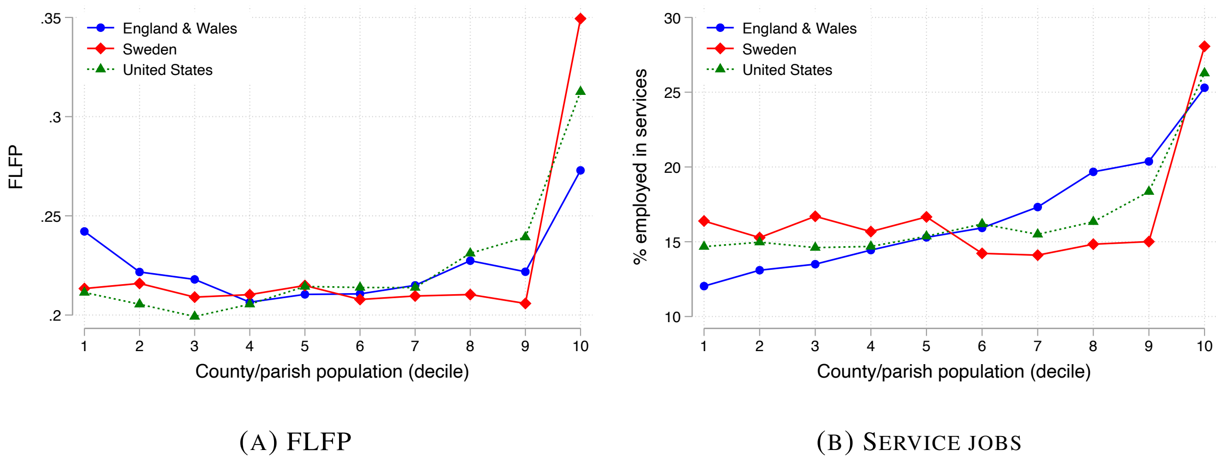
Our results show that social interactions with other working women are strongly and positively associated with female migrants' labor-force participation. Notably, the role of social interactions is mainly driven by migrants that are less likely to have pre-existing social networks in Stockholm. To provide direct evidence on one channel that social interactions may operate through, we show that migrants are particularly likely to take on the same occupation as the women in their building of first arrival. For example, a migrant that first arrives in a building with a larger fraction of seamstresses is more likely to become a seamstress herself. The fact that these effects are occupation-specific is seemingly most consistent with information diffusion (e.g., job referrals), rather than a diffusion of social norms.

Our paper provides new evidence of large spatial heterogeneities in FLFP within Europe and the United States more than a century ago. While industrialization is often associated with an aggregate decline in FLFP, we show that large cities during industrialization provided unique opportunities for female labor-market advances due to their precocious shift toward the service sector. In that sense, our paper contributes to an influential literature that studies aggregate trends in FLFP documenting that it follows U-shape over the course of economic development (Boserup, 1970; Goldin, 1990, 1995; Olivetti, 2014).[8] While the rise of working women is mainly seen as a post-World War II phenomenon in this literature, we document that that these aggregate changes conceal significant spatial variation. While some work document a key role for social and cultural influences on FLFP (Fernández et al., 2004; Fernández, 2013; Olivetti et al., 2020), we examine a setting where cultural and social norms were relatively similar and fixed. Thus, our results appear more consistent with 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). At the same time, we provide suggestive evidence of the role of social interactions for women's labor-market outcomes (Fogli and Veldkamp, 2011), which contributes to a recent literature that examines the role of gender-specific social interactions in shaping individual outcomes (Bell et al., 2019; Tan, 2022).

Lastly, our paper also provides novel evidence on the historical returns to migration for women. While a vast literature studies the returns to migration among male migrants (Long, 2005; Kennan and Walker, 2011; Abramitzky et al., 2012; Collins and Wanamaker, 2014; Ward, 2020), we show

[7] A lack of sorting is consistent with the fact that housing markets are very thin within narrow urban geographies. See Bayer et al. (2008) and Tan (2022) for similar methodological approaches.

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




*Notes:* This figure shows that FLFP (A) and the fraction of service jobs (B) was higher in more populous areas in England and Wales, Sweden, and the United States in the early 20th century. To construct the figures, we divide geographical units (parish or county) in each country into deciles based on their population and plot the mean FLFP and mean services share within each bin after absorbing country fixed effects.

FIGURE 1: FLFP AND SERVICE JOBS IN THE EARLY 20TH CENTURY.

that focusing on the returns for female migrants alter several stylized facts in the literature.[9] Most importantly, while male migrants faced a trade-off between improved economic outcomes and worse health outcomes, female migrants saw improved economic and unchanged health outcomes in the city. More fundamentally, we show that migration to the emerging metropolises of the early 20th century shaped women's employment, marriage, and fertility decisions in ways that ultimately led to the rise of what Goldin (2006) termed the "independent female worker".

# 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 20th century and was significantly higher in the most populous places in both Europe and the United States. To construct the figure, we use individual-level data on millions of women enumerated in the population censuses of Britain (1911), Sweden (1910), and the United States (1910).[10] We calculate the fraction of prime-aged (20–55) women that reported any gainful

[9]In particular, we relate to the literature on the returns to rural-to-urban migration (Young, 2013; Bryan et al., 2014; Hamory et al., 2021).

[10]To estimate FLFP, we use individual-level data drawn from the 100% 1911 census for England and Wales, the full-count 1910 census for Sweden, and a 1-in-100 random sample from the 1910 U.S. census obtained from IPUMS (2020).

occupation by their place of residence.[11] We divide counties and parishes into deciles based on their total population and plot the the mean FLFP rate within each bin normalized by the FLFP rate in each country. Thus, Figure 1A shows that FLFP rates in the most populous parishes in Sweden were about two thirds higher than in other parishes. Notably, for all three countries, there is a discontinuous increase in FLFP between the $9^{th}$ and $10^{th}$ decile showing that FLFP rates in Europe and the United States were substantially higher in the most populous areas.

What explains this variation in FLFP rates? An influential literature documents that FLFP follows a U-shape with respect to economic development at the country-level (Goldin, 1995; Olivetti, 2013). As a country industrializes, FLFP declines due to the separation of work from the home and the expansion of dirty and physically demanding industrial jobs. Conversely, the rising part of the U-shape is explained by the shift towards service jobs with less of a social stigma for female workers. Consequently, the labor-market entry of women in the aggregate has been seen mainly as a post-World War II phenomenon when the tertiary sector expanded.[12] However, the focus on aggregate structural change conceals the significant spatial variation in occupational composition.

A key factor explaining the higher FLFP rates in big cities is that they experienced the shift toward services more than half a century before less populated areas. Figure 1B shows that the largest places in England, Sweden, and the United States had a substantially higher fraction of service jobs in the early-20th century.[13] Again, we plot the mean share of employment in services across deciles of county or parish populations, which are normalized by the mean services share in each country. The earlier shift toward the service sector in more populous places is an empirical regularity that is evident across Europe. Figure 2A 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.[14] Thus, it took more than half a century for less densely populated areas to reach the service share attained in the most densely populated region already by the early 20th century.

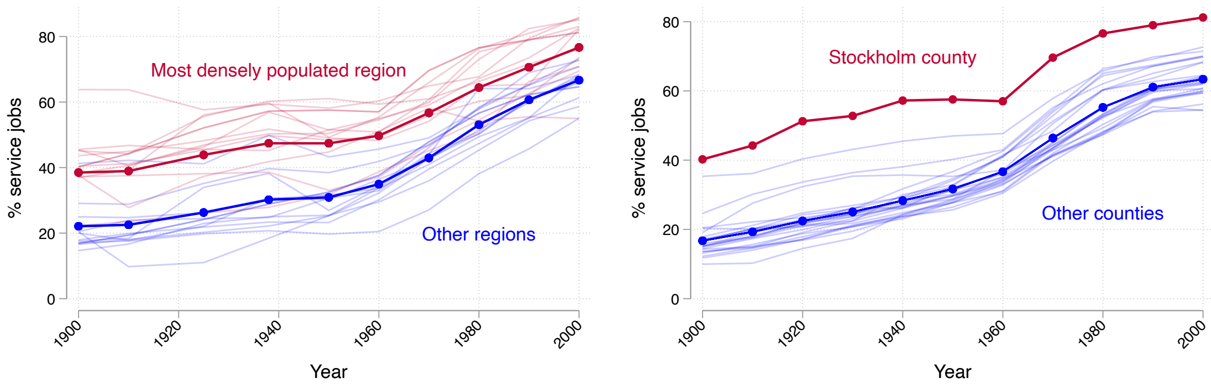
Sweden's rapidly growing capital — Stockholm — is a case in point. Already by the turn

[11] We assign individuals to the smallest administrative geographical unit available in the IPUMS data. In Britain and Sweden, individuals are assigned to their (civil) parish of enumeration where the total parish population is calculated directly from each respective census. For the U.S., we assign individuals to their county of enumeration (excluding overseas military installations) and add data on county populations from the 1910 census obtained via ICPSR.

[12] Several papers also argue for a direct role of World War II in accounting for the rising participation rates among women (Acemoglu et al., 2004; Fernández et al., 2004; Goldin and Olivetti, 2013).

[13] 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 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 4 (*Clerks*) and 5 (*Service workers and shop and market sales*) in the International Standard Classification of Occupations (ISCO) scheme.

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




(A) EUROPEAN (NUTS-2) REGIONS

(B) SWEDISH (NUTS-3) REGIONS

*Notes:* This figure shows that the most densely populated region in Sweden and other European countries had a higher share of employment in the service sector compared to less densely populated regions. A: The figure displays the share of service jobs in the most densely populated region over the period 1900–2000 in Austria, Belgium, Denmark, Finland, France, Germany, Italy, Norway, Portugal, Spain, Sweden, and the UK, as well as the unweighted average employment share in services for all other regions in the country. Data on the employment share of services is drawn from Rosés and Wolf (2018). B: The figure displays the share of service jobs in Stockholm county (containing the capital) and the (unweighted average) for all other counties. Data on the employment shares for Swedish counties is drawn from Enflo et al. (2014).

FIGURE 2: SERVICE JOBS IN EUROPEAN REGIONS, 1900–2000.

of the century, Stockholm had seen a shift towards the services sector at a time when the rest of the country was still rapidly industrializing. Figure 2B shows that Stockholm county (containing the capital) had achieved an employment share in services of about 40 percent by the early 20th century, which was matched by other counties only in the 1970s. The early shift towards services is also mirrored in very high FLFP rates: in 1910, more than 50 percent of prime-aged (20–55) women in Stockholm were part of the labor force (Appendix Table A.1). Notably, less than 40 percent of women residing in Stockholm were married or had at least one child. Stockholm thus sharply contrasts rural areas where about about 60 percent of women were married and had a child respectively, while just about 18 percent were part of the labor force.[15] Stockholm was also majority female, with 122 women per 100 men in 1910, which mirrors the fact that European cities tend to be majority female while rural areas have a surplus of men still today.

Our descriptive results in this section documents that there existed significant spatial variation in FLFP rates in early-20$^{th}$ century Europe and the United States. In particular, FLFP rates were substantially higher in large cities. However, a potential explanation for the higher FLFP rates in these urban areas is that women residing in large cities may be inherently different — e.g., in terms

[15]Similarly, Stockholm have substantially higher FLFP rates and lower marriage and fertility rates compared to women residing in other urban areas (Appendix Table A.1).

of ambition, human capital, or family background — from those residing in rural locations. We next describe how we link Swedish census data that allows us track women from their childhood households into the locations where they reside as adults, which allows us to tease out the effects of large cities on women's economic and social outcomes.

## 2 Data and Sample

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

**Linking procedure.** Focusing on children observed in their childhood home in 1880, we link these to the 1910 census when they are in their adulthood. To do so, we rely on probabilistic linking methods. 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.[16] The detail and accuracy of these time-invariant variables allow us to construct a relatively small set of candidate links. Second, we evaluate these candidate links by comparing first and last names. Importantly, censuses typically recorded women's maiden name instead of or together with their married name, 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 match within the same sex×birth year×place of birth cell that satisfies a Jaro-Winkler threshold of at least 0.85 for both the first and the last name and require that there is no close runner-up.[17] To complement our linked sample between the 1880 and 1910 censuses, we add links from the 1880 census to 1890 and 1900 based on Wisselgren et al. (2014), obtained from IPUMS International.

**Demographic and occupational data.** The census contains individual level data on year of birth, civil status, occupation, as well as family relationships between members of households. It also separates families within households in cases where multiple families reside together. Using this information, we identify siblings using a combination of parent and family indicators. Occupations are classified according to the HISCO system which identifies sector of work (Leeuwen et

[16]Since the Swedish censuses are not plagued by age-heaping, we only consider potential matches among candidates with the same exact birth year.

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

al., 2002). We also match occupations to HISCLASS, which categorizes occupations in terms of their skill requirements (Leeuwen and Maas, 2011). We use IPUMS's definition to measure (female) labor force participation. Individuals are defined as in the labor force if they have a reported occupation (excluding students and titles capturing noble ranks), her occupational title is not related to another household member (e.g., widow or daughter), and are above 15 years old. We also construct an alternative measure including unreported female work (see Section 4.1). Using data on the family in 1880, we collect data on fathers' and mother's occupations, as well as a number of characteristics at the household level.[18] Apart from parents' occupations, we obtain for each household the number of families and generations, siblings, servants, and (un)related members.

**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. We follow a non-parametric approach similar to Ward (2020).[19] 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 is 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.[20] In a second step, income scores are then adjusted by county-level CPI to account for regional cost differences, as well as urban-rural price differences within counties based on Collin (2016). At the same time, income scores are adjusted using estimated urban and Stockholm-specific income premia. These premia are calculated using a sample of 5,000 tax records from 1900, which include information on parish of residence, allowing us to distinguish income premia by urban and rural areas, including a specific premium for Stockholm. The premia for urban areas and Stockholm are 41 and 66 percentage points, respectively.

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

[18] Note, however, that less than 3% of our sample individuals' mothers have reported occupations in 1880.

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

[20] 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.

**Linked census sample.** Our sample consists of all individuals from rural parishes that were between 0 and 16 years of age in 1880 and that are also observed in 1910. To be able to compare siblings within families, we restrict our attention to individuals that have an identified same-sex sibling. Starting from the 1910 census, we are able to identify 66 percent of individuals to their household in 1880. This compares very favorably to the existing literature using linked historical census data, which has linked 20.3–21.9% in Britain and the US (Long, 2005) and 37% in Norway (Modalsli, 2017).[21] Linkage rates are similar for men and women with 65 and 67 percent, respectively.

In our main analysis, we keep migrants who live outside of their childhood parish in 1910. Comparing only migrants allows us to study the role of the destination and, in particular, large cities, rather than the role of migration itself. After this restriction, we end up with 57,652 women (and 50,874 men used for comparison). In Appendix Tables A.14–A.15, we also show results when comparing sisters migrating only to towns as well as an extended sample where we include also non-migrants.

While we are able to achieve relatively large match rates, it is possible that matched individuals differ systematically from those that are unmatched, possibly yielding unrepresentative estimates. For example, it is easier to link individuals with uncommon names, and name commonality has been linked to traits such as individualism and socio-economic status. With this in mind, Appendix Table A.17 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.18).[22]

**Administrative data.** To examine the effects of social interactions, we link females in the census data to the *Roteman* database, an administrative database covering all residents of Stockholm between 1878 and 1926, including information on individuals' building of residence, household, marriage status, and occupation. Observations are event-based and are updated upon significant life changes such as births, marriages, or when individuals move to a new building. In this application, we match our sample to the Roteman database and create a yearly individual panel data set covering the first five years of residence in Stockholm.

In order to link individuals observed in the 1910 census to the Roteman database, we proceed in two main steps. First, we make use of the links from the Death Index to the 1910 census (see above). Second, we link individuals from the Roteman database to the Death Index. The

[21] Ward (2020) links 9.1 % of linkable sons in a triple-linked sample from US censuses 1910–1940.

[22] 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).

main reason for taking this detour is that both the Roteman database and the Death Index have detailed information on birth dates (and not only birth years), which allows us to improve on the match rate by only comparing name similarities between same sex individuals born on the same day.[23] The forward match rate for women born 1864–1880 in step 2 is 73 percent (71 percent for men). Considering the subset of individuals observed in Stockholm in the 1910 census, we can ultimately match 80 percent of women (79 percent of men) to the Roteman database using our two-step linking procedure.

# 3 Empirical Strategy

We start our analysis by estimating the effects of migration to destinations of different population size using our full sample of sisters living in rural areas in 1880:

$$Y_{ih} = \sum_{p=1}^{100} \beta^p Migrant_i^p + \gamma \mathbf{X}_i + \delta \mathbf{X}_h + \varepsilon_{ih} \tag{1}$$

where $Y_{ih}$ is the outcome of interest for individual $i$ from household $h$. $Migrant_{ih}^p$ is a binary variable taking value 1 if the individual has moved to a parish at population percentile $p$ by 1910, and value 0 if the individual remains in the origin parish. Thus, $\beta^p$ is a set of coefficients estimated for each population percentile. For simplicity of exposition, we present these estimates graphically below. Moreover, $\mathbf{X}_i$ is a matrix of individual fixed effects for birth year, birth order, and being the eldest sister, while $\mathbf{X}_h$ is a matrix of household-level controls.[24]

To isolate the role of migrating to the largest urban areas (motivated by the results from estimating the above equation), we focus on a sample of migrants and estimate the effect of migration to Stockholm, the largest and fastest growing city during our period of study. To do so, we estimate

$$Y_{ih} = \beta_1 SthlmMigrant_i + \beta_2 OtherUrbanMigrant_{is} + \gamma \mathbf{X}_i + \delta \mathbf{X}_h + \varepsilon_{ih} \tag{2}$$

where $SthlmMigrant_i$ is an indicator taking value 1 if the individual lives in Stockholm in 1910, and 0 otherwise. To differentiate between rural migrants and migrants to other urban areas, we also introduce an additional indicator variable in our main regressions capturing if the sister migrates to another urban area than Stockholm. This allows us to interpret $\beta_1$ as the difference in outcomes

[23] At the same time, the information on birth parish (which we use when linking census individuals) in the Roteman Archive is less accurate.

[24] Household-level controls include: a full set of origin county fixed effects; father's percentile income rank; family size; a set of dummies capturing the mother's LFP, the father's major (1-digit) HISCO group, the presence of servants in the household and whether a household is a married/cohabitating couple with children, single-parent family, extended family (relatives only), or composite (family and non-relatives) as well as whether the family is multigenerational.

for migrants to Stockholm relative to the counterfactual of remaining in a rural area.[25]

While the above models have the advantage of controlling for a number of pre-migration variables at the individual and the household level that might influence both the decision to migrate and outcomes such as employment or income score, there may still be many unobserved factors that can lead to bias. In order to improve on this, we can limit the analysis to comparing outcomes between siblings who live in the same household in 1880.[26] By restricting attention to within-siblings variation, we hold constant all common parental influences on children, whether environmental or inherited, that may otherwise have a direct effect on both migration and our outcomes of interest. The empirical model thus becomes

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

where $s$ identifies sets of siblings and $\phi_s$ is the sibling fixed effect. Since the sibling effect absorbs any variation that does not vary within the family, the household level controls are subsumed in this specification. All models use cluster-robust standard errors at the level of 1880 households.

In the within-sibling model, the identifying assumption is that migration status in 1910 should be as good as randomly assigned across individuals of the same sex within a family, conditional on birth year and birth order. However, it remains possible that migrants to Stockholm are inherently different from their siblings migrating elsewhere, for example in terms of risk preferences or labor market ability.[27] In order to account for such possibilities, we also estimate first-difference models that control for adult pre-migration outcomes. Making use of the panel structure of our data, we also estimate models that restrict attention to individuals who migrate after 1900, allowing us to control for outcomes in 1900, when our sample is 20–36 years old. The model then becomes

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

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

[25]In Appendix Tables A.14–A.15, we show that results are similar when we include non-migrating sisters in the sample as well as when focusing only on migrants to urban areas.

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

[27]In Appendix Table A.3, we show limited selection in terms of observables (age, birth order, or being the eldest sister.

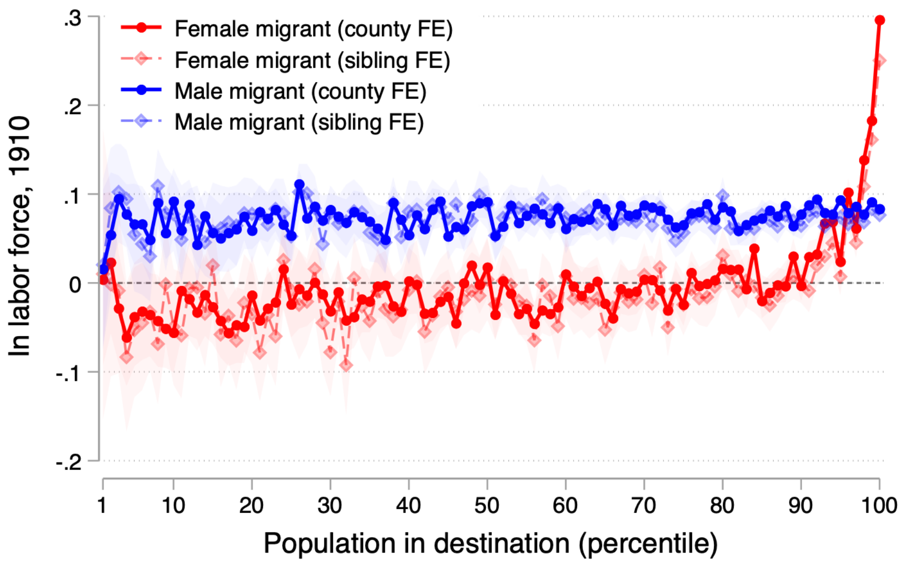



FIGURE 3: MIGRATION AND EMPLOYMENT BY POPULATION IN DESTINATION

*Notes:* This figure displays OLS estimates of Equation (1) where the outcome is an indicator capturing whether an individual is part of the labor force in 1910. The figure plots point estimates and 95 percent confidence intervals that capture the returns to migrating to different destinations ranked by their population size. We report separate estimates for females (red) and males (blue). Solid lines denote a specification using county fixed effects, while shaded lines correspond to estimates from a regression that includes family fixed effects. Standard errors are clustered at the family level.

and family formation.

# 4 Results

## 4.1 Main results: FLFP, marriage, and fertility

Figure 3 displays results from estimating Equation (1) with separate coefficients by population percentile of the destination parish. Moreover, the dashed shaded lines indicate estimates including sibling fixed effects, which are generally very similar to the baseline model estimates. The figure displays a marked non-linearity for females and indicates that migrants moving to the largest parishes were much more likely to transition into formal employment compared to migrants moving to more sparsely populated areas. Interestingly, the difference in employment is essentially the same for non-migrants and migrants up until the $90^{th}$ percentile. However, the difference in employment at the top percentile of population rises to approximately 30 percentage points compared to non-movers. In sharp contrast to female migrants, the relationship between migration and labor force participation is always positive for men but remains flat across the population distribution.

Motivated by the sharp discontinuity above, we turn to Equation 2 and estimate the effect of moving to Stockholm in particular, the largest and most densely populated city. To isolate the role of the city rather than migration itself, we make use of a sample of only migrants. Table 1 displays estimates for the effect of migration to Stockholm on having a reported occupation in 1910. Column 1 of Panel A shows that female migrants to Stockholm were 49.6 percentage points more likely to have an occupation recorded in the census, relative to all other female migrants. To compare Stockholm migrants with women that stayed in the rural economy, we add an indicator variable capturing migration to other towns than Stockholm. In line with Figure 3, this has only a modest effect on our estimate, which increases slightly to 53.9 percentage points.[28] While migrating to another urban area also implied an increase in FLFP, it is not close to the sizeable relationship with migrating to Stockholm. Column 3 adds fixed effects for the individuals' birth year, birth order, being the eldest sister and a variety of household characteristics in the 1880 census (see Section 3). Adding these controls makes only minor changes to the estimate. Finally, column 4 introduces sibling fixed effects as in equation 3, and thus bases its estimate by only comparing siblings of the same sex who lived in the same household in childhood. Capturing all common nature and nurture effects that may influence both migration status and the outcome, the estimated effect of migrating to Stockholm is marginally reduced to 49.7 percentage points. Relative to the average outcomes for women at 22.7 percent, the effect of migrating to Stockholm thus leads to more than a doubling of reported employment.

**Underreporting of occupations.** A key empirical concern when studying employment of women is that (informal) female work is often under-reported in historical census data (Goldin, 1990; Stanfors, 2014). In particular, women in agricultural households are especially believed to have engaged in work that was not recorded in the census due to its informal nature. To address this issue, Table 1, column 5, drops all women who lives with a male household head (either her father or husband) engaged in farming. This yields an estimate of 44.8 percentage points, which is strikingly similar in magnitude to our preferred estimate in column 4.

While our deliberate focus is on market-oriented work, we also provide estimates using an adjusted measure of female labor force participation including informal work in column 6. To account for informal work, FLFP is here defined to be equal to one also for women living with a male household head who is either engaged in farming or a working proprietor in sales or services.[29] With this conservative adjustment, the magnitude drops to 22.9 percentage points. Although the new outcome variable deviates from our focus on market-oriented work, which arguably entailed

[28] In Appendix Table A.14, we alter the sample and document similar results when using an extended sample including also non-migrants as well as a smaller sample using only urban migrants.

[29] The adjustment for women living with working-proprietor household heads has only minor implications for the estimates.

TABLE 1: MIGRANTS' LABOR FORCE PARTICIPATION

| Dependent variable: | In labor force (=1) | | | | | |
|---|---|---|---|---|---|---|
| | | | | | HH not farmer | Adjusted |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Migrant: Stockholm (=1) | 0.491*** | 0.536*** | 0.548*** | 0.497*** | 0.448*** | 0.229*** |
| | (0.007) | (0.007) | (0.007) | (0.010) | (0.012) | (0.011) |
| Migrant: other urban area (=1) | | 0.172*** | 0.166*** | 0.132*** | 0.087*** | -0.093*** |
| | | (0.004) | (0.004) | (0.005) | (0.007) | (0.006) |
| Individual controls | No | No | Yes | Yes | Yes | Yes |
| Household controls | No | No | Yes | No | No | No |
| Sibling fixed effects | No | No | No | Yes | Yes | Yes |
| Observations | 79850 | 79850 | 71301 | 79850 | 49815 | 79850 |
| Mean outcome | 0.227 | 0.227 | 0.227 | 0.227 | 0.307 | 0.491 |

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

a different type of independent labor force participation, it is noteworthy that our estimate for migration to Stockholm still has a sizeable positive magnitude as Stockholm migrants are compared to sisters staying in rural areas. In contrast, sisters migrating to other urban areas see a negative relationship with FLFP including informal work.

In sum, these results show that female migrants to large cities in general and Stockholm in particular experienced substantial increases in employment. However, women in the early-20th century typically faced a choice between having a family and having a job, as virtually all women exited the labor force upon marriage.[30] We next show that the vast majority of female migrants to Stockholm that took up formal employment remained unmarried and childless.

**Marriage, fertility, and FLFP.** Figure 4 displays estimates based on Equation (1) where the outcome is an indicator capturing whether an individual is married or has at least one child in 1910.[31] Again, we report separate coefficients by the population percentile in the destination. The estimates in Figure 4A and 4B show that female migrants moving to destinations in the bottom-90

[30] In our main sample, only 1.7 percent of married women in 1910 had a reported occupation.
[31] See Appendix Figure A.2 for the male sample.


(A) MARRIAGE

(B) ANY CHILD

FIGURE 4: FEMALE MIGRANTS WERE LESS LIKELY TO MARRY AND HAVE CHILDREN IN BIG CITIES.

*Notes:* This figure displays OLS estimates of Equation (1) where the outcome is an indicator capturing whether an individual is married (A) or has any child (B) in 1910. The figure plots point estimates and 95 percent confidence intervals that capture the returns to migrating to different destinations ranked by their population size. Solid lines denote a specification using county fixed effects, while shaded lines correspond to estimates from a regression that includes family fixed effects. Standard errors are clustered at the family level.

percent of the population distribution were significantly *more* likely to marry and have at least one child respectively. In sharp contrast, migrants to the most populous parishes were significantly *less* likely both to marry and have a child. Estimates are very similar when including sibling fixed effects reported as dashed shaded lines in both figures.

We next examine the marriage and fertility decisions among female migrants to Stockholm. Table 2 column 1 indicates that women in our within-sibling design are 50.5 percentage points less likely to be married in 1910 if they migrated to Stockholm between 1880 and 1910 as compared to sisters migrating to rural locations. Relative to the average among women of 70.9 percent, this represents a large decrease of about 70 percent. Similarly, female migrants were 47.1 percentage points less likely to have any children by 1910, which again is a large effect compared to the sample mean of 70.6 percent (column 2). Analogously to the case of FLFP, migrants to other urban areas see much more modest decreases. While most female migrants that took up formal employment in Stockholm at the same time remained single and childless, the increases in employment are not solely driven by the lower marriage rates among migrants. Table 2, columns 3, reports estimates where we limit the sample to sister pairs where both were married in 1910. While the the coefficient is much more modest as compared to the main sample, it represents a large increase compared to the sample mean.

A potential concern regards the timing of marriage and migration decisions. For instance, did single sisters move to Stockholm and married sisters to rural areas? The last two columns of Table

TABLE 2: MIGRANTS' MARRIAGE, FERTILITY, AND FLFP

| Dependent variable: | Married (=1) | Any child (=1) | In labor force (=1) | | |
|---|---|---|---|---|---|
| | | | Married 1910 | Unmarried 1910 | Unmarried 1900 pre-migration |
| | (1) | (2) | (3) | (4) | (5) |
| Migrant: Stockholm (=1) | -0.505*** | -0.471*** | 0.031*** | 0.201*** | 0.555*** |
| | (0.010) | (0.010) | (0.009) | (0.019) | (0.057) |
| Migrant: other urban area (=1) | -0.119*** | -0.102*** | 0.016*** | 0.128*** | 0.147*** |
| | (0.006) | (0.006) | (0.003) | (0.016) | (0.033) |
| Individual controls | Yes | Yes | Yes | Yes | Yes |
| Sibling fixed effects | Yes | Yes | Yes | Yes | Yes |
| Observations | 79850 | 79850 | 46102 | 11595 | 3228 |
| Mean outcome | 0.709 | 0.706 | 0.017 | 0.739 | 0.300 |

*Notes:* OLS regressions. All outcomes are measured in 1910. *Stockholm migrant* is an indicator taking value 1 if the individual lives in Stockholm city, and 0 if in a rural area. *Individual controls* include fixed effects for birth year and birth order, and an indicator for eldest sister. *Sibling fixed effects* is a fixed effect for same sex siblings. Standard errors are given in parentheses and are clustered at the 1880 household level. *** - $p < 0.01$, ** - $p < 0.05$, * - $p < 0.1$.

2 address this issue by focusing solely on unmarried women. In column 4, we restrict the sample to sisters where both were unmarried in 1910. This restriction cuts our main estimate by more than half (see Table 1), which confirms that a substantial part of the employment decisions is related to marital status. However, even compared to single sisters migrating elsewhere, those migrating to Stockholm were 20.1 percentage points more likely to report an occupation. In column 5, we instead restrict the sample to sister pairs observed as unmarried in 1900 prior to migration. While both estimates are positive and substantial, the latter documents a larger and more similar magnitude as compared to our main estimate in Table 1. This is consistent with the notion that the positive effects on FLFP for Stockholm migrants are not driven by marriage choices prior to migration.

**Robustness.** One worry about the empirical strategy of comparing siblings is that individual differences even between siblings may be important determinants of migration or labor market outcomes. While estimates in Table 1 include controls for age, birth order, and being the eldest sister, we can additionally use pre-migration employment status for the subset of individuals who move to Stockholm later in life. We show in Appendix A.1 that our results are robust among the subset of migrants that move after 1900 where we can directly control for a pre-migration lagged outcomes that may differ within siblings pairs. We further show in Appendix A.2 that our findings are similar when applying twin rather than sibling fixed effects to further account for unobserved within-sibling differences.[32]

To further examine the potential of selection on unobservables in determining our results, we also apply the method of Oster (2019). This method compares the differences in models with and without controls to give an estimate of what the estimate would be if one could control for all unobserved characteristics. In our main sample, we find small differences between the controlled and uncontrolled regressions. For our key outcome capturing FLFP, the method suggests a corrected estimate of 54.0 percentage points, which is close to our main estimate 49.7 (see Table 1, column 4).[33] Using our sample of later migrants, which allows us to add pre-migration lagged outcomes, we find a similar pattern. The corrected estimate becomes 57.1 percentage points, which can be compared to 54.9 (see Appendix Table A.19, column 2).[34] We conclude that taking into account potential unobserved characteristics at the individual level that may vary within sibling groups, our estimate still remains highly economically and statistically significant.

We address a number of additional empirical concerns in the Online Appendix. First, Stockholm is known for having started the trend of cohabitation as an alternative to marriage (so-called *Stockholm marriages*). To the extent that Stockholm migrants chose to cohabit instead of marrying, the results for marriage in Table 2 may be misleading. We show in Online Appendix Table A.5 that the increase in cohabitation for female migrants is small in magnitude.[35] Second, our census data contains information on children born out of wedlock, which can be used as an alternative measure of informal marriages. Using these data, column 3 of Online Appendix Table A.5 shows that female migrants are less likely to have any children even when restricting attention to those born out of wedlock. Third, we show that the results pertaining to marriage and fertility are not driven by the fact that migrants tend to be younger than their siblings, and thus are less likely to form families.[36] Fourth, we also show that our results are unlikely to be driven by the fact that

[32]Additionally, Appendix Figure A.3 documents that our results are similar when accounting for: i) potential age differences between migrants and their siblings; and ii) whether the family sent only one or at least two migrants; and iii) whether a migrant's family size is above or below the median.

[33]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 $R^2$.

[34]For our outcomes on married and any children, we get -43.3 and -44.0 percentage points as corrected estimates, respectively.

[35]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.

[36]Panel A of Appendix Figure A.3 displays results when restricting the sample to include only siblings that are either younger or older than the oldest migrant. Results are very stable across these sample splits. To test more directly for the role of age, we estimate heterogeneous effects of migration by age in 1910, exploiting the fact that migrants are between 30 and 46 years old when observed in that year. If age is the explanatory factor, we would expect older migrants to have different outcomes than younger ones. Table A.7 shows our results, which indicate only minor difference in marriage or having children across older or younger migrants. The remaining panels of Figure A.3 perform similar sample splits as above for two other variables: whether the migrant's family size is above or below the median (panel B), and whether the family sent only one or at least two migrants (panel C. Across both variables and for all outcomes, estimates are similar and indicate little heterogeneity.

migrants may move away from Stockholm once they are ready to start a family.[37]

## 4.2 Cities and FLFP: The Role of Services

We documented above that the substantially higher FLFP rates in the most populous areas in America and Europe is mirrored in a higher fraction of service jobs in these areas. Stockholm was a case in point and had experienced a significant shift towards the service sector by the early-20th century, which would be matched in other places in the country only in the late 20th century. We next proceed to document the key role of the service sector to account for the higher FLFP rates among female migrants to Stockholm.

Figure 5 displays estimates from Equation (1) where the outcome is the probability that a migrant works in the service sector. As above, we report separate coefficients by the population percentile in the destination. Female migrants were typically not more likely to work in the service sector, apart from in the most populous places where migrants were more than 20 percentage points more likely to transition into service work.

We next corroborate these results by showing that the increase in FLFP among female migrants to Stockholm is similarly driven by an increased transition into service jobs. Table 3, column 1, shows that service employment increases by 41 percentage points among female migrants, while the increase in industrial occupations is more muted (column 2) and similar to migrants to other urban areas.

To further substantiate these results, Figure 6 displays the heterogeneous effect of migration on FLFP based on the supply of service jobs in the destination displayed along the horizontal axis. We control for family fixed effects and the same set of individual controls as above, which means that the results are identified from sisters that migrate to destinations with a different supply of service jobs. Clearly, female migrants that moved to areas with a larger fraction of services jobs were substantially more likely to enter the formal labor market compared to their sisters moving to other locations.

While we find very large increases in service employment among female migrants to Stockholm, an important question is whether they transitioned into lower-skill service jobs (e.g., domestic services), or whether they also took on more qualified employment (e.g., as teachers)? To

[37] To investigate the role of migrants who move away prior to 1910, we include all ever-movers to Stockholm in the sample in Appendix Table A.8. While effect sizes are in the similar direction, they are substantially smaller than for women who remain in Stockholm in 1910. Thus, it is possible that some women move away in order to start families, although on average this group still is less likely to be married or have children compared to never-movers. For men, however, the equivalent effect on marriage is in fact positive compared to non-movers, and zero for having children. Thus, the results indicate that males who leave Stockholm largely converge to the outcomes of non-migrants, while female migrants' family formation is significantly lower. This is consistent with the fact that even temporary female migrants have higher labor force participation of 5 percentage points, as shown in column 1, indicating a substitution away from family formation.

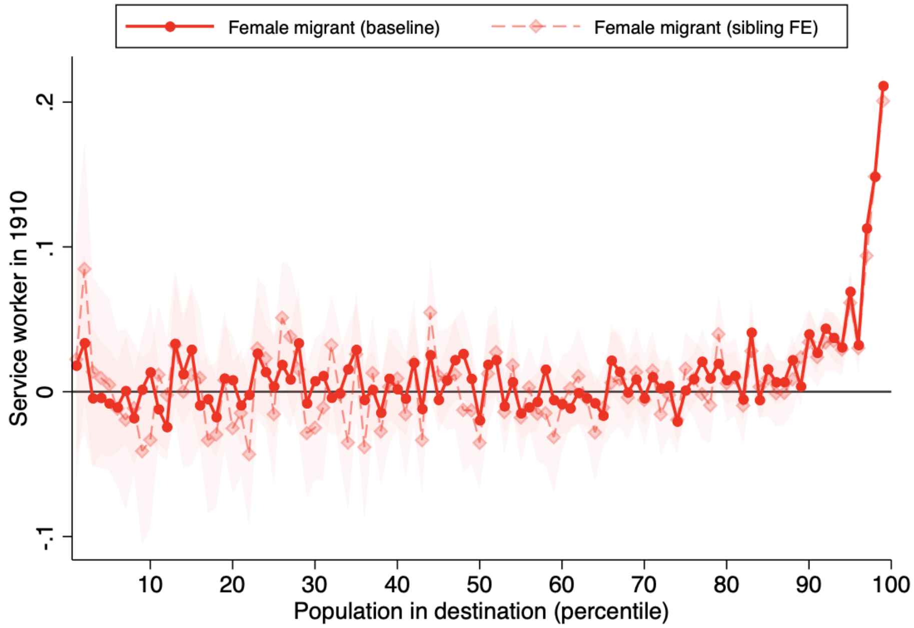



FIGURE 5: MIGRATION AND SERVICE EMPLOYMENT BY POPULATION IN DESTINATION

*Notes:* This figure displays OLS estimates of Equation (1) where the outcome is an indicator capturing whether an individual works in a service occupation in 1910. The figure plots point estimates and 95 percent confidence intervals that capture the returns to migrating to different destinations ranked by their population size. Solid lines denote a specification using county fixed effects, while shaded lines correspond to estimates from a regression that includes family fixed effects. Standard errors are clustered at the family level.

examine differences in job "quality", we proceed to study the skill-requirements of occupations, incomes, and health risks associated with urban jobs.


FIGURE 6: MIGRATION, FLFP, AND THE SUPPLY OF SERVICE JOBS IN DESTINATION

*Notes:* This figure presents a binned scatterplot of the probability that a female migrant is in the labor force based on the services share in the destination parish. We control for family fixed effects and include the full set of individual controls described in the main text. We also show a best fit line estimated on the underlying data.

TABLE 3: MIGRANTS' INCOME AND JOB QUALITY

| Dependant variable: | Sector of employment | | Occupational skill | | | Occ. income score | | | Age at |
|---|---|---|---|---|---|---|---|---|---|
| | Services | Industry | High-skill | Low-skill | Unskilled | ln(Income) | | ln(H. income) | death |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| Migrant: Stockholm (=1) | 0.404*** | 0.077*** | 0.073*** | 0.390*** | 0.034*** | 0.071*** | 0.221*** | 0.205*** | -0.103 |
| | (0.009) | (0.006) | (0.007) | (0.009) | (0.004) | (0.023) | (0.012) | (0.011) | (0.349) |
| Migrant: other urban area (=1) | 0.092*** | 0.046*** | 0.026*** | 0.095*** | 0.014*** | -0.180*** | -0.169*** | -0.077*** | 0.315 |
| | (0.004) | (0.003) | (0.004) | (0.004) | (0.002) | (0.017) | (0.007) | (0.006) | (0.195) |
| Individual controls | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Sibling fixed effects | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Occupation fixed effects | No | No | No | No | No | No | Yes | No | No |
| Observations | 79850 | 79850 | 79850 | 79850 | 79850 | 8353 | 8238 | 68952 | 61734 |
| Mean outcome | 0.114 | 0.061 | 0.080 | 0.131 | 0.013 | 7.002 | 7.000 | 6.846 | 73.636 |

*Notes:* OLS regressions. All outcomes are measured in 1910. *Stockholm migrant* is an indicator taking value 1 for residents of Stockholm city and 0 for rural residents. *Individual controls* include fixed effects for birth year and birth order, and an indicator for eldest sister. Standard errors, in parentheses, are clustered at the 1880 household level. *** - $p < 0.01$, ** - $p < 0.05$, * - $p < 0.1$.

**Skills.** We first document that while female migrants took on jobs across the skill spectrum, the most substantial effects are found among low-skill jobs. To classify non-agricultural occupations by their skill requirements, we use the HISCLASS scheme following Maas and van Leeuwen (2005). Table 3, column 3, shows that migrating women are 7.1 percentage points more likely to work in high (including medium) skilled occupations. Columns 4 and 5 show that female migrants are respectively 39.5 and 3.4 percentage points more likely to work in low-skilled and unskilled occupations compared to their sisters migrating to rural locations.[38]

**Income.** We next document that female migrants experienced significant increases in income. Zooming in on working women, Table 3 column 6 shows that female migrants have about 7.3 percent higher income scores compared to their sisters migrating to rural locations and even higher compared to sisters migrating to other urban areas.[39] Column 7 shows that this effect is even stronger when we introduce a fixed effect for the occupation of the migrant. Thus, female migrants to Stockholm obtained higher income also within their occupation.

While female migrants to Stockholm that engaged in market work earned more than those migrating elsewhere, the differential marriage rates could still imply that they experienced lower living standards. Column 8 shows results from a regression with the log of household income per adult household member, depicting an increase of 20 percent.[40] Thus, if anything female migrants to Stockholm appear to have increased their disposable income.[41] In Appendix Table A.16, we also show that the intra-household ratio of female to male income is higher for Stockholm migrants.[42] This is suggestive of that female migrants to Stockholm enjoyed an increased economic independence.

[38]Table A.4 shows results when including agricultural occupations in the skill categories.

[39]A related question is whether female migrants to Stockholm experienced higher rates of intergenerational mobility. Appendix Figure A.4 displays the association between children's and parents income ranks. Along the horizontal axis, we plot the father's income rank in the 1880 census. On the vertical axis, we display the mean income rank attained by female (non-)migrants in adulthood (i.e., in the 1910 census). Red dots indicate the average income ranks attained by Stockholm migrants, while blue dots correspond to the mean income of non-migrants. Migrants on average attain higher income ranks conditional on their father's income, which correspond to a higher level of absolute mobility. Figure A.4 also presents estimates of relative mobility, as captured by the rank-rank slope of daughters' and fathers' income ranks, which corresponds to the correlation between parents and children's place in the income distribution. The smaller rank-rank associations among migrants indicates a lower degree of intergenerational persistence (i.e., a higher level of relative mobility). Together, these results show that migrating to Stockholm was an avenue to upward mobility that weakened the influence of family background on earnings.

[40]In contrast, migration to Stockholm is associated with less household income when we do not adjust for household size. This is expected given that migrants to rural areas are more likely to marry.

[41]While this assumes that households shared their income equally, we show in Appendix Table A.16 that the effect is even more pronounced if we down-weight the contribution of the spouse. This suggests that the increase in household income is not driven by Stockholm migrants matching with spouses with higher income.

[42]For completeness, we set this ratio equal to one if the woman lives in her own household.

**Health.** Urban areas were historically characterized by worse health and shorter lifespans, which make the welfare gains from urban migration more ambiguous. We next document that female migrants in Stockholm did *not* experience worse health outcomes than sisters migrating to rural destinations. Table 3 column 9 displays estimates showing that female migrants in Stockholm by 1910 lived as long as their sisters migrating elsewhere. In contrast, male migrants to Stockholm on average died about 2 years earlier than their brothers (Appendix Table A.7), which reflects the well-known health hazards of big cities. The differential mortality rates by sex implies that general health hazards at the city level (e.g., poor sanitation and hygiene) are not the main explanation for mortality differences among migrants. Instead, a potential explanation for the positive health effects among female migrants is that they transitioned into less hazardous service jobs. Indeed, adding fixed effects at the occupation level decreases the estimate although it remains insignificant (Appendix Table A.7, column 2).[43] In any case, these results reveal that female migrants to Stockholm did not face a trade-off between economic and health outcomes.

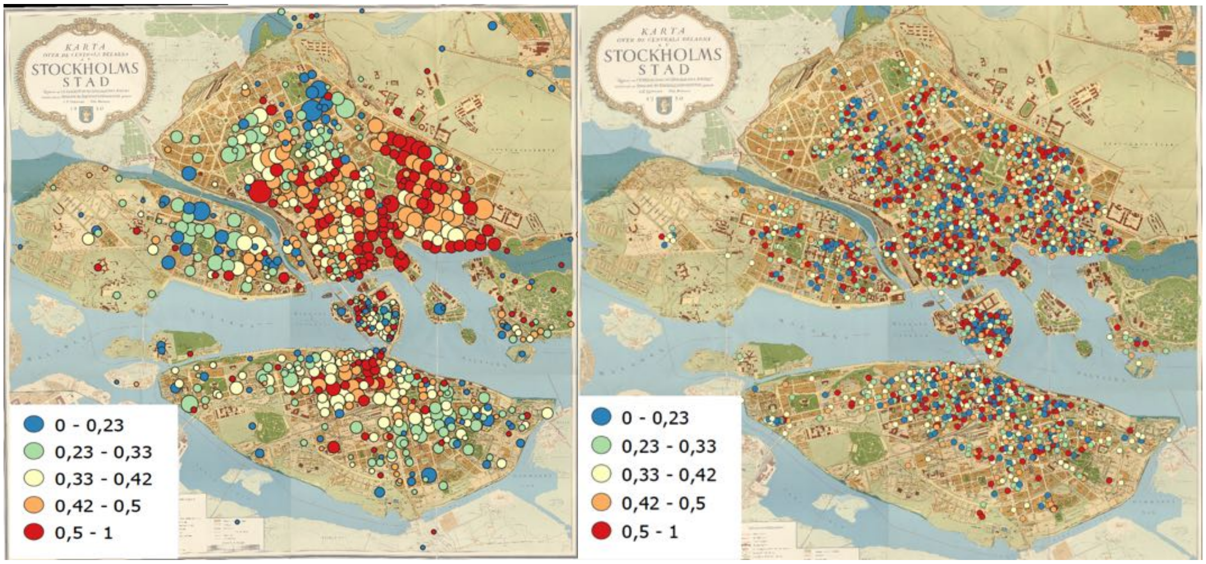
## 4.3 Cities and FLFP: The role of social interactions

Our results in the previous sections show that big cities such as Stockholm provided substantial economic and social opportunities for women in the early 20th century. While the earlier structural shift toward services goes some way in accounting for these differences, big cities may also have offered several other advantages. A particular advantage is that big cities may facilitate social interactions. The role of such social interactions may be particularly important for migrants if they facilitate the diffusion of information regarding job opportunities or social norms between working women (Fogli and Veldkamp, 2011). We next show that such social interactions promoted female migrants' entry into the formal labor market in Stockholm.

**Empirical strategy.** Our analysis combines the considerable spatial variation in FLFP in early-$20^{th}$ century Stockholm with the temporal and spatial granularity of our data (see Figure 7). For each migrant, we observe the building (the median building has about 70 adult inhabitants) and city block or "neighborhood" (about 500 adults) that she resides in during her first year after arriving in Stockholm. To measure potential social interactions, we calculate the FLFP rate among women (excluding individuals from the own household) in a female migrant's building of residence during her first year in Stockholm. We estimate OLS regressions of the following form:

$$Y_{ibnt} = \gamma_{nt} + \beta FLFP_{bnt} + \mathbf{X}'_{ibnt}\delta + \varepsilon_{ibnt} \quad (5)$$

[43] An alternative explanation is that female migrants were less likely to have children (see above), which in turn may have reduced the risk to to die in child birth, or related complications. Yet, Appendix Figure A.5 shows that the mortality differential for female migrants are not apparently related to women's child-bearing age.




(A) BLOCK-LEVEL MEANS

(B) BUILDING-LEVEL MEANS

FIGURE 7: SPATIAL VARIATION IN FLFP

*Notes:* This figure displays the FLFP rate across Stockholm in 1900. Panel A displays mean FLFP by block, with larger circles representing blocks with more residents. Panel B displays mean FLFP by building as deviations from the block mean.

where $Y_{ibnt}$ is an outcome for a female migrant $i$, whose first building of residence is $b$, which is located in block $n$. $FLFP_{bnt}$ is the share of female neighbors outside of the own household who have reported occupations in the year of first arrival in Stockholm, $t$. The vector $\mathbf{X}_{ibnt}$ adds the full set of individual and household-level controls described above. Throughout, we restrict our attention to female migrants that were aged 15 or above at the time of arrival.

Interpreting $\beta$ as the causal effect of social interactions requires that more industrious migrants did not select into buildings with higher FLFP. While this assumption may seem bold, the key idea of our identification strategy is to leverage variation in FLFP rates within small residential areas where housing markets are extremely thin (Bayer et al., 2008; Tan, 2022).[44] That is, while migrants may select into a particular neighborhood with higher FLFP rates, the scope for selection into specific buildings *within* narrowly defined neighborhoods is arguably much smaller. Thus, we include a full set of fixed effects for arrival year-by-block ($\gamma_{nt}$) in Equation 5, so that the identifying variation comes solely from variation in FLFP rates across buildings within the same block among

[44]Stockholm's housing market during industrialization was no exception, with large numbers of residents sharing small flats being typical (Thörn, 1997). In 1910, real rents were at a historical high, being 40 percent higher than in 1880 (Blöndal, 2015). Apart from housing shortages making sorting harder, for female migrants to sort into buildings they would also need to accurately be able to observe differences in FLFP. This appears unlikely given that we focus on migrants' residential decisions made during their arrival year when they likely have a limited knowledge about building and neighborhood heterogeneity.

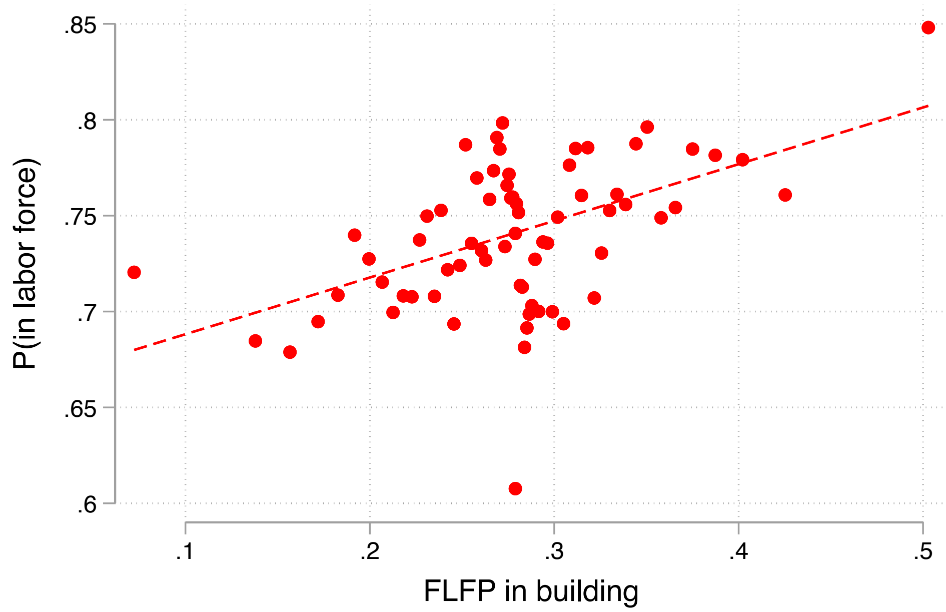



FIGURE 8: SOCIAL INTERACTIONS AND FLFP AMONG MIGRANTS.

*Notes:* This figure presents a binned scatterplot of the probability that a female migrant is working during her first year in Stockholm and the FLFP rate in the building that she resides in (excluding the own household) when arriving to Stockholm. We include a full set of arrival year-by-block fixed effects and individual (fixed effects for age, birth order, and being the eldest sister) and household controls (based on the 1880 census: i) a full set of origin county fixed effects, father's percentile income rank, and family size; and ii) a set of dummies capturing: the mother's LFP, the father's major (1-digit) HISCO group, the presence of servants in the household and whether a household is a married/cohabitating couple with children, single-parent family, extended family (relatives only), or composite (family and non-relatives) as well as whether the family is multigenerational). We also show a best fit line estimated on the underlying data.

migrants arriving in the same year.

**Tests for selection.** Figure 7 visualizes the spatial variation in FLFP within Stockholm in 1900 that captures the identifying variation we rely on. Panel A displays the variation in FLFP across neighborhoods, which reveals substantial sorting of working women across neighborhoods. Panel B instead displays how the FLFP of each building deviates from the level observed in the broader neighborhood. Notably, the within-neighborhood variation in FLFP is much more dispersed and does not indicate a systematic pattern.

We next show that there is no association between a wide range of observable migrant characteristics and the FLFP rate in the building a migrant resides in upon arriving in Stockholm, once we control for arrival year-by-neighborhood fixed effects. To formally test for sorting across and within neighborhoods, we examine the association between FLFP rates in a building and predetermined migrant characteristics (e.g., age or family background) among all migrants in our baseline sample, as well as a richer set of pre-migration outcomes (e.g, FLFP and marriage) for the subsample of individuals that are linked to the 1900 census. Appendix Figure A.6A shows that differences in pre-determined migrant characteristics such as family background attenuate once we condition on arrival year-by-block fixed effects. Similarly, Appendix Figure A.6B shows that younger women, those that came from more advantaged backgrounds, and those who worked prior to migration (i.e., in 1900) reside in buildings in Stockholm with higher FLFP rates on average. Conversely, women that were married or had children prior to moving to Stockholm ended up in buildings with significantly lower FLFP rates. Thus, there is clear evidence of selection into different neighborhoods. However, after conditioning on year of arrival-by-neighborhood fixed effects, the associations between observable migrant characteristics and the building-level FLFP rates sharply attenuate.[45] Thus, the scope for sorting within blocks is seemingly much more limited based on observable pre-migration characteristics.

**Results.** Figure 8 presents a binned scatterplot of the probability that a female migrant is in the labor force based on the FLFP rate in her building of residence in the year of arrival in Stockholm. Because we include year of arrival-by-neighborhood fixed effects and the full set of individual and household controls, the positive relationship indicates that a female migrant residing in a building with higher FLFP is more likely to be working compared to another female migrant arriving in the same year and residing in the same neighborhood but in a building with lower FLFP.

To substantiate these results, Table 4, panel A, presents estimates from Equation 5. Column 1 shows that a female migrant residing in a building with a 1 SD higher FLFP rate was about 3.5 percentage points more likely to work. Notably, the estimate is very similar when including the full set of individual and household controls in column 2, which again indicates limited selection within arrival year-by-block cells. Column 3 shows that the employment effects are gender specific: the LFP rate among males in a building is not correlated with female migrants' employment.

A key concern is that of reverse causality: rather than social interactions facilitating the employment of women, more industrious women may have selected into buildings with high FLFP.[46] Table 4, panel B, restricts the sample to individuals that we match to the 1900 census that allows us to directly control for whether a migrant was already working prior to arriving in Stockholm. Column 4 presents an analogous estimate to column 1 for this subsample showing that a 1 SD

[45] The associations with age, FLFP, and father's income score is reduced entirely, while the associations with marriage and fertility rates are sharply reduced in magnitude. We document below that our main results are robust to limiting our sample to female migrants that were unmarried prior to moving to Stockholm.

[46] A related concern is that working women may have been provided with accommodation through their employer. We explore this issue in Appendix Table A.12 where we exclude individuals who live in the same household as their employer showing that the results are stable to dropping this subset. Moreover, we also present estimates in Appendix Table A.12 showing that the effect of the FLFP rate in a building on a migrant's labor-force participation is negative or close to zero among those women who are employed by their household head, which suggests that this subset is not driving our main results.

| | In labor force (=1) | | | | | | |
|---|---|---|---|---|---|---|---|
| | A. All migrants | | | B. Migrants linked to 1900 census | | | |
| | | | | All | All | Working, 1900 | Not working, 1900 |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| FLFP in building | 0.035*** | 0.039*** | 0.041*** | 0.062*** | 0.056*** | 0.037 | 0.079** |
| | (0.008) | (0.007) | (0.010) | (0.017) | (0.017) | (0.028) | (0.037) |
| Male LFP in building | | | 0.003 | | | | |
| | | | (0.009) | | | | |
| Year of arrival-by-block FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Individual controls | No | Yes | Yes | Yes | Yes | Yes | Yes |
| Household controls | No | Yes | Yes | Yes | Yes | Yes | Yes |
| Pre-migration controls | No | No | No | No | Yes | Yes | Yes |
| Observations | 10862 | 10862 | 10862 | 2525 | 2525 | 946 | 854 |
| Mean dep. var. | 0.759 | 0.759 | 0.759 | 0.663 | 0.663 | 0.846 | 0.470 |

*Notes:* Individual-level OLS regressions. *FLFP in building* measures the FLFP rate among women (excluding the own household) in the building that a migrant resides in at arrival in Stockholm, which is standardized to have a mean of zero and a standard deviation of one. *Individual controls* include fixed effects for age and birth order, and an indicator for eldest sister. *Household controls* include a set of controls based on the 1880 census: i) a full set of origin county fixed effects, father's percentile income rank, and family size; and ii) a set of dummies capturing: the mother's LFP, the father's major (1-digit) HISCO group, the presence of servants in the household and whether a household is a married/cohabitating couple with children, single-parent family, extended family (relatives only), or composite (family and non-relatives) as well as whether the family is multigenerational. *Pre-migration controls* includes a set of dummies based on the 1900 census capturing whether an individual was part of the labor force, married, had any child, resided in an urban area, and had moved away from her county of birth. Standard errors clustered at year of arrival-by-neighborhood level. *** - $p < 0.01$, ** - $p < 0.05$, * - $p < 0.1$.

TABLE 4: SOCIAL INTERACTIONS AND FLFP AMONG MIGRANTS.

increase in FLFP in a building is associated with a 6.2 percentage point increase in labor-force participation for a female migrant.[47] Column 5 adds the full set of pre-migration controls characteristics — including labor force participation, marriage, and fertility — that has a limited effect on the estimated magnitude. To further examine the potential role of selection of more industrious female migrants into buildings with higher FLFP rates, we compare the impact of social interactions on female migrants that were (not) working prior to arriving in Stockholm. If our results mainly reflect a selection of more industrious women into buildings with higher FLFP, we would expect to see stronger results among the subset of women working prior to migration. In contrast, the estimates reported in the final two columns of Table 4 show that the results are stronger among women that were *not* working prior to arriving in Stockholm, which appears inconsistent with the argument that the effects are solely driven by a selection of more industrious migrants.

We present additional results in the Appendix. First, we show that the short-run effects on

[47]The difference in magnitude compare to the baseline estimate in column 1 mainly reflects the fact that the subsample of migrants in column 4 by definition arrived later (i.e., after 1900); the corresponding estimate (standard error) for the full migrant sample among the subset of migrants arriving after 1900 is 0.061 (0.014), which is very close to that in the linked subsample.

FLFP we find persist at least over the first five years that a female migrant resides in Stockholm (Appendix Table A.10). Second, to further support the interpretation that the effects reflect social interactions we show that our main results are mainly driven by female migrants that were more likely to lack pre-existing social networks (proxied by their place of birth and previous migrant flows) in Stockholm (Appendix Table A.11). Third, while migrants that were already married or had a child prior to moving to Stockholm were less likely to reside in a building with high FLFP, we show that our results are similar when restricting the sample to female migrants that were single and childless prior to moving to Stockholm (Appendix Table A.9). Thus, a potential selection based on marriage or fertility is unlikely to drive our results. While our findings sorting based on observables is unlikely to drive our results, sorting may have taken place based on unobservables. Reassuringly, we find that our controlled estimate in column 5 of Table 4 (0.056) is close in magnitude to the estimate corrected for potential selection on unobservables (0.054) obtained using the Oster (2019) method. In sum, these results suggest that the higher employment observed among female migrants that ended up in buildings with higher FLFP rates in Stockholm is not mainly driven by a sorting of migrants.

**Potential mechanisms.** A number of mechanisms may account for the positive association between social interactions and the improved employment outcomes among female migrants. To shed light on one potential mechanism, we examine whether female migrants are particularly likely to take up employment in the same occupations as their female neighbors. To do this, we create individual-occupation cells (at the 2-digit HISCO level, which corresponds to 74 unique occupations) where the outcome is an indicator taking the value 1 if a female migrant $i$ works in a particular occupation $o$, while the independent variables capture the building-level FLFP rate within the same occupation and all other occupations, respectively.

Table 5, column 1, shows that a female migrant that ends up in a building with a 1 SD higher FLFP rate in a particular occupation is 0.4 percentage points more likely to work in that occupation. Considering that the mean of the outcome is 1.0 percentage points this is a large magnitude. Column 2 shows that this association remains stable in magnitude and precision when controlling for the FLFP rate in all *other* occupations, which itself has a small negative impact on working in occupation $o$. In column 3, we show that female migrants' occupational choice is more tightly associated with the occupation of working women than working men in the same building. Again, we find similar results in columns 4 and 5 that presents estimates for the subset of female migrants linked to the 1900 census, which allows us to control for a wide range of pre-migration characteristics.

Together, these results provide suggestive evidence of the role of social interactions in shaping female migrants labor-market outcomes in Stockholm. While we cannot disentangle all underlying

| | Outcome: P(Working in occupation) | | | | |
|---|---|---|---|---|---|
| | A. All migrants | | | B. Linked to 1900 census | |
| | (1) | (2) | (3) | (4) | (5) |
| FLFP in building (same occ.) | 0.004*** | 0.004*** | 0.004*** | 0.003*** | 0.003*** |
| | (0.000) | (0.000) | (0.000) | (0.001) | (0.001) |
| FLFP in building (other occ.) | | -0.001*** | -0.001*** | -0.001* | -0.001* |
| | | (0.000) | (0.000) | (0.001) | (0.001) |
| Male LFP in building (same occ.) | | | 0.001*** | 0.001*** | 0.001*** |
| | | | (0.000) | (0.000) | (0.000) |
| Male LFP in building (other occ.) | | | 0.000 | 0.001 | 0.001 |
| | | | (0.000) | (0.001) | (0.001) |
| Year of arrival-by-block FE | Yes | Yes | Yes | Yes | Yes |
| Occupation FE | Yes | Yes | Yes | Yes | Yes |
| Individual controls | Yes | Yes | Yes | Yes | Yes |
| Household controls | Yes | Yes | Yes | Yes | Yes |
| Pre-migration controls | No | No | No | No | Yes |
| Observations | 1186812 | 1186812 | 1186146 | 353350 | 353350 |
| Mean dep. var. | 0.010 | 0.010 | 0.010 | 0.009 | 0.009 |

*Notes:* Individual-by-occupation-level OLS regressions. *FLFP in building* measures the FLFP rate among women (excluding the own household) in the building that a migrant resides in at arrival in Stockholm, which is standardized to have a mean of zero and a standard deviation of one. *Individual controls* include fixed effects for age and birth order, and an indicator for eldest sister (all measured in 1880). *Household controls* include a set of controls based on the 1880 census: i) a full set of origin county fixed effects, father's percentile income rank, and family size; and ii) a set of dummies capturing: the mother's LFP, the father's major (1-digit) HISCO group, the presence of servants in the household and whether a household is a married/cohabitating couple with children, single-parent family, extended family (relatives only), or composite (family and non-relatives) as well as whether the family is multigenerational. *Pre-migration controls* includes a set of dummies based on the 1900 census capturing whether an individual was part of the labor force, married, had any child, resided in an urban area, and had moved away from her county of birth. Standard errors clustered at year of arrival-by-neighborhood. *** - $p < 0.01$, ** - $p < 0.05$, * - $p < 0.1$.

TABLE 5: SOCIAL INTERACTIONS WITHIN/ACROSS OCCUPATIONS AND FLFP AMONG MIGRANTS.

mechanisms, the fact that the effects are occupation-specific point to the role of job referrals or role-model effects, rather than for example the diffusion of social norms regarding female work.

# 5 Conclusions

Our paper studies the central role of big cities in accounting for the historical advances in economic and social opportunities of women. Using historical census data, we document significant spatial variation in FLFP rates and structural change within both European countries and the United States. Already by the early 20th century, women had made significant strides in large cities where they were substantially more likely to be employed in the formal labor market.

Our main analysis focuses on the case of Stockholm analyzing the economic and social outcomes of migrants to Sweden's rapidly growing capital. Moving to the city had radical impacts on women's economic and social outcomes: migrants were substantially more likely to enter the formal labor market and to remain single and childless. We find little evidence that these gains reflect a selection of more ambitious or independent women, which suggests that the urban environment had large effects on the economic and social outcomes of women.

A key explanation for the higher FLFP rates in large cities is the fact that Stockholm — similar to other American and European cities — had experienced a precocious structural shift towards the service sector already by the early 20th century. The supply of service jobs, which were deemed "respectable" for women, potentially facilitated the entry of women into paid work. Indeed, the majority of female migrants to Stockholm transitioned into service jobs, which often meant significant increases in income and the absence of negative health effects in sharp contrast to that of male industrial workers in the city. Another advantage of large cities is that they facilitated social interactions between working women, which further promoted their transition into formal employment. Female migrants that upon arrival in Stockholm were exposed to working women were themselves more likely to transition into employment. These effects appear to be driven by job referrals, as evident from the fact that female migrants were more likely to transition into the same occupations as other working women in their proximity.

Our findings more broadly provide historical evidence of significant spatial variation in women's economic and social advances and the key role played by big cities in facilitating the rise of working women. Notably, the concentration of a growing number of working women in large cities likely provided fertile ground for the emergence of the women's rights movements, which fundamentally shaped economic, social, and political outcomes over the next 100 years. We call for future work to examine this further.

# References

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

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

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

**Acemoglu, Daron, David H Autor, and David Lyle**, "Women, war, and wages: The effect of female labor supply on the wage structure at midcentury," *Journal of political Economy*, 2004, *112* (3), 497–551.

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

**Bayer, Patrick, Stephen L Ross, and Giorgio Topa**, "Place of work and place of residence: Informal hiring networks and labor market outcomes," *Journal of Political Economy*, 2008, *116* (6), 1150–1196.

**Bell, Alex, Raj Chetty, Xavier Jaravel, Neviana Petkova, and John van Reenen**, "Who Becomes an Inventor in America? The Importance of Exposure to Innovation," *The Quarterly Journal of Economics*, 2019, *134* (2), 647–713.

**Blöndal, Sölvi**, "Speculative bubbles in the Stockholm housing market 1875-1935," *Stockholm Papers in Economic History*, 2015, *16.*

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

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

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

**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.

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

**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.

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

**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.

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

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

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

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

**Goldin, Claudia**, *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*, University of Chicago Press,

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

— **and Claudia Olivetti**, "Shocking labor supply: A reassessment of the role of World War II on women's labor supply," *American Economic Review*, 2013, *103* (3), 257–62.

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

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

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

**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.

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

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

**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.

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

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

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

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

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

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

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

**Tan, Hui Ren**, "How Widespread Are Social Network Effects? Evidence from the Early Twentieth-Century United States," *Journal of Labor Economics*, 2022, *40* (1), 187–237.

**Thörn, Kerstin**, "En bostad för hemmet: idéhistoriska studier i bostadsfrågan 1889-1929." PhD dissertation, Umeå universitet 1997.

**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.

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

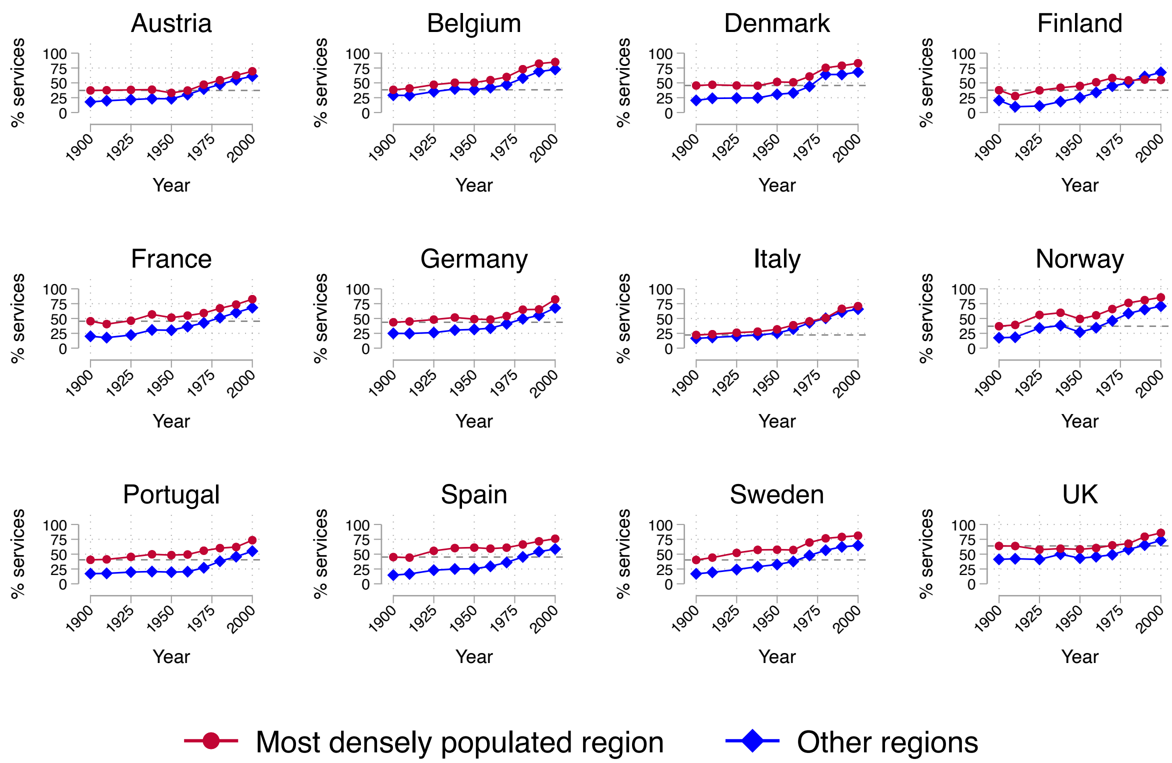
**Wisselgren, Maria J., Sören Edvinsson, Mats Berggren, and Maria Larsson**, "Testing Methods of Record Linkage on Swedish Censuses," *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 2014, *47* (3), 138–151.

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

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

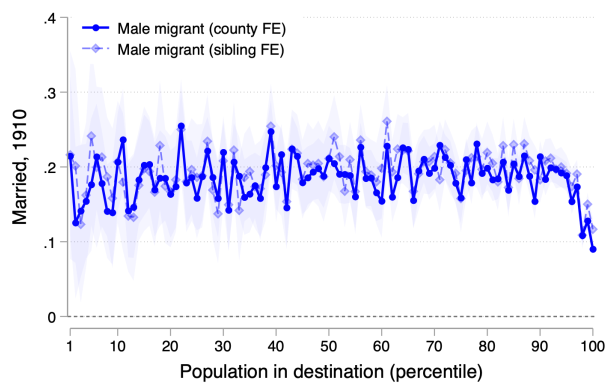
# Appendix

## A Robustness and Additional Material

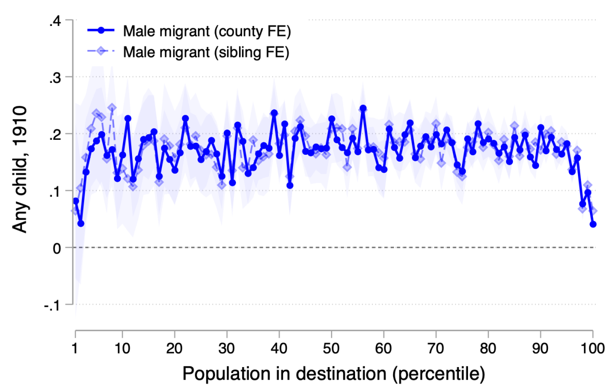



*Notes:* This figure shows that the most densely populated region in all European countries had a higher share of employment in the service sector compared to other regions. Data on the employment share of services is drawn from Rosés and Wolf (2018). For each country, we report the employment share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

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



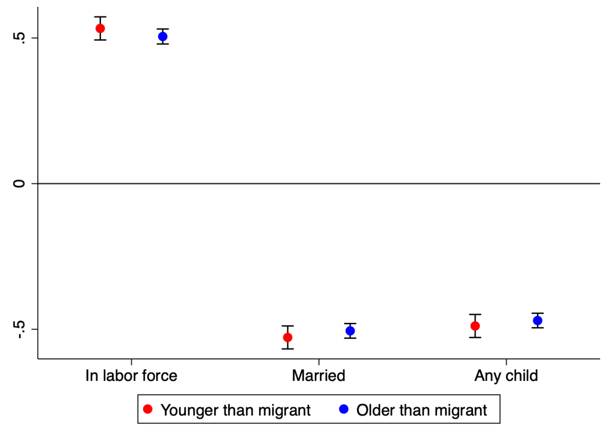

(A) MARRIAGE



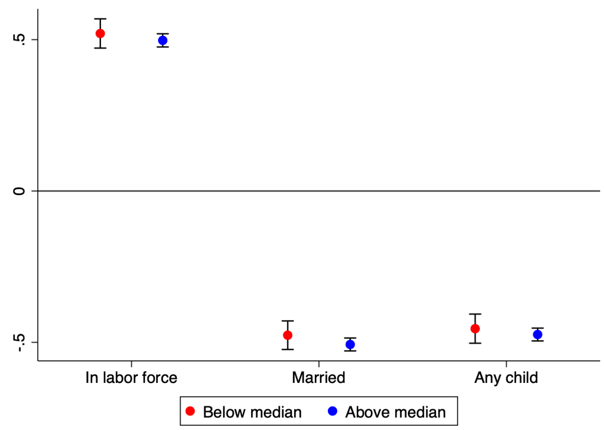

(B) ANY CHILD

FIGURE A.2: FAMILY FORMATION AMONG MALE MIGRANTS BY POPULATION IN DESTINATION

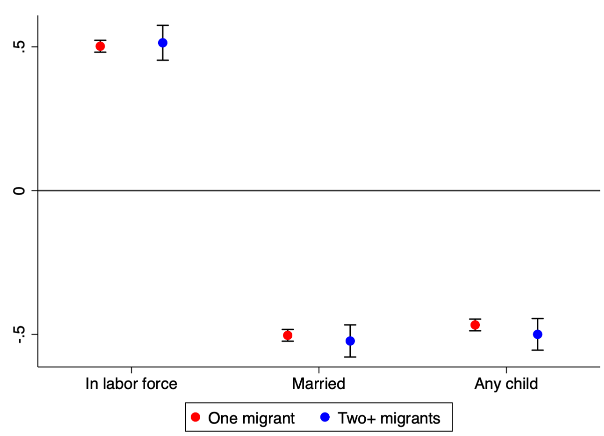
*Notes:* This figure displays OLS estimates of Equation (1) where the outcome is an indicator capturing whether an individual is married (A) or has any child (B) in 1910. The figure plots point estimates and 95 percent confidence intervals that capture the returns to migrating to different destinations ranked by their population size. Solid lines denote a specification using county fixed effects, while shaded lines correspond to estimates from a regression that includes family fixed effects. Standard errors are clustered at the family level.




(A) SIBLINGS YOUNGER/OLDER THAN MIGRANT




(B) BELOW/ABOVE MEDIAN FAMILY SIZE




(C) ONE/MORE MIGRANTS

FIGURE A.3: SPLITTING SAMPLES

*Notes:* This figure displays regression coefficients when splitting the main sample in various ways. All outcomes refer to 1910. All specifications include sibling fixed effects.

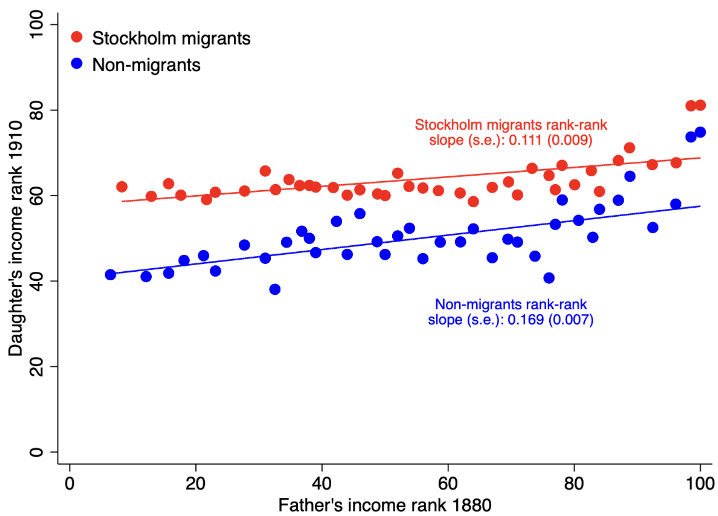



FIGURE A.4: INTERGENERATIONAL INCOME MOBILITY

*Notes:* This figure displays binned scatter plots of the association between father's income rank and two outcomes for children: their own income rank (panels A and B) and the probability of having a reported employment (panels C and D). Ranks are based on the average income for the occupation held by the father in 1880, son in 1910, or daughter's spouse in 1910. Red dots indicate the correlation among migrants to Stockholm. Blue dots indicate the correlation among non-migrants.

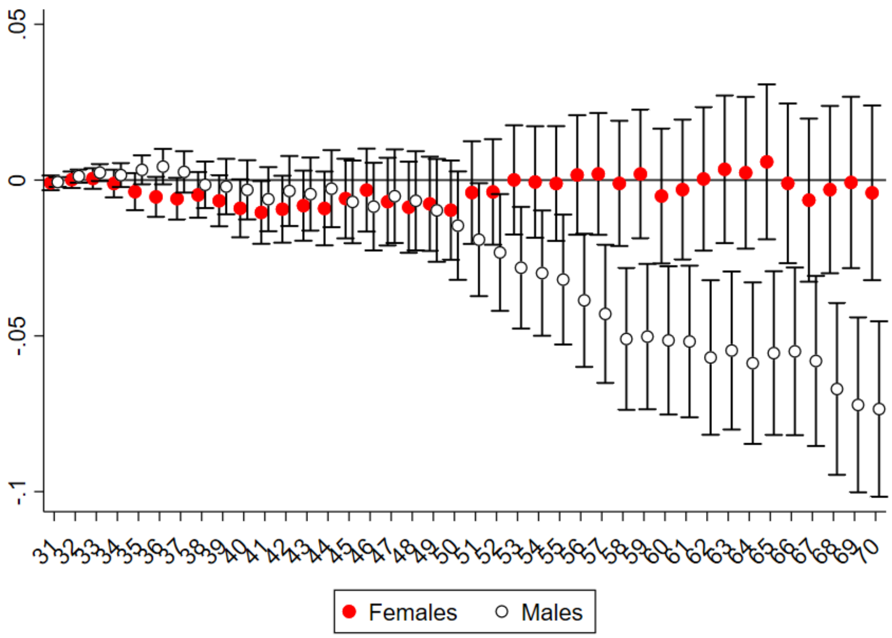
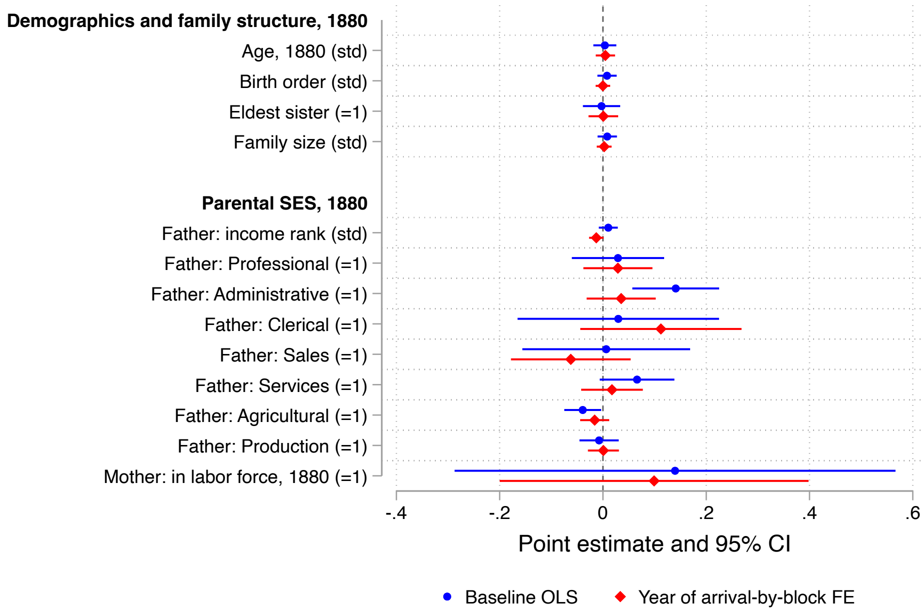
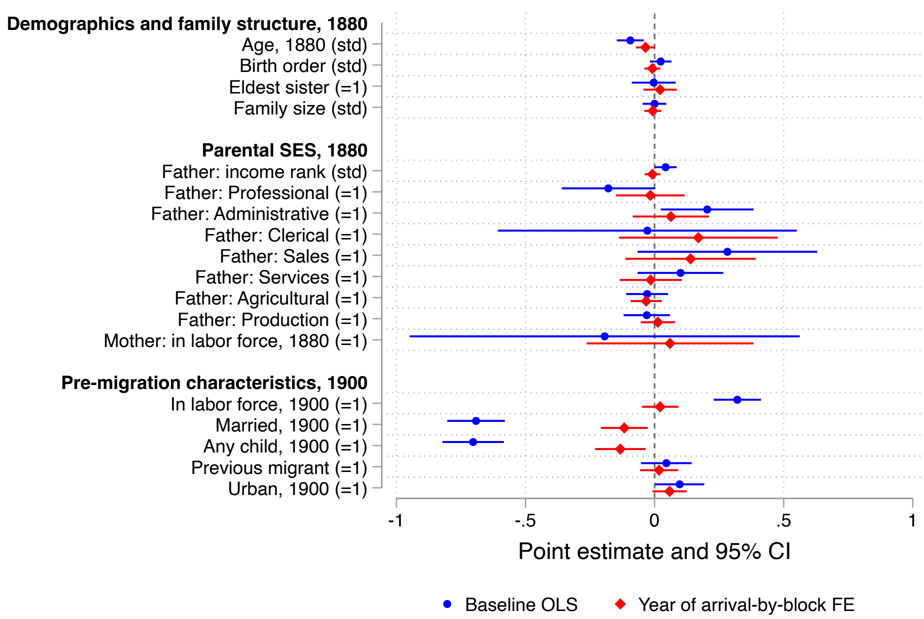



FIGURE A.5: MIGRATION AND MORTALITY

*Notes:* This figure displays regressions coefficients for the effect of moving to Stockholm by 1910 on the probability of surviving past a given age. Each coefficient is from a separate regression including individual and sibling fixed effects. Standard errors are clustered at the 1880 family level.



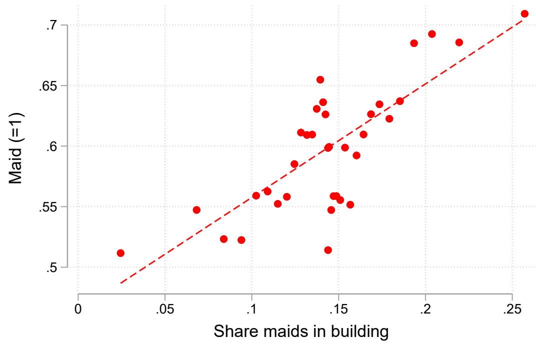

(A) FULL SAMPLE



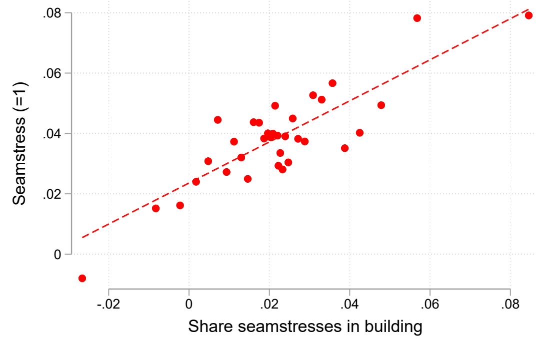

(B) SAMPLE LINKED TO 1900 CENSUS

*Notes:* This figure displays point estimates and 95% CIs from a set of individual-level regressions where the outcome is the FLFP rate in the building that a female migrant resides in during her first year in Stockholm on pre-migration characteristics observed in the 1900 and 1880 censuses. All non-binary variables are standardized. Panel A presents results for our baseline sample and the sample in panel B is restricted to female migrants who migrate after 1900 and are linked to the 1900 census. Estimates in blue correspond to bivariate associations and estimates in red include year of arrival-by-block fixed effects. Standard errors are clustered at the year of arrival-by-block level.

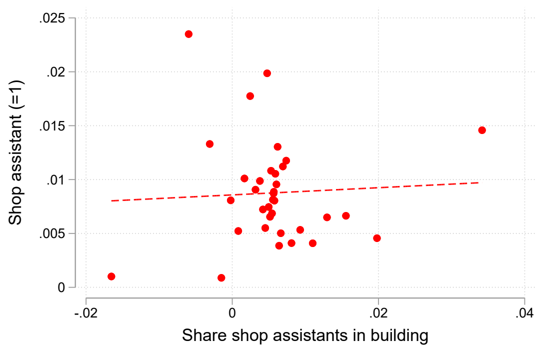
FIGURE A.6: TESTS FOR SELECTION




(A) MAIDS




(B) SEWERS




(C) SHOP ASSISTANTS AND DEMONSTRATORS


(D) LAUNDERERS AND PRESSERS

*Notes:* This figure presents a binned scatterplot of the probability that a female migrant is working in a specific occupation in her first year in Stockholm and the FLFP rate of that occupation in the building that she resides in (excluding the own household) when arriving to Stockholm. We control for the female employment share (excluding the own household) and include a full set of arrival year-by-block fixed effects and individual (fixed effects for age, birth order, and being the eldest sister) and household controls (based on the 1880 census: i) a full set of origin county fixed effects, father's percentile income rank, and family size; and ii) a set of dummies capturing: the mother's LFP, the father's major (1-digit) HISCO group, the presence of servants in the household and whether a household is a married/cohabitating couple with children, single-parent family, extended family (relatives only), or composite (family and non-relatives) as well as whether the family is multigenerational). We also show a best fit line estimated on the underlying data.

FIGURE A.7: SPECIFIC OCCUPATIONS IN BUILDING IN Y1 AND LATER OUTCOMES FOR FEMALE MIGRANTS.

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

| | (1) Stockholm mean | (2) Other urban area mean | (3) Rural area mean | (4) Stockholm (migrant) mean | (5) Stockholm (native) mean |
|---|---|---|---|---|---|
| FLFP | 0.51 | 0.37 | 0.18 | 0.51 | 0.50 |
| Any occupation | 0.54 | 0.39 | 0.20 | 0.55 | 0.52 |
| 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: SUMMARY STATISTICS OF MIGRANT SAMPLE

| *Panel A: Women* | Stockholm mean | Other urban mean | Rural parish mean |
|---|---|---|---|
| In labor force 1910 | 0.690 | 0.316 | 0.144 |
| Married 1910 | 0.238 | 0.635 | 0.797 |
| Any child 1910 | 0.247 | 0.629 | 0.785 |
| Age at death | 73.699 | 73.693 | 73.550 |
| Age 1880 | 6.448 | 6.841 | 7.158 |
| Birthorder | 3.224 | 3.131 | 3.070 |
| Eldest sister | 0.264 | 0.297 | 0.323 |
| Father's income score | 7.520 | 7.517 | 7.487 |
| Mother in labor force 1880 | 0.002 | 0.002 | 0.002 |
| Observations | 4466 | 17993 | 53437 |

| *Panel B: Men* | mean | mean | mean |
|---|---|---|---|
| In labor force 1910 | 0.977 | 0.974 | 0.967 |
| Married 1910 | 0.615 | 0.754 | 0.804 |
| Any child 1910 | 0.505 | 0.680 | 0.750 |
| Age at death | 68.747 | 70.824 | 72.503 |
| Age 1880 | 6.387 | 6.761 | 7.161 |
| Birthorder | 3.187 | 3.099 | 3.039 |
| Eldest brother | 0.274 | 0.298 | 0.314 |
| Father's income score | 7.523 | 7.533 | 7.478 |
| Mother in labor force 1880 | 0.003 | 0.003 | 0.002 |
| Observations | 5475 | 13351 | 44375 |

*Notes:* Summary statistics 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 our main migrant sisters sample. Panel B displays mean values for the male migrant brothers sample.

TABLE A.3: MIGRANT SELECTION

| Dependent variable: | Stockholm | Other urban | Rural parish |
|---|---|---|---|
| *Panel A: Within parishes* | (1) | (2) | (3) |
| Age | -0.001*** | -0.002*** | 0.001*** |
| | (0.000) | (0.000) | (0.000) |
| | (1) | (2) | (3) |
| Birthorder | 0.002*** | 0.002*** | -0.003*** |
| | (0.000) | (0.001) | (0.001) |
| | (1) | (2) | (3) |
| Eldest sister | -0.010*** | -0.013*** | 0.016*** |
| | (0.001) | (0.002) | (0.003) |
| *Panel B: Within families* | (1) | (2) | (3) |
| Age | -0.001*** | -0.002*** | 0.003*** |
| | (0.000) | (0.000) | (0.000) |
| | (1) | (2) | (3) |
| Birthorder | 0.004*** | 0.006*** | -0.009*** |
| | (0.000) | (0.001) | (0.001) |
| | (1) | (2) | (3) |
| Eldest sister | -0.009*** | -0.014*** | 0.018*** |
| | (0.001) | (0.002) | (0.003) |

*Notes:* OLS regressions. Each cell displays a separate regression with the dependent variable Stockholm migrant in column 1, Other urban area migrant in column 2, and Rural parish migrant in column 3. Panel A includes fixed effects for childhood parish. Panel B includes sibling fixed effects.

TABLE A.4: INCLUDING AGRICULTURAL OCCUPATIONS IN SKILL MEASURES

| Dependent variable: | High skilled | Low skilled | Unskilled |
|---|---|---|---|
| | (1) | (2) | (3) |
| Migrant: Stockholm (=1) | 0.069*** | 0.388*** | 0.004** |
| | (0.007) | (0.009) | (0.002) |
| Migrant: other urban area (=1) | 0.023*** | 0.092*** | -0.001 |
| | (0.004) | (0.004) | (0.001) |
| Individual controls | Yes | Yes | Yes |
| Sibling fixed effects | Yes | Yes | Yes |
| Observations | 79850 | 79850 | 79850 |
| Mean outcome | 0.085 | 0.134 | 0.006 |

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

TABLE A.5: COHABITATION AND ILLEGITIMATE CHILDREN

| Dependent variable: | Cohabit | Cohabit or married | Any illegitime child | Any child |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Migrant: Stockholm (=1) | 0.005* | -0.501*** | -0.066*** | -0.471*** |
| | (0.002) | (0.010) | (0.004) | (0.010) |
| Migrant: other urban area (=1) | -0.000 | -0.119*** | 0.004 | -0.102*** |
| | (0.001) | (0.006) | (0.003) | (0.006) |
| Individual controls | Yes | Yes | Yes | Yes |
| Sibling fixed effects | Yes | Yes | Yes | Yes |
| Observations | 79850 | 79850 | 79850 | 79850 |
| Mean outcome | 0.007 | 0.716 | 0.070 | 0.706 |

*Notes:* OLS regressions. All outcomes are measured in 1910. *Stockholm migrant* is an indicator taking value 1 if the individual lives in Stockholm city, and 0 if in a rural area. *Cohabit* is an indicator for households in which there are no married adults, and exactly one single adult female and one single adult male. *Any illegit. children* is an indicator for having any children born out of wedlock. *Individual controls* include fixed effects for birth year and birth order, and an indicator for eldest sister. Standard errors are given in parentheses and are clustered at the 1880 household level. *** - $p < 0.01$, ** - $p < 0.05$, * - $p < 0.1$.

TABLE A.6: MIGRANTS' HEALTH OUTCOMES

| Dependent variable: | Age at death | | | |
|---|---|---|---|---|
| | Women | | Men | |
| | (1) | (2) | (3) | (4) |
| Migrant: Stockholm (=1) | -0.103 | -0.487 | -2.053*** | -1.724*** |
| | (0.349) | (0.390) | (0.349) | (0.368) |
| Migrant: other urban area (=1) | 0.315 | 0.274 | -1.116*** | -0.781*** |
| | (0.195) | (0.199) | (0.233) | (0.248) |
| Individual controls | Yes | Yes | Yes | Yes |
| Sibling fixed effects | Yes | Yes | Yes | Yes |
| Occupation fixed effects | No | Yes | No | Yes |
| Observations | 61734 | 61607 | 49400 | 49253 |
| Mean outcome | 73.636 | 73.636 | 71.783 | 71.787 |

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

TABLE A.7: HETEROGENEITY BY AGE

| Dependent variable: | In labor force | Married | Any children |
|---|---|---|---|
| | (1) | (2) | (3) |
| Migrant: Stockholm (=1) | 0.551*** | -0.538*** | -0.496*** |
| | (0.016) | (0.015) | (0.015) |
| Stockholm migrant × Age | -0.008*** | 0.005*** | 0.004** |
| | (0.002) | (0.002) | (0.002) |
| Migrant: other urban area (=1) | 0.132*** | -0.119*** | -0.102*** |
| | (0.005) | (0.006) | (0.006) |
| Individual controls | Yes | Yes | Yes |
| Sibling fixed effects | Yes | Yes | Yes |
| Observations | 79850 | 79850 | 79850 |
| Mean outcome | 0.227 | 0.709 | 0.706 |

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

TABLE A.8: EFFECTS OF TEMPORARY MOVING TO STOCKHOLM

| Dependent variable: | In labor force | Married | Any children |
|---|---|---|---|
| | (1) | (2) | (3) |
| Migrant: Ever in Stockholm | 0.065*** | -0.060*** | -0.118*** |
| | (0.014) | (0.015) | (0.015) |
| Migrant: In Stockholm 1910 | 0.440*** | -0.453*** | -0.366*** |
| | (0.016) | (0.016) | (0.017) |
| Migrant: other urban area (=1) | 0.132*** | -0.119*** | -0.102*** |
| | (0.005) | (0.006) | (0.006) |
| Individual controls | Yes | Yes | Yes |
| Sibling fixed effects | Yes | Yes | Yes |
| Observations | 79850 | 79850 | 79850 |
| Mean outcome | 0.227 | 0.709 | 0.706 |

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

| | Outcome: in labor force (=1) | | |
|---|---|---|---|
| | All | Not married in 1900 | No child in 1900 |
| | (1) | (2) | (3) |
| FLFP in building | 0.056*** | 0.049*** | 0.050*** |
| | (0.017) | (0.019) | (0.018) |
| Year of arrival-by-block FE | Yes | Yes | Yes |
| Individual controls | Yes | Yes | Yes |
| Household controls | Yes | Yes | Yes |
| Pre-migration controls | Yes | Yes | Yes |
| Observations | 2525 | 2013 | 2091 |
| Mean dep. var. | 0.663 | 0.759 | 0.739 |

*Notes:* Individual-level OLS regressions. *FLFP in building* measures the FLFP rate among women (excluding the own household) in the building that a migrant resides in at arrival in Stockholm, which is standardized to have a mean of zero and a standard deviation of one. *Individual controls* include fixed effects for age, birth order, and being the eldest sister. *Household controls* include a set of controls based on the 1880 census: i) a full set of origin county fixed effects, father's percentile income rank, and family size; and ii) a set of dummies capturing: the mother's LFP, the father's major (1-digit) HISCO group, the presence of servants in the household and whether a household is a married/cohabitating couple with children, single-parent family, extended family (relatives only), or composite (family and non-relatives) as well as whether the family is multigenerational. *Pre-migration controls* includes a set of dummies based on the 1900 census capturing whether an individual was part of the labor force, married, had any child, resided in an urban area, and had moved away from her county of birth. Standard errors clustered at year of arrival-by-neighborhood. *** - $p < 0.01$, ** - $p < 0.05$, * - $p < 0.1$.

TABLE A.9: SOCIAL INTERACTIONS AND FLFP AMONG MIGRANTS: FOCUSING ON SINGLE AND CHILDLESS MIGRANTS.

| | Outcome: In labor force in ... year in Stockholm (=1) | | | | |
|---|---|---|---|---|---|
| | First | Second | Third | Fourth | Fifth |
| | (1) | (2) | (3) | (4) | (5) |
| FLFP in building | 0.039*** | 0.044*** | 0.035*** | 0.067*** | 0.078*** |
| | (0.007) | (0.009) | (0.011) | (0.013) | (0.020) |
| Individual controls | Yes | Yes | Yes | Yes | Yes |
| Household controls | Yes | Yes | Yes | Yes | Yes |
| Observations | 10862 | 8652 | 6845 | 5187 | 2648 |
| Mean dep. var. | 0.759 | 0.735 | 0.707 | 0.668 | 0.608 |

*Notes:* Individual-level OLS regressions. *FLFP in building* measures the FLFP rate among women (excluding the own household) in the building that a migrant resides in at arrival in Stockholm, which is standardized to have a mean of zero and a standard deviation of one. *Individual controls* include fixed effects for age, birth order, and being the eldest sister. *Household controls* include a set of controls based on the 1880 census: i) a full set of origin county fixed effects, father's percentile income rank, and family size; and ii) a set of dummies capturing: the mother's LFP, the father's major (1-digit) HISCO group, the presence of servants in the household and whether a household is a married/cohabitating couple with children, single-parent family, extended family (relatives only), or composite (family and non-relatives) as well as whether the family is multigenerational. Standard errors clustered at year of arrival-by-neighborhood. *** - $p < 0.01$, ** - $p < 0.05$, * - $p < 0.1$.

TABLE A.10: SOCIAL INTERACTIONS AND FLFP AMONG MIGRANTS: MEDIUM-TERM EFFECTS.

| | Childhood county (1880) | | Past migration to Stockholm | |
|---|---|---|---|---|
| | Stockholm | Other | High | Low |
| | (1) | (2) | (3) | (4) |
| FLFP in building | -0.023 | 0.043*** | 0.014 | 0.062*** |
| | (0.029) | (0.008) | (0.011) | (0.013) |
| Year of arrival-by-block FE | Yes | Yes | Yes | Yes |
| Individual controls | Yes | Yes | Yes | Yes |
| Household controls | Yes | Yes | Yes | Yes |
| Observations | 681 | 8702 | 4185 | 4193 |
| Mean dep. var. | 0.764 | 0.759 | 0.788 | 0.743 |

*Notes:* Individual-level OLS regressions. *FLFP in building* measures the FLFP rate among women (excluding the own household) in the building that a migrant resides in at arrival in Stockholm, which is standardized to have a mean of zero and a standard deviation of one. *Individual controls* include fixed effects for age, birth order, and being the eldest sister. *Household controls* include a set of controls based on the 1880 census: i) a full set of origin county fixed effects, father's percentile income rank, and family size; and ii) a set of dummies capturing: the mother's LFP, the father's major (1-digit) HISCO group, the presence of servants in the household and whether a household is a married/cohabitating couple with children, single-parent family, extended family (relatives only), or composite (family and non-relatives) as well as whether the family is multigenerational. Standard errors clustered at year of arrival-by-neighborhood. *** - $p < 0.01$, ** - $p < 0.05$, * - $p < 0.1$.

TABLE A.11: SOCIAL INTERACTIONS AND FLFP AMONG MIGRANTS: HETEROGENEITY BASED ON PRE-EXISTING NETWORKS IN STOCKHOLM.

| | Outcome: in labor force (=1) | | | | | |
|---|---|---|---|---|---|---|
| | A. All migrants | | | B. Linked to 1900 census | | |
| Employed by HH head: | Included (1) | Excluded (2) | Restricted to (3) | Included (4) | Excluded (5) | Restricted to (6) |
| FLFP in building | 0.039*** | 0.034*** | -0.009*** | 0.056*** | 0.033* | -0.002 |
| | (0.007) | (0.009) | (0.003) | (0.017) | (0.019) | (0.012) |
| Year of arrival-by-block FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Individual controls | Yes | Yes | Yes | Yes | Yes | Yes |
| Household controls | Yes | Yes | Yes | Yes | Yes | Yes |
| Pre-migration controls | No | No | No | Yes | Yes | Yes |
| Observations | 10862 | 7902 | 1946 | 2525 | 2034 | 266 |
| Mean dep. var. | 0.759 | 0.702 | 0.987 | 0.663 | 0.612 | 0.974 |

*Notes:* Individual-level OLS regressions. *FLFP in building* measures the FLFP rate among women (excluding the own household) in the building that a migrant resides in at arrival in Stockholm, which is standardized to have a mean of zero and a standard deviation of one. *Individual controls* include fixed effects for age, birth order, and being the eldest sister. *Household controls* include a set of controls based on the 1880 census: i) a full set of origin county fixed effects, father's percentile income rank, and family size; and ii) a set of dummies capturing: the mother's LFP, the father's major (1-digit) HISCO group, the presence of servants in the household and whether a household is a married/cohabitating couple with children, single-parent family, extended family (relatives only), or composite (family and non-relatives) as well as whether the family is multigenerational. *Pre-migration controls* includes a set of dummies based on the 1900 census capturing whether an individual was part of the labor force, married, had any child, resided in an urban area, and had moved away from her county of birth. Standard errors clustered at year of arrival-by-neighborhood. *** - $p < 0.01$, ** - $p < 0.05$, * - $p < 0.1$.

TABLE A.12: SOCIAL INTERACTIONS AND FLFP AMONG MIGRANTS: EXLUDING MIGRANTS WORKING FOR HOUSEHOLD HEAD.

| | Outcome: P(Working in same occupation) | | |
|---|---|---|---|
| | Professional, administrative, managerial | Clerical, sales, services | Production, operators, laborers |
| | (1) | (2) | (3) |
| FLFP in building (same occ.) | 0.004*** | 0.004*** | 0.003*** |
| | (0.001) | (0.001) | (0.001) |
| FLFP in building (other occ.) | -0.000* | -0.001*** | -0.000 |
| | (0.000) | (0.000) | (0.000) |
| Year of arrival-by-block FE | Yes | Yes | Yes |
| Occupation FE | Yes | Yes | Yes |
| Individual controls | Yes | Yes | Yes |
| Household controls | Yes | Yes | Yes |
| Observations | 320760 | 352836 | 449064 |
| Mean dep. var. | 0.002 | 0.029 | 0.002 |

*Notes:* Individual-by-occupation-level OLS regressions. *FLFP in building* measures the FLFP rate among women (excluding the own household) in the building that a migrant resides in at arrival in Stockholm, which is standardized to have a mean of zero and a standard deviation of one. *Individual controls* include fixed effects for age and birth order, and an indicator for eldest sister. *Household controls* include a set of controls based on the 1880 census: i) a full set of origin county fixed effects, father's percentile income rank, and family size; and ii) a set of dummies capturing: the mother's LFP, the father's major (1-digit) HISCO group, the presence of servants in the household and whether a household is a married/cohabitating couple with children, single-parent family, extended family (relatives only), or composite (family and non-relatives) as well as whether the family is multigenerational. Standard errors clustered at year of arrival-by-neighborhood. *** - $p < 0.01$, ** - $p < 0.05$, * - $p < 0.1$.

TABLE A.13: SOCIAL INTERACTIONS AND FLFP AMONG MIGRANTS: RESULTS BY SECTOR.

TABLE A.14: MIGRANTS' EMPLOYMENT USING ALTERNATIVE SAMPLES

| Dependent variable: | In labor force (=1) | | | |
|---|---|---|---|---|
| *Panel A: Extended sample* | (1) | (2) | (3) | (4) |
| Migrant: Stockholm (=1) | 0.548*** | 0.546*** | 0.555*** | 0.514*** |
| | (0.006) | (0.006) | (0.007) | (0.009) |
| Migrant: other urban area (=1) | 0.175*** | 0.174*** | 0.167*** | 0.135*** |
| | (0.003) | (0.003) | (0.004) | (0.004) |
| Observations | 156525 | 156525 | 140172 | 156525 |
| Mean outcome | 0.188 | 0.188 | 0.187 | 0.188 |
| *Panel B: Migrant sample* | (1) | (2) | (3) | (4) |
| Migrant: Stockholm (=1) | 0.536*** | 0.533*** | 0.547*** | 0.497*** |
| | (0.007) | (0.007) | (0.007) | (0.010) |
| Migrant: other urban area (=1) | 0.172*** | 0.170*** | 0.166*** | 0.132*** |
| | (0.004) | (0.004) | (0.004) | (0.005) |
| Observations | 79850 | 79850 | 71301 | 79850 |
| Mean outcome | 0.227 | 0.227 | 0.227 | 0.227 |
| *Panel C: Urban migrant sample* | (1) | (2) | (3) | (4) |
| Migrant: Stockholm (=1) | 0.353*** | 0.352*** | 0.375*** | 0.362*** |
| | (0.010) | (0.010) | (0.011) | (0.018) |
| Individual controls | No | Yes | Yes | Yes |
| Household controls | No | No | Yes | No |
| Sibling fixed effects | No | No | No | Yes |
| Observations | 14601 | 14601 | 13114 | 14601 |
| Mean outcome | 0.422 | 0.422 | 0.422 | 0.422 |

*Notes:* OLS regressions. All outcomes are measured in 1910. *Stockholm migrant* is an indicator taking value 1 for residents of Stockholm city and 0 for rural residents. *Individual controls* include fixed effects for birth year and birth order, and an indicator for eldest sister. *Household controls* include fixed effects for the following within the 1880 household: father's and mother's occupation, family size, number of families, generations, mothers, fathers, couples, servants, unrelated members, as well as an indicator for farming households. Standard errors, in parentheses, are clustered at the 1880 household level. *** - $p < 0.01$, ** - $p < 0.05$, * - $p < 0.1$.

TABLE A.15: MIGRANTS' MARRIAGE AND FERTILITY USING ALTERNATIVE SAMPLES

| Dependent variable: | Married (=1) | | | | Any child (=1) | | | |
|---|---|---|---|---|---|---|---|---|
| *Panel A: Extended sample* | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Migrant: Stockholm (=1) | -0.481*** | -0.475*** | -0.489*** | -0.441*** | -0.470*** | -0.462*** | -0.470*** | -0.422*** |
| | (0.006) | (0.006) | (0.006) | (0.009) | (0.006) | (0.006) | (0.006) | (0.009) |
| Migrant: other urban area (=1) | -0.082*** | -0.079*** | -0.072*** | -0.046*** | -0.090*** | -0.087*** | -0.076*** | -0.049*** |
| | (0.004) | (0.004) | (0.004) | (0.005) | (0.004) | (0.004) | (0.004) | (0.005) |
| Observations | 156525 | 156525 | 140172 | 156525 | 156525 | 156525 | 140172 | 156525 |
| Mean outcome | 0.686 | 0.686 | 0.685 | 0.686 | 0.690 | 0.690 | 0.687 | 0.690 |
| *Panel B: Migrant sample* | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Migrant: Stockholm (=1) | -0.545*** | -0.541*** | -0.551*** | -0.505*** | -0.524*** | -0.518*** | -0.524*** | -0.471*** |
| | (0.006) | (0.006) | (0.007) | (0.010) | (0.006) | (0.006) | (0.007) | (0.010) |
| Migrant: other urban area (=1) | -0.155*** | -0.154*** | -0.146*** | -0.119*** | -0.149*** | -0.146*** | -0.137*** | -0.102*** |
| | (0.004) | (0.004) | (0.004) | (0.006) | (0.004) | (0.004) | (0.004) | (0.006) |
| Observations | 79850 | 79850 | 71301 | 79850 | 79850 | 79850 | 71301 | 79850 |
| Mean outcome | 0.709 | 0.709 | 0.707 | 0.709 | 0.706 | 0.706 | 0.703 | 0.706 |
| *Panel C: Urban migrant sample* | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Migrant: Stockholm (=1) | -0.372*** | -0.371*** | -0.388*** | -0.388*** | -0.370*** | -0.368*** | -0.383*** | -0.363*** |
| | (0.009) | (0.009) | (0.010) | (0.018) | (0.009) | (0.009) | (0.010) | (0.018) |
| Individual controls | No | Yes | Yes | Yes | No | Yes | Yes | Yes |
| Household controls | No | No | Yes | No | No | No | Yes | No |
| Sibling fixed effects | No | No | No | Yes | No | No | No | Yes |
| Observations | 14601 | 14601 | 13114 | 14601 | 14601 | 14601 | 13114 | 14601 |
| Mean outcome | 0.512 | 0.512 | 0.511 | 0.512 | 0.519 | 0.519 | 0.516 | 0.519 |

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

TABLE A.16: 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) |
| Migrant: Stockholm (=1) | 0.205*** | 0.325*** | 0.480*** | 0.698*** | 1.071*** | 0.508*** |
| | (0.011) | (0.011) | (0.012) | (0.014) | (0.020) | (0.010) |
| Migrant: other urban area (=1) | -0.077*** | -0.048*** | -0.010 | 0.043*** | 0.134*** | 0.124*** |
| | (0.006) | (0.006) | (0.006) | (0.008) | (0.011) | (0.006) |
| Individual controls | Yes | Yes | Yes | Yes | Yes | Yes |
| Sibling fixed effects | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 68952 | 68952 | 68952 | 68952 | 68952 | 79850 |
| Mean outcome | 6.846 | 6.680 | 6.466 | 6.165 | 5.651 | 0.300 |

*Notes:* OLS regressions. All outcomes are measured in 1910. *Stockholm migrant* is an indicator taking value 1 for residents of Stockholm city and 0 for rural residents. *Individual controls* include fixed effects for birth year and birth order, and an indicator for eldest sister. Standard errors, in parentheses, are clustered at the 1880 household level. *** - $p < 0.01$, ** - $p < 0.05$, * - $p < 0.1$.

| | All mean | Linked mean |
|---|---|---|
| Age | 7.419 | 7.210 |
| Birthorder | 2.772 | 2.743 |
| Eldest sister (=1) | 0.475 | 0.481 |
| Father's age | 42.838 | 42.724 |
| Mother's age | 39.673 | 39.416 |
| Mother in labor force (=1) | 0.002 | 0.002 |
| Father in labor force (=1) | 0.915 | 0.929 |
| Father white collar (=1) | 0.136 | 0.139 |
| Family members in household | 6.285 | 6.196 |
| Any servants (=1) | 0.167 | 0.217 |
| Multigenerational family (=1) | 0.058 | 0.060 |
| HH type: Extended family (=1) | 0.050 | 0.047 |
| HH type: Composite (=1) | 0.198 | 0.250 |
| Observations | 301943 | 287456 |

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

TABLE A.17: SUMMARY STATISTICS FOR FULL POPULATION AND LINKED SAMPLE IN 1880 CENSUS

| Dependent variable: | In labor force | Married | Any children |
|---|---|---|---|
| | (1) | (2) | (3) |
| Migrant: Stockholm (=1) | 0.504*** | -0.515*** | -0.481*** |
| | (0.010) | (0.010) | (0.010) |
| Migrant: other urban area (=1) | 0.133*** | -0.120*** | -0.104*** |
| | (0.005) | (0.006) | (0.006) |
| Individual controls | Yes | Yes | Yes |
| Sibling fixed effects | Yes | Yes | Yes |
| Observations | 77233 | 77233 | 77233 |
| Mean outcome | 0.225 | 0.712 | 0.710 |

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

TABLE A.18: OLS REGRESSIONS: WEIGHTING REGRESSIONS WITH THE INVERSE PROBABILITY OF BEING LINKED ACROSS CENSUSES

## A.1 Accounting for between-sibling variation: controlling for pre-migration characteristics

Table A.19 presents estimates for the subset of individuals who move after 1900 where we control for outcomes in that year, when our sample individuals are 20–36 years old and thus in working age. Column 1 shows that the baseline employment effect of migration is 42.4 percentage points in this subsample. Column 2 then adds a control for the lagged outcome in 1900, prior to migration. As expected, the lagged outcome has a strong, positive association with reporting an occupation in 1910. However, the coefficient on migration remains highly statistically and economically significant at 35.9 percentage points. The decreased coefficient indicates that within families, migrants were indeed positively selected in terms of labor force participation. However, selection along this margin explains only a minor share of the migration effect, which remains sizeable and highly statistically significant.

TABLE A.19: CONTROLLING FOR PRE-MIGRATION OUTCOMES

| Dependent variable: | In labor force | | Married | | Any children | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Migrant after 1900: Stockholm (=1) | 0.547*** | 0.549*** | -0.458*** | -0.433*** | -0.479*** | -0.446*** |
| | (0.053) | (0.052) | (0.054) | (0.054) | (0.051) | (0.051) |
| Migrant after 1900: other urban area (=1) | 0.144*** | 0.139*** | -0.135*** | -0.112*** | -0.115*** | -0.093*** |
| | (0.027) | (0.027) | (0.028) | (0.027) | (0.029) | (0.028) |
| Lagged outcome | | 0.142*** | | 0.302*** | | 0.333*** |
| | | (0.028) | | (0.021) | | (0.022) |
| Individual controls | Yes | Yes | Yes | Yes | Yes | Yes |
| Sibling fixed effects | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 4692 | 4692 | 4692 | 4692 | 4692 | 4692 |
| Mean outcome | 0.233 | 0.233 | 0.677 | 0.677 | 0.647 | 0.647 |

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

## A.2 Accounting for between-sibling variation: a twin design

There may be additional individual level variation between siblings that can bias results and that are not captured by the inclusion of sibling fixed effects. These should be smallest between twins, who either have identical or near-identical genetics. Using data on birth year and birth month, we are able to identify 1539 sets of twins. Interestingly, Stockholm migration rates in the twin sample are similar to the full sample and close to 5% for both males and females. Table A.20 replicates our results using twin fixed effects. Due to a much reduced sample size, precision is lower for some, but not all, outcomes. Nevertheless, the results confirm our main results. All have the same estimated sign as for the full sample and effect sizes are larger for female than male migrants. 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.20: TWIN FIXED EFFECTS

| Dependent variable: | In labor force | Married | Any children |
|---|---|---|---|
| | (1) | (2) | (3) |
| Migrant: Stockholm (=1) | 0.781*** | -0.763*** | -0.623*** |
| | (0.118) | (0.117) | (0.129) |
| Migrant: other urban area (=1) | 0.084 | -0.036 | 0.006 |
| | (0.097) | (0.104) | (0.102) |
| Individual controls | Yes | Yes | Yes |
| Twin fixed effects | Yes | Yes | Yes |
| Observations | 381 | 381 | 381 |
| Mean outcome | 0.270 | 0.688 | 0.677 |

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