UNIVERSITY OF OXFORD

Oxford Economic and Social History Working Papers

Number 214, January 2025

Explaining gender differences in migrant sorting: Evidence from Canada-US migration

DAVID ESCAMILLA-GUERRERO, MIKO LEPISTÖ, CHRIS MINNS

Explaining gender differences in migrant sorting: Evidence from Canada-US migration

David Escamilla-Guerrero[*]
University of St Andrews and IZA

Miko Lepistö[†]
Paris School of Economics

Chris Minns[‡]
London School of Economics

December 4, 2024

Keywords: Migration, sorting, gender, Canada, United States
JEL Codes: J61, N31, N32

Acknowledgments: We thank Amera Dam-Elhana, Catherine Lee, and Shloka Sahoo for excellent research assistance. We have benefited from comments on the draft from Ian Keay, Kris Inwood, Diego Battiston, David Jaeger, Margaret Leighton, Edward Kosack, Zach Ward, seminar participants at Bonn, LSE and St. Andrews, and conference participants at the Scotland and Northern England Workshop in Applied Microeconomics, the Canadian Economics Association, the European Historical Economics Society, and the Economic History Association. Miko Lepistö thanks the support of a French government subsidy managed by the Agence Nationale de la Recherche under the framework of the Investissements d'Avenir program reference ANR-17-EURE-0001. Competing interests: the authors declare none.

[*]Department of Economics, University of St Andrews, The Scores, St Andrews KY16 9AZ. e-mail: david.escamilla-guerrero@st-andrews.ac.uk.

[†]Paris School of Economics, Université Paris 1 Panthéon-Sorbonne. 48 Boulevard Jourdan, 75014 Paris, France. e-mail: miko.lepisto@psemail.eu.

[‡]Department of Economic History, London School of Economics, Houghton Street, London WC2A 2AE. e-mail: c.minns@lse.ac.uk. Corresponding Author.

Explaining gender differences in migrant sorting: Evidence from Canada-US migration

Abstract

This paper uses newly digitized Canada-Vermont border crossing records from the early twentieth century to document substantial differences in how female and male migrants sorted across US destination counties by earnings potential. Income maximization largely explains sorting patterns among men. For single women, gender-based labor market constraints were important, with locations offering more work opportunities attracting women with higher earnings capacity. Among married women, destination choices were much less influenced by labor market characteristics. These findings reveal how labor market constraints based on gender and marriage influence the allocation of migrant talent across destinations.

1. Introduction

Understanding how migrants sort across destinations by skill has important implications for innovation, trade, and the fiscal costs of immigration in receiving societies. (Kerr et al., 2016; Hunt and Gauthier-Loiselle, 2010; Moser, Voena and Waldinger, 2014; Egger, Von Ehrlich and Nelson, 2012; Gould, 1994; Parsons and Vézina, 2018; Dustmann and Frattini, 2014; Storesletten, 2000). Although women comprise nearly half of all international migration flows (Artuç et al., 2015), research on destination sorting typically focuses on men or does not distinguish between women and men.[1] There are good reasons to ask whether female migrants exhibit a skill profile across destinations similar to their male counterparts. Maternal skill plays a major role in shaping the labor market outcomes of second-generation immigrants (Rosenzweig and Wolpin, 1994; Black, Devereux and Salvanes, 2005). Research has also linked female migrants' destination choices to local economic growth, as US regions with more children of immigrant mothers show higher per capita income (von Berlepsch, Rodríguez-Pose and Lee, 2019).

Recent scholarship has documented substantial differences in destination choices between female and male migrants in contemporary settings (Aksoy and Poutvaara, 2021; Bertoli, Fernández-Huertas Moraga and Ortega, 2013). Economic historians have also recognized the existence of "sex differentials" in the migration decisions made by men and women in the Age of Mass Migration (1850–1920) (Greenwood, 2008; Thomas, 1938), a period during which roughly 30 million migrants moved to the United States (Abramitzky and Boustan, 2017; Hatton and Williamson, 1998). But neither strand of research has examined how male and female migrants sort across destinations by skill, nor how this sorting relates to gender-specific labor market characteristics.[2]

This paper examines how female and male French Canadian migrants sorted across US counties by earnings potential between 1905 and 1925. This period provides an excellent environment to study migrant sorting, as entry to the United States was virtually unrestricted for Canadians even after the introduction of national origin quotas in 1921, which allows us to observe migrant destination choices in the absence of direct immigration policy effects. In addition, Canada-to-US migration is of significant historical importance. About 1.1 million Canadian immigrants resided in the United States in 1920, 51 percent of whom were female (Truesdell, 1943; Willcox, 1929). French Canadians were disproportionately represented in Canadian emigration and mostly settled in the US Northeast

[1] Cobb-Clark (1993); Docquier, Lowell and Marfouk (2009); Docquier et al. (2012) are examples of the limited literature that studies the migration decisions of women.

[2] Partial exceptions can be found in Tortorici and Fernández Sánchez (2023), who examine selection into migration of Portuguese migrants between 1850 and 1930, and Dribe, Eriksson and Helgertz (2023), who examine class selection by gender among Swedish migrants.

(Ramirez, 1986; Ramirez and Otis, 2001).[3] This region was attractive to French Canadian migrants due its proximity and wide array of manufacturing jobs that did not require previous experience in industry or the ability to speak good English.[4] However, economic prospects across Northeast counties varied substantially by gender due to the spatial distribution of female-intensive industries and the presence of social norms regarding the role of women in the workplace. Contemporary observers noted that migrants' destination choices were shaped by differences in local employment structures and the presence of French-Canadian (Catholic) parishes in New England (MacDonald, 1898).

We examine migrant sorting using a sample of newly digitized Canada-Vermont border crossing records. These records consist of immigration cards reporting individual data including physical stature (height) at time of crossing. Physical stature is a function of childhood conditions (nutrition, disease environment, and work assignments) that are highly correlated with human capital and earnings potential in adulthood (e.g., Borrescio-Higa, Bozzoli and Droller, 2019; Komlos and Baten, 2004; Schultz, 2002; Schneider and Ogasawara, 2018). Unlike occupation or earnings, height was measured for migrants irrespective of whether or not they were previously employed or intending to work in the United States. Height also has advantages over other measures of human capital such as literacy, given that about 95 percent of early 20th century French Canadian migrants were able to read and write.[5] We collapse the individual-level height data into Quebec region-US county pairs (migration corridors) and use full-count US census data to construct county-level labor market and socioeconomic characteristics for each destination. This allows us to evaluate which destination characteristics influenced sorting patterns among male and female migrants.

We find substantial gender differences in migrant sorting across Northeast destinations, with few counties attracting female and male migrants of an equivalent height profile. As in contemporary studies (e.g., Grogger and Hanson, 2011), income maximization largely explains sorting patterns of men. In contrast, we find that returns to skill were much less important in explaining sorting patterns among female migrants. Single women were sorted positively on employment probabilities, with high-earnings-potential females more likely to settle in locations where it was easier for women to find work. This implies that women faced a strong trade-off between matching their skills to local labor market demand and the possibility to work. When we estimate models that pool together married and single migrants, we find almost complete attenuation of sorting on labor market characteristics among

[3]French Canadians were a distinctive immigrant community, with important linguistic, cultural, and religious differences to the rest of English-speaking North America.

[4]Train fares from Montreal to Boston were about 6.50 in 1900, roughly equivalent to a few days of income for a laborer (Green, MacKinnon and Minns, 2005).

[5]Stature is correlated with other measures of economic status and earnings potential in our data. Figure B.5 shows that height is correlated with the amount of cash migrants declared carrying at entry to the United States.

women. This finding is in line with previous research showing that tied migration attenuates female migration responses to economic signals (Greenwood, 2008; Tortorici and Fernández Sánchez, 2023). For men there are few substantive differences between the two sets of results. These findings support previous research arguing that women tend to be tied migrants, but also underscores the importance of considering the choices of single women separately to understand the dynamics of female migration.

This paper contributes to the literature on migrant destination choice by demonstrating that female and male migrants sorted themselves distinctively across destinations by skill, with different responses to labor market conditions in those destinations. We show how gender-based labor market constraints and tied migration shape women's sorting across destinations. These constraints —few opportunities to work after marriage and limited options when single— are found in contemporary labor markets, for example major migrant destinations in the Gulf States (Bursztyn, González and Yanagizawa-Drott, 2020; ILO, 2023; World Bank, 2016). An implication of our findings is that labor market discrimination based on personal traits is likely to induce an inefficient allocation of migrant talent by creating a wedge between returns to skill and the skill profile of immigrants. This misallocation may have important implications for the performance of local economies (Lee, 2024).

Our results also provide new insights into the migration patterns of women during the Age of Mass Migration, a period when the United States experienced unparalleled population transfers (Hatton and Williamson, 1994, 1998). While previous studies have provided evidence on how source characteristics influenced emigration patterns, we are able to document differences in settlement patterns by skill, gender, and marital status, which may have had long-run implications for the integration of Franco Americans and other immigrant communities. (MacKinnon and Parent, 2012).

2. Historical Background

About 30 million immigrants moved to the United States during the Age of Mass Migration (1850-1920). Although most research addressing this period has focused on transatlantic flows, significant migrations also took place across the Canada-US border (Abramitzky and Boustan, 2017; Hatton and Williamson, 1998).[6] While some European migrants transited through Canada en route to the United States, the vast majority of border crossings consisted of native-born Canadians of either British or French ancestry. As a result over one million Canadians were resident in the United States in 1920.[7] Emigration rates were particularly high in Quebec, with approximately 20 percent of all French Canadians living in the

[6] Mexican immigration to the United States also increased gradually from the 1880s (Gratton and Merchant, 2015).
[7] At least until 1900, Canadian gross out-migration to the United States completely offset European gross immigration to Canada (McInnis, 1994).

US Northeast (Ramirez and Otis, 2001), where large French Canadian communities featured parochial French-language schools, French-language newspapers, and Franco-American credit unions. (Brault, 1986).

The conventional view among economic historians is that Canada-US immigration reflected differences in economic prospects between the two countries. Real wages for men in most occupations were 5 to 20 percent higher in the United States (Green, MacKinnon and Minns, 2002), and the vast majority of men were likely to find a job in any location. Comparisons of wage ratios between clerical and production workers suggest that skill premia were also larger in the United States from 1900 to 1920, particularly for women (Goldin and Katz, 1999; Inwood, MacKinnon and Minns, 2010).[8] However, skill premiums provide a partial picture of the labor market opportunities for migrant women in early 20th century North America. Unlike men, women faced significant barriers to employment in both countries. Job opportunities for women were somewhat greater in the United States in 1920, where adult labor force participation rates were on average 6 percentage points higher than in Canada (23.7 percent v. 17.7 percent), but less than ten percent of married women were employed in 1920 (Goldin, 1990).[9] Within the US Northeast, employment opportunities for women varied significantly, with county-level female labor force participation rates ranging from 9 to 55 percent.[10] This variation partly reflected the distribution of female-accessible industries within the region. Across destination counties, the proportion of the female labor force employed in female-intensive sectors—defined as those employing more than 75 percent of the female labor force in the US Northeast—ranged from 11 to 71 percent.

Entry to the United States was largely unrestricted for Canadians during the early 20th century. This allows us to observe how French Canadian migrants sorted across destinations by skill in the absence of direct immigration policy effects or substantial out-of-pocket migration costs.[11] Canadian immigration was gender balanced, with 51 percent of flows being female as compared to only 33 percent of overseas immigrant flows between 1900 and 1924 (Truesdell, 1943; Willcox, 1929). As a result, French Canadian migration to the US Northeast provides an excellent opportunity to assess sorting patterns of both female and male migrants in an environment with low barriers to entry, substantial differences in terms of culture and language from home, and notable variation in the labor market opportunities available to both sexes across destinations.

[8] Table C.1 shows calculations of skill premia for men and women in both countries.

[9] Low labor force participation rates of women were influenced by social norms about their role in the workplace, such as marriage and pregnancy bars (Hyland, Djankov and Goldberg, 2020; Goldin, 1988).

[10] Authors calculations using the full-count 1920 US Census.

[11] Train fares from Montreal to Boston were only $6.50 and $8.50 from Halifax to Boston circa 1900. These fares were roughly equivalent to a few days of income for an unskilled laborer (Green, MacKinnon and Minns, 2005).

3. Conceptual Framework

Economic theory predicts that migrants sort themselves across destinations by skill depending on the benefits and costs attached to each potential destination. Grogger and Hanson (2011) develop a model focusing on the earnings difference between high- and low-skilled workers (absolute returns to skill), while accounting for skill-specific migration costs in the manner of Chiquiar and Hanson (2005) and McKenzie and Rapoport (2010).[12] Their theoretical approach predicts that destinations offering higher absolute returns to skill should attract a higher-skilled mix of migrants. This framework assumes that men and women respond similarly to returns to skill and thus sort across destinations in a symmetrical manner. However, contemporary research shows that gender pay gaps partly result from differences in how men and women sort across firms and occupations (Card, Cardoso and Kline, 2016). Given the restrictive social norms regarding women's employment a century ago, female migrants likely faced greater challenges in matching their skills and earnings potential to destination labor markets.

One explanation for gender differences in migrant sorting is that women may respond less strongly than men to variation in returns to skill across destinations. A second explanation would arise from the presence of labor market constraints faced by women. For women seeking work circa 1920, the main constraint was the extent to which women could readily work in different destination labor markets. As we later show, employment probabilities for women varied widely across the US Northeast. This variation could have significantly influenced female migrants' sorting, as the likelihood of finding work directly affects the expected returns to migration (Hatton, 1995, 2010).

While some destinations may offer more opportunities for women to work across the breadth of the labor market, another way in which the structure of labor markets could influence women's sorting across destinations is through local differences in the presence of feminized occupations, where there were lower barriers to female employment. The presence of female-intensive industries would be expected to reduce search costs, but primarily for a subset of "women's jobs" characterized by limited prospects for promotion or earnings growth with experience (Abbott, 1906; Goldin, 2006). If such positions were biased towards less skilled workers with relatively short work horizons, then the presence of feminized occupations would mostly make a destination more attractive to women with lower earnings potential.

Marital status is another potential driver of women's sorting patterns. Previous research suggests that female migration flows in the lat 19th and early 20th century were less sensitive to labor market

[12] This approach contrasts with earlier work focusing on relative returns to skill to explain the selection and sorting of migrants (Borjas, 1987).

conditions than those of male migrants (Tortorici and Fernández Sánchez, 2023; Greenwood, 2008). This pattern is consistent with many women—particularly those married at the time of migration—being tied movers or secondary earners (Mincer, 1978; McKinnish, 2008). In such cases, labor market characteristics may not explain sorting patterns, as the destination choice primarily reflects their male partners' income maximization decisions. [13]

4. Data

Our analysis of migrant sorting draws on data from immigration cards issued by the US Immigration and Naturalization Service (INS) and maintained by the National Archives and Records Administration (NARA). These documents were the main administrative tool to record the flow of immigrants from and via Canada during the early 20th century (Ramirez and Otis, 2001, p. 190).[14] We digitize a twenty percent sample (8,336 individual border crossings) of the immigration cards in NARA's publication M1462.[15] We also use US Census data and other sources to calculate county-level characteristics that may have influenced immigrants' destination choices.

4.1 Outcome Variable

The immigration cards report a wide range of individual demographic and geographic information, including age, height, literacy, marital status, nationality, occupation, race, sex, locality of birth, last permanent residence, and intended destination.[16] They also report the intended time to remain, previous immigration experience, and contacts in the United States. For our purposes in this paper, the key data are the origin and destination locations, from which we identify migration corridors (origin-destination pairs), and the height of migrants, which we use to make inferences about migrant sorting. The cards provide individual information which includes marital status, but we typically do not observe information about the spouse. For this reason we model migration as an individual decision only, rather than formally considering joint migration of couples in our empirical exercise.[17]

[13] In the Online Appendix (Section A), we present a formal extension of Grogger and Hanson's (2011) model that accounts for gender differences in migrant sorting and provides predictions for single and married migrants.

[14] The recording of immigrants started in 1895 and was formalized under the Immigration Act of 1903, which instructed the inspection of aliens along the borders of Canada and Mexico (US Congress, 1903, p. 1221).

[15] The publication M1462—"Alphabetical Index to Canadian Border Entries through Small Ports in Vermont, 1895-1924"—contains 41,679 cards arranged alphabetically across 6 microfilm reels. The records in our sample span from 1880 to 1954, with the bulk being from 1917 and 1923 (see Figure B.4). The Online Appendix (Section D) provides further information about these data.

[16] We classify localities of birth and last residence into Canadian regions (see Figure B.1) and destination localities into US counties. We follow St-Hilaire et al. (2007) to classify Canadian localities. The Online Appendix (Section F) provides further details on the classification method.

[17] A manual search of the complete records allowed us to match up over 100 spouse pairs, which we use to describe the extent of assortative matching in marriage in Appendix Figure B.6.

We restrict the sample to individuals reporting complete geographic data and reporting counties in New England, New York, New Jersey, and Pennsylvania as intended destinations. French Canadian parishes were established in these regions over the course of the 19th century, attracting the bulk of French Canadian migrants to the United States (MacDonald, 1898). We also limit the sample to French Canadians based on the individuals' name (given name and family name) and "race" specified on the immigration card.[18] We further restrict the sample to individuals who had passed their pubertal growth spurt before being observed: males aged 16—65 years and females aged 14—65 years. This refinement avoids capturing growing and shrinkage effects, which can distort sorting estimates based on physical stature (Spitzer and Zimran, 2018). Following these restrictions, our final sample contains 3,224 individual border crossings (2,033 men and 1,191 women) spanning from 1905 to 1925. We then collapse the individual-level data to origin-destination pairs to estimate the average height of men and women (single or single and married combined) for up to 302 migration corridors. Height serves as an ideal outcome variable because it allows us to assess sorting patterns regardless of migrants' labor market status—particularly valuable for studying female migrants, who rarely reported an occupation. Extensive research demonstrates that physical stature correlates with earnings potential (see Spitzer and Zimran, 2018, for a review). This relationship persisted in the United States throughout the 20th century (Case and Paxson, 2008), with Schultz (2002) finding that each additional centimeter of adult height corresponded to wage increases of 0.45 and 0.31 percent for men and women, respectively, in the early 1990s.[19]

4.2 Explanatory Variables

We use the full-count 1910 US Census to estimate county-level characteristics that approximate the labor market conditions that migrants would anticipate on arrival. Following Grogger and Hanson (2011), we compute absolute returns to skill as the difference between prospective earnings for the top and bottom 20 percent of the income distribution. Since earnings are not reported in the 1910 US Census, we use full-count data from the 1940 Census and the Saavedra and Twinam (2020) LASSO procedure to assign individual earnings in 1910 by occupation code for each county in New England, New Jersey, New York, and Pennsylvania.[20] We then use the distribution of occupations in each county in 1920 to estimate local occupational earning scores at the 80th and 20th percentiles. We then compute

[18]We use this information to classify immigrants as British (English, Irish, Scotch, or Welsh) or French Canadians following the *Dictionary of Races or Peoples* (Folkmar, 1911). The Online Appendix (Section E) provides a full description of the methodology we follow to identify French surnames.

[19]Height correlates with other economic indicators in our data. Figure B.5 shows a positive relationship between migrants' stature and the cash they declared at entry.

[20]This approach uses variation in earnings by demographic characteristics in 1940 to predict individual earnings in earlier censuses. The occupational coding is the IPUMS default 1950 Census occupational classification system.

their difference to get our measure of absolute returns to skill in each destination. We apply this method for both employed men and employed women to generate sex-specific, local, absolute returns to skill.

To measure sex-specific employment probabilities, we calculate the shares of men and women aged 16-65 who reported being employed in 1910. While employment probabilities were high for men in all counties, there were significant differences for women, which we interpret as representing variation in the accessibility of work for women across destinations.[21] Variation in female employment may also be related to the presence of industries with more feminized occupations. We define female-intensive industries as those employing more than 75 percent of the female labor force in 1910: manufacturing of non-durable goods, retail trade, personal services, and telecommunications, and calculate the share of these sectors in the overall employment for each destination.

We calculate a set of additional destination-specific control variables. These include the working-age (16-65) population share of French Canadians, the population share living in cities (25K+ inhabitants), the population share living in rural areas, the mortality rate, the share of illiterate adults (10+ years old), the share of children attending school (6-14 years old), the shares of farm and non-farm homes owned, the average value of owner-operated farms, the adult female-to-male ratio, and the population share of Catholics. As our period of analysis overlaps with the first wave of the Great Migration, we also include the population share of black Americans to control for potential labor market competition between black workers and immigrants (Boustan, 2009; Collins and Wanamaker, 2015). We compute all variables from the 1910 US Census except for information on Catholicism, which we obtain from Manson et al. (2023). Finally, we estimate the origin-destination distance for each corridor and create an indicator for corridors with contiguous origin-destination locations. These variables control for out-of-pocket migration costs and border dynamics which can influence destination choice (Hatton and Williamson, 1998).

4.3 Migration patterns and destination characteristics

Figure 1 displays migration corridors (Canadian region-US county pairs) for unmarried individuals observed in our sample. Corridors with more intense migration, captured by brighter lines, originate mostly in Southern Quebec—especially in locations between the Richelieu and Chaudière rivers—and have as destinations counties in New England (Maine, Vermont, New Hampshire, Massachusetts, Connecticut, and Rhode Island). We also observe corridors originating north of the St. Lawrence River and ending in the states of New York, New Jersey, and Pennsylvania, though with less intense migration.

[21] We cannot distinguish between employment rates of skilled and unskilled women with the 1910 Census data.

These origin-destination patterns are similar for both female and male migrants. Figure 1 also reports the weighted mean height across corridors. Female migrants were on average 63.3 inches (161 cm) tall, while their male counterparts were just under 66.7 inches (170 cm). Comparisons to mean heights for non-migrants reported in Pett and Ogilvie (1957) suggest that migrants were noticeably taller than their counterparts who remained in Quebec. Both French Canadian migrant men and women were at least 0.8 inches taller than non-migrants of the same birth cohort. This suggests that French Canadian migrants were positively selected from the source population (see Table C.2).

Figure 2 presents evidence on the sorting of single migrants. It compares standardized mean heights of female and male migrants by US destination. We find that despite broadly similar spatial patterns of migration, few counties attracted female and male migrants of an equivalent height profile. This is clear evidence that the destination sorting on economic potential was quantitatively different between sexes. For example, women migrants arriving in Southern New Hampshire were drawn from the upper ranks of the migrant height distribution, whereas their male counterparts were predominantly drawn from the lower end of the distribution. The opposite is observed in Grafton County and Carroll County, located in the north of the state.

Figure 3 displays spatial patterns of our two main explanatory variables. The evidence suggests that labor market conditions for women varied widely, and the destinations offering the greatest employment opportunities for women differed from those most attractive to prospective male migrants. Panel A shows that destinations offering the highest returns to skill for women were in a belt running from Maine to Northern New York. In contrast, returns to skill for men were relatively homogeneous across counties. Counties in the south of New England offered the highest employment possibilities for women—a region where feminized industries were also concentrated (see Figure B.3)—whereas employment possibilities for men were varied little across the Northeast (see Panel B).[22]

5. Empirical Approach

We use Equation 1 to estimate the relationship between destination characteristics and migrant sorting on height:

$$h_{od} = \alpha + \beta \cdot rskill_d + \lambda \cdot empl_d + \delta \cdot fintensive_d + \mathbf{Y}'_d \cdot \mathbf{\Delta} + \mathbf{X}'_{od} \cdot \mathbf{\Gamma} + \theta_o + e_{od}, \tag{1}$$

where h_{od} is the average height of migrants from Quebec region o to US county d. $rskill_d$, $empl_d$, and $fintensive_d$ are the returns to skill, employment rates, and preponderance of female-intensive

[22]Table C.3 presents descriptive statistics of our main explanatory variables for the top 10 migration corridors.

industries at the destination. $\mathbf{Y}'_d$ is a vector of control variables that capture differences across counties in French Canadian population shares, urbanization, literacy and schooling enrollment, mortality, adult female-to-male ratios, home and farm ownership, farm value, racial composition, and Catholicism. $\mathbf{X}'_{od}$ is a vector of corridor-level controls that include the origin-destination distance (linear and quadratic) and an indicator variable for migration corridors between contiguous locations. We also include region of origin fixed effects θ_o to absorb region-specific factors that may influence height. We standardize all continuous explanatory variables and cluster standard errors at the US county level.[23]

To evaluate whether sorting patterns varied by sex, we estimate Equation 1 separately for men and women using weighted least squares (WLS), where observations are weighted by the number of migrants per corridor.[24] We present results for single migrants separately from those for the full sample of men and women (married and unmarried). This approach allows us to assess the extent to which tied migration dilutes the effect of labor market conditions and migrant sorting. While the interpretation of the coefficients on returns to skill follows Grogger and Hanson (2011), the employment variables require careful reading. The share of employed women depends on local employer preferences (Delgado Helleseter, Kuhn and Shen, 2020; Kuhn and Shen, 2013) and the industrial or occupational segmentation of labor markets by sex (Góes, Lopez-Acevedo and Robertson, 2023). In our preferred model, which includes employment probabilities and the share of female-intensive industries, the coefficient on employment probabilities captures employers' willingness to hire and retain women, conditional on the local industrial structure.

5.1 Results

Table 1 presents our main results for men (Panel A) and women (Panel B). Columns 1 to 3 show estimates for single migrants, while columns 4 to 5 show estimates for the full sample that includes both single and married migrants.[25]

Men

In line with predictions of Grogger and Hanson (2011), we find that single men sorted positively on returns to skill: counties with large skill-related wage differentials received taller men with above-average earnings potential. Column 1 shows estimates of our baseline specification. A one standard

[23]We standardize the values for men and women separately using the distribution parameters of all counties belonging to New England, Pennsylvania, New York, and New Jersey.
[24]Unweighted OLS regressions are available on request
[25]We use the working-age (16-65) population share of French Canadians to control for the presence of French Canadian enclaves. However, using sex-specific enclave measures yield very similar results (see Table C.4).

deviation increase in returns to skill is associated with a 0.15 inch (0.38 cm) increase in the height profile of single men. Controlling for the presence of female-intense industries and French Canadian enclaves (columns 2-3) adjusts the coefficients upwards—albeit by small margins—with estimates ranging from 0.20 to 0.22 inches (0.51 – 0.56 cm). While mostly statistically insignificant, the coefficient on employment probabilities suggests that destinations offering higher employment probabilities attracted migrants with below-average earnings potential, which is also consistent with Grogger and Hanson (2011)'s contemporary findings.

Column 4 presents results using the full sample of single and married men. Our results are similar to those in columns 1-3, though more precisely estimated, implying that single and married men sorted across destinations in the same way. Given that men were the primary earner in the vast majority of migrant couples in the early 20th century, it is unsurprising that marital ties did not constrain the destination choices of male migrants. One concern is that the changes in the number of migrant corridors may be driving this finding. Column 5 shows this is not the case, as we find very similar results when restricting the estimation to corridors with both married and single migrants.

Interestingly, the coefficients on female-intensive industry in columns 4-5 become statistically significant. The positive point estimates suggest that taller men moved to destinations with a greater presence of feminized industries. Figure B.6 shows that height-based assortative matching characterized French Canadian marriages, with taller men marrying taller women who had greater employment opportunities in locations with female-intensive industries.[26] One explanation for the patterns in columns 4 and 5 is that wives may have served as secondary earners, leading married men to choose destinations where family migration gains could be maximized (Mincer, 1978).

Women

We find limited evidence that returns to skill were a key driver of migrant sorting for for single women. The coefficient on returns to skill is positive and significant in columns 1 and 2, but loses statistical power when we control for the presence of French Canadian enclaves in column 3. Contrary to men, we find that variables related to female employment have a powerful effect on the sorting patterns of women in all specifications. Estimates in column 1 show that a one standard deviation increase in the share of employed women is associated with a 1.5 inch (3.8 cm) increase in the height profile of single women. This implies that taller women with above-average earnings potential prioritized employment

[26] Curtis (2022) provides further evidence on the existence of strong assortative matching in marriage in nineteenth-century Quebec. In the Online Appendix (Section A), we formally show that assortative matching in marriage plays an important role in explaining sorting patterns of migrant couples.

probabilities over returns to skill. Controlling for the presence of female-intensive industries does not significantly change the coefficient on employment probabilities. However, the coefficient on female-intensive industries is highly significant and negative, implying that destinations with greater preponderance of feminized industries disproportionately attracted shorter female migrants. This is consistent with feminized industries usually offering jobs with limited earnings growth and attracting migrants with different economic potential as a result. These results show that the destination choices of single women were highly responsive to labor market conditions, but in a strikingly different way to their male counterparts. Furthermore, our results are in line with recent research showing that frictions associated with job search due to gender-based discrimination play a major role in how female skill is allocated across space (Lee, 2024).

Column 4 presents results using the full sample of single and married women. We find no significant relationship between labor market variables and sorting across destinations in this specification. The attenuation of the point estimates is consistent with most married women in the early 20th century having relatively little economic agency in migration: arriving in the United States and settling in a destination through a tie to their spouse, whose labor market opportunities determined the couples' destination choice. This finding echoes results of earlier studies that are unable to observe the marital status of female migrants (Tortorici and Fernández Sánchez, 2023; Greenwood, 2008). When restricting the analysis to corridors with single and married migrants (column 5), the labor market variables recover statistical significance with only moderate attenuation.

5.2 Addressing Endogeneity

Our analysis based on Equation 1 reveals significant gender-specific correlations between destination labor market characteristics and migrant heights. Although we control for a rich set of destination characteristics with origin fixed effects, interpreting these results as causal effects remains challenging. One might be particularly concerned about the endogeneity of employment probabilities—the most significant factor explaining differences in sorting patterns between men and women (see Table 1)—as these simultaneously depend on the supply and demand of labor across different destinations.

We address the potential endogeneity of employment probabilities through a shift-share instrumental variable strategy (SSIV). Our approach instruments $empl_{d,1910}$ with the sum of workers in 1900 and the

SSIV-predicted change from 1900 to 1910, divided by the working-age population in 1900:[27]

$$Z = \widehat{empl_{d,1910}} = \left(\frac{workers_{d,1900} + \Delta \widehat{workers_{d,1900-1910}}}{workagepop_{d,1900}} \right). \tag{2}$$

The predicted change in the number of workers is computed following a standard SSIV structure as in Blanchard et al. (1992):

$$\Delta \widehat{workers_{d,1900-1910}} = \sum_k share_{d,k}^{1900} \cdot \theta_k^{1900-1910}, \tag{3}$$

where k indexes industries.[28] We predict county-level employment growth by multiplying each industry's aggregate employment growth in the Northeast, $\theta_k^{1900-1910}$, by its corresponding county employment share, $share_{d,k}^{1900}$. This distributes regional growth in employment across counties based on an earlier distribution, so that identification relies on exogenous shifts (Borusyak, Hull and Jaravel, 2022) instead of exogenous shares (Goldsmith-Pinkham, Sorkin and Swift, 2020). Note that we construct the instrument separately for men and women using sex-specific variables.

We present SSIV estimates in the Appendix (Table C.5).[29] Consistent with our OLS results, we find that employment probabilities did not influence the sorting of men regardless of their marital status (Panel A). However, this result should be interpreted with caution, as it may suffer from weak instrument bias reflected in the low Kleibergen-Paap F-statistics. For women, the instrument is sufficiently strong, and the estimates confirm our previous findings: employment probabilities were the main driver of migrant sorting instead of returns to skill (Panel B). The point estimates for our preferred specification show that increasing female labor force participation by one standard deviation—as an example, from 33.4% in York, Maine to 40.7% in Androscoggin, Maine—would induce a 2.6 inch increase in the height profile of single migrant women. One difference from the WLS results is that we find significant positive effects of employment probabilities on sorting when combining married and single women. An earlier literature emphasized that French Canadian wives played an important role as secondary earners in New England (Ramirez, 1986). The SSIV results are consistent with the notion that wives' employment potential may have mattered for destination choice.

[27] As the 1900 full-count census does not record employment status, we proxy employment with occupational responses to *occ1950*.

[28] We define industries based on the *ind1950* variable aggregated to the 2-digit level, which identifies 17 different industries.

[29] All estimates are robust to overrejection bias, a common issue in identification strategies based on shift-share instruments (see Adao, Kolesár and Morales, 2019).

6. Conclusion

Using data on migrant stature from border crossing records, we study sorting patterns of French Canadian immigrants to the United States in the early 20th century. To our knowledge, this is the first study to demonstrate how migrants sorted across US destinations by skill during this period, and the first in any period to compare sorting patterns between women and men.

Our analysis yields several important findings for understanding gender differences in the destination choice of migrants. First, sorting across destinations by skill differed strikingly between men and women. Tall women with high earnings potential chose different destinations than tall men. Second, the determinants of sorting varied notably by gender. As in contemporary settings, male sorting responded strongly to differences in absolute return to skill (Grogger and Hanson, 2011). However, returns to skill were much less important for female sorting patterns. Conditional on local economic structure, employment probabilities were the main driver: taller women were drawn to destinations with greater changes for women to find work. We interpret this finding as evidence that in settings with pervasive labor market discrimination against women, female migrants seeking work would face a trade-off between matching their skills to labor market returns and access to the workplace.

Although drawn from a historical setting, our findings have broader implications for migration research. Women's employment opportunities remain limited by institutional and social norms in many regions, including immigration-dependent economies such as the Persian Gulf states (Bursztyn, González and Yanagizawa-Drott, 2020; ILO, 2023; World Bank, 2016).[30] Our results suggest that destinations with high gender discrimination may attract a lower-skilled mix of female migrants despite offering high returns to skill. Following Lee (2024), this suboptimal allocation of female human capital could significantly affect the productivity of major immigrant destinations. This raises an important question for future research: what are the long-run implications for immigrant integration in destinations where discrimination discourages the arrival of skilled migrant women?

[30] Immigrants represent about 38.7 and 88.1 percent of the population in Saudi Arabia and the UAE, respectively (United Nations, 2020).

References

Abbott, Edith. 1906. "The history of industrial employment of women in the United States: An Introductory Study." *Journal of Political Economy*, 14(8): 461–501.

Abramitzky, Ran, and Leah Boustan. 2017. "Immigration in American Economic History." *Journal of Economic Literature*, 55(4): 1311–45.

Adao, Rodrigo, Michal Kolesár, and Eduardo Morales. 2019. "Shift-share designs: Theory and inference." *The Quarterly Journal of Economics*, 134(4): 1949–2010.

Aksoy, Cevat Giray, and Panu Poutvaara. 2021. "Refugees' and irregular migrants' self-selection into Europe." *Journal of Development Economics*, 152: 102681.

Artuç, Erhan, Frédéric Docquier, Çaglar Özden, and Christopher Parsons. 2015. "A global assessment of human capital mobility: the role of non-OECD destinations." *World Development*, 65: 6–26.

Bertoli, Simone, Jesús Fernández-Huertas Moraga, and Francesc Ortega. 2013. "Crossing the border: Self-selection, earnings and individual migration decisions." *Journal of Development Economics*, 101: 75–91.

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

Blanchard, Olivier Jean, Lawrence F Katz, Robert E Hall, and Barry Eichengreen. 1992. "Regional evolutions." *Brookings Papers on Economic Activity*, 1992(1): 1–75.

Borjas, George J. 1987. "Self-Selection and the Earnings of Immigrants." *American Economic Review*, 77(4): 531–553.

Borjas, George J, and Stephen G Bronars. 1991. "Immigration and the Family." *Journal of Labor Economics*, 9(2): 123–148.

Borrescio-Higa, Florencia, Carlos Guillermo Bozzoli, and Federico Droller. 2019. "Early life environment and adult height: The case of Chile." *Economics & Human Biology*, 33: 134–143.

Borusyak, Kirill, Peter Hull, and Xavier Jaravel. 2022. "Quasi-experimental shift-share research designs." *The Review of Economic Studies*, 89(1): 181–213.

Boustan, Leah Platt. 2009. "Competition in the Promised Land: Black Migration and Racial Wage Convergence in the North, 1940–1970." *The Journal of Economic History*, 69(3): 755–782.

Brault, Gerard J. 1986. *The French-Canadian Heritage in New England.* McGill-Queen's University Press.

Bursztyn, Leonardo, Alessandra L González, and David Yanagizawa-Drott. 2020. "Misperceived social norms: Women working outside the home in Saudi Arabia." *American Economic Review*, 110(10): 2997–3029.

Card, David, Ana Rute Cardoso, and Patrick Kline. 2016. "Bargaining, sorting, and the gender wage gap: Quantifying the impact of firms on the relative pay of women." *Quarterly Journal of Economics*, 131(2): 633–686.

Case, Anne, and Christina Paxson. 2008. "Stature and status: Height, ability, and labor market outcomes." *Journal of Political Economy*, 116(3): 499–532.

Chiquiar, Daniel, and Gordon H Hanson. 2005. "International migration, self-selection, and the distribution of wages: Evidence from Mexico and the United States." *Journal of Political Economy*, 113(2): 239–281.

Cobb-Clark, Deborah A. 1993. "Immigrant selectivity and wages: the evidence for women." *The American Economic Review*, 83(4): 986–993.

Collins, William J, and Marianne H Wanamaker. 2015. "The Great Migration in Black and White: New Evidence on the Selection and Sorting of Southern Migrants." *The Journal of Economic History*, 75(4): 947–992.

Curtis, Matthew. 2022. "The her in inheritance: marriage and mobility in Quebec 1800–1970." Université Libre de Bruxelles, ECARES Working Paper 2022-38.

Delgado Helleseter, Miguel, Peter Kuhn, and Kailing Shen. 2020. "The Age Twist in Employers' Gender Requests: Evidence from Four Job Boards." *Journal of Human Resources*, 55(2): 428–469.

Docquier, Frédéric, Abdeslam Marfouk, Sara Salomone, and Khalid Sekkat. 2012. "Are skilled women more migratory than skilled men?" *World Development*, 40(2): 251–265.

Docquier, Frédéric, B Lindsay Lowell, and Abdeslam Marfouk. 2009. "A gendered assessment of highly skilled emigration." *Population and Development Review*, 35(2): 297–321.

Dribe, Martin, Björn Eriksson, and Jonas Helgertz. 2023. "From Sweden to America: migrant selection in the transatlantic migration, 1890–1910." *European Review of Economic History*, 27: 24–44.

Dustmann, Christian, and Tommaso Frattini. 2014. "The Fiscal Effects of Immigration to the UK." *The*

Economic Journal, 124(580): F593–F643.

Egger, Peter H, Maximilian Von Ehrlich, and Douglas R Nelson. 2012. “Migration and Trade.” *The World Economy*, 35(2): 216–241.

Folkmar, Daniel. 1911. “Dictionary of Races or Peoples.” In *Reports of the Immigration Commission.* Vol. 5, , ed. William Paul Dillingham. U.S. Government Printing Office.

Geneanet. 2024. “Geneanet Family Name Search Engine.” https://en.geneanet.org/surnames/, Accessed: 2021-08-13.

Góes, Carlos, Gladys Lopez-Acevedo, and Raymond Robertson. 2023. “Gender-Segmented Labor Markets and Trade Shocks.” IZA Institute of Labor Economics Discussion Paper Series No. 15892.

Goldin, Claudia. 1986. “The Female Labor Force and American Economic Growth, 1890-1980.” In *Long-term factors in American economic growth.* 557–604. Chicago: University of Chicago Press.

Goldin, Claudia. 1988. “Marriage Bars: Discrimination Against Married Women Workers, 1920’s to 1950’s.” National Bureau of Economic Research Working Paper 2747.

Goldin, Claudia. 1990. *Understanding the gender gap: An economic history of American women.* Oxford University Press.

Goldin, Claudia. 2006. “The Quiet Revolution That Transformed Women’s Employment, Education, and Family.” *American Economic Review*, 96(2): 1–21.

Goldin, Claudia, and Lawrence F. Katz. 1999. “The Returns to Skill in the United States across the Twentieth Century.” National Bureau of Economic Research NBER Working Papers 7126.

Goldsmith-Pinkham, Paul, Isaac Sorkin, and Henry Swift. 2020. “Bartik Instruments: What, When, Why, and How.” *American Economic Review*, 110(8): 2586–2624.

Gould, David M. 1994. “Immigrant Links to the Home Country: Empirical Implications for US Bilateral Trade Flows.” *The Review of Economics and Statistics*, 302–316.

Gratton, Brian, and Emily Merchant. 2015. “An Immigrant’s Tale: The Mexican American Southwest 1850 to 1950.” *Social Science History*, 39(4): 521–50.

Green, Alan G, Mary MacKinnon, and Chris Minns. 2002. “Dominion or Republic? Migrants to North America from the United Kingdom, 1870–1910.” *The Economic History Review*, 55(4): 666–696.

Green, Alan, Mary MacKinnon, and Chris Minns. 2005. “Conspicuous by their absence: French Canadians and the Settlement of the Canadian West.” *Journal of Economic History*, 65(3): 822–849.

Greenwood, Michael J. 2008. “Family and sex-specific US immigration from Europe, 1870–1910: A panel data study of rates and composition.” *Explorations in Economic History*, 45(4): 356–382.

Grogger, Jeffrey, and Gordon H. Hanson. 2011. “Income maximization and the selection and sorting of international migrants.” *Journal of Development Economics*, 95(1): 42–57.

Hatton, Timothy J. 1995. “A Model of UK Emigration, 1870-1913.” *The Review of Economics and Statistics*, 407–415.

Hatton, Timothy J. 2010. “The Cliometrics of International Migration: A Survey.” *Journal of Economic Surveys*, 24(5): 941–969.

Hatton, Timothy J, and Jeffery G Williamson. 1994. “International migration and world development: A historical perspective.” In *Economic Aspects of International Migration.* 3–56. Springer.

Hatton, T. J, and J. G. Williamson. 1998. *The Age of Mass Migration: Causes and Economic Impact.* New York:Oxford University Press.

Hunt, Jennifer, and Marjolaine Gauthier-Loiselle. 2010. “How much does immigration boost innovation?” *American Economic Journal: Macroeconomics*, 2(2): 31–56.

Hyland, Marie, Simeon Djankov, and Pinelopi Koujianou Goldberg. 2020. “Gendered laws and women in the workforce.” *American Economic Review: Insights*, 2(4): 475–90.

ILO. 2023. “ILO Modelled Estimates Database.”

Inwood, Kris, Mary MacKinnon, and Chris Minns. 2010. “Labour market dynamics in Canada, 1891-1911: a first look from new census samples.” In *The dawn of Canada’s century: hidden histories.* , ed. A. G. Darroch. Kingston, Ontario: McGill-Queen’s Press.

Kerr, Sari Pekkala, William Kerr, Çağlar Özden, and Christopher Parsons. 2016. “Global Talent Flows.” *Journal of Economic Perspectives*, 30(4): 83–106.

Komlos, John, and Jörg Baten. 2004. “Looking Backward and Looking Forward: Anthropometric Research and the Development of Social Science History.” *Social Science History*, 28(2): 191–210.

Kuhn, Peter, and Kailing Shen. 2013. "Gender Discrimination in Job Ads: Evidence from China." *The Quarterly Journal of Economics*, 128(1): 287–336.

Lee, Munseob. 2024. "Allocation of Female Talent and Cross-Country Productivity Differences." *The Economic Journal*, ueae056.

MacDonald, William. 1898. "The French Canadians in New England." *The Quarterly Journal of Economics*, 12(3): 245–279.

MacKinnon, Mary, and Daniel Parent. 2012. "Resisting the melting pot: The long term impact of maintaining identity for Franco-Americans in New England." *Explorations in Economic History*, 49(1): 30–59.

Manson, Steven, Jonathan Schroeder, David Van Riper, Katherine Knowles, Tracy Kugler, Finn Roberts, and Steven Ruggles. 2023. "1906 Census of Religious Bodies."

McInnis, Marvin. 1994. "Immigration and Emigration: Canada in the Late Nineteenth Century." In *Migration and the International Labor Market, 1850-1939.* , ed. Timothy J. Hatton and Jeffrey G. Williamson, 139–155. Routledge.

McKenzie, David, and Hillel Rapoport. 2010. "Self-Selection Patterns in Mexico-US Migration: The Role of Migration Networks." *The Review of Economics and Statistics*, 92(4): 811–821.

McKinnish, Terra. 2008. "Spousal Mobility and Earnings." *Demography*, 45: 829–849.

Mincer, Jacob. 1978. "Family migration decisions." *Journal of Political Economy*, 86(5): 749–773.

Moser, Petra, Alessandra Voena, and Fabian Waldinger. 2014. "German Jewish émigrés and US invention." *American Economic Review*, 104(10): 3222–3255.

Parsons, Christopher, and Pierre-Louis Vézina. 2018. "Migrant Networks and Trade: The Vietnamese Boat People as a Natural Experiment." *The Economic Journal*, 128(612): F210–F234.

Pett, L.B., and G.F. Ogilvie. 1957. "The report on Canadian average weights, heights and skinfolds." *Canadian Bulletin on Nutrition*, 5: 1–81.

Ramirez, Bruno. 1986. *On the Move: French-Canadian and Italian Migrants in the North Atlantic Economy, 1860-1914.* McClelland and Stewart.

Ramirez, Bruno, and Yves Otis. 2001. *Crossing the 49th parallel: Migration from Canada to the United States, 1900-1930.* Cornell University Press.

Rosenzweig, Mark R, and Kenneth I Wolpin. 1994. "Are there increasing returns to the intergenerational production of human capital? Maternal schooling and child intellectual achievement." *Journal of Human Resources*, 670–693.

Ruggles, Steven, Catherine A. Fitch, Ronald Goeken, J. David Hacker, Matt A. Nelson, Evan Roberts, Megan Schouweiler, and Matthew Sobek. 2021. "IPUMS Ancestry Full Count Data: Version 3.0 [dataset]."

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

Schneider, Eric B, and Kota Ogasawara. 2018. "Disease and child growth in industrialising Japan: critical windows and the growth pattern, 1917-39." *Explorations in Economic History*, 69: 64–80.

Schultz, T Paul. 2002. "Wage gains associated with height as a form of health human capital." *American Economic Review*, 92(2): 349–353.

Smith, Marian L. 2000. "By way of Canada: US Records of immigration across the US-Canadian border, 1895-1954 (St. Albans Lists)." *Prologue: Quarterly of the National Archives*, 32(3): 192–199.

Spitzer, Yannay, and Ariell Zimran. 2018. "Migrant self-selection: Anthropometric evidence from the mass migration of Italians to the United States, 1907–1925." *Journal of Development Economics*, 134: 226–247.

St-Hilaire, Marc, Byron Moldofsky, Laurent Richard, and Mariange Beaudry. 2007. "Geocoding and mapping historical census data: The geographical component of the Canadian Century Research Infrastructure." *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 40(2): 76–91.

Storesletten, Kjetil. 2000. "Sustaining Fiscal Policy Through Immigration." *Journal of Political Economy*, 108(2): 300–323.

Thomas, Dorothy S. 1938. *Research Memorandum on Migration Differentials.* Social Science Research Council, Bulletin 43.

Tortorici, Gaspare, and Martín Fernández Sánchez. 2023. "Male and Female Self-Selection During the Portuguese Mass Migration, 1885-1930." Social Science Research Network, SSRN Working Paper 4180590.

Truesdell, Leon Edgar. 1943. *The Canadian Born in the United States: an Analysis of the Statistics of the Canadian Element in the Population of the United States, 1850 to 1930.* Yale University Press.

United Nations. 2020. "Global Migration Database."

US Congress. 1903. "Immigration Act, 32 Stat. 1213, 57th Congress."

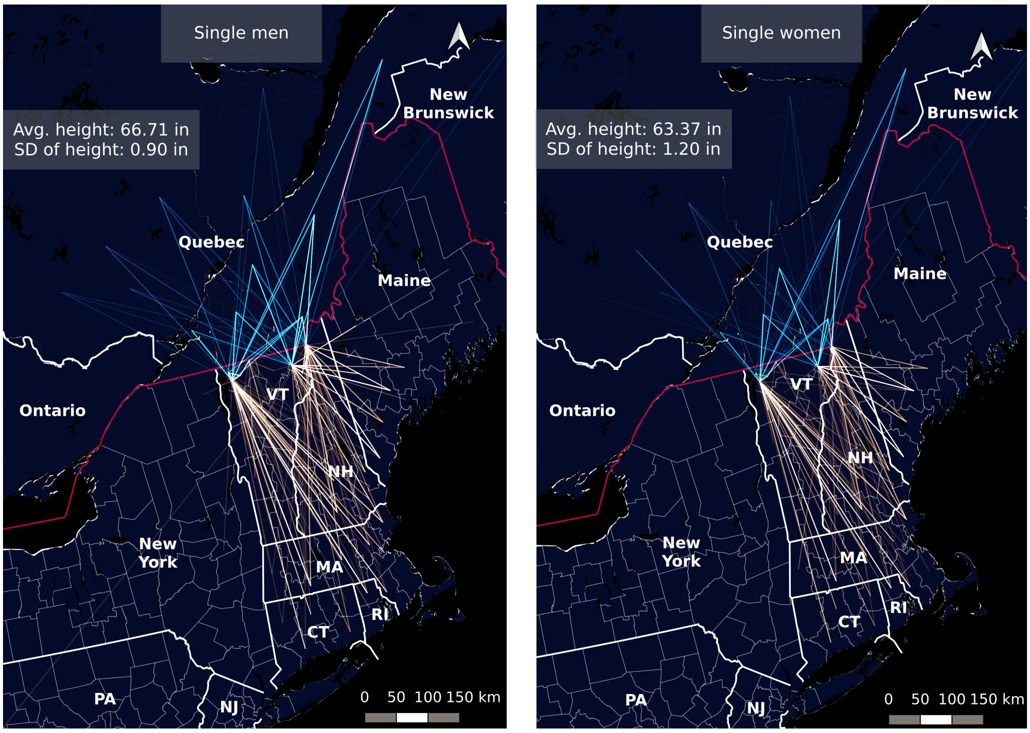
von Berlepsch, Viola, Andrés Rodríguez-Pose, and Neil Lee. 2019. "A woman's touch? Female migration and economic development in the United States." *Regional Studies*, 53(1): 131–145.

Willcox, Walter F. 1929. "Statistics of migrations, national tables, United States." In *International Migrations, Volume I: Statistics.* NBER.

World Bank. 2016. *Paving the Way for Women's Economic Inclusion in the Gulf Cooperation Council.* World Bank Group.

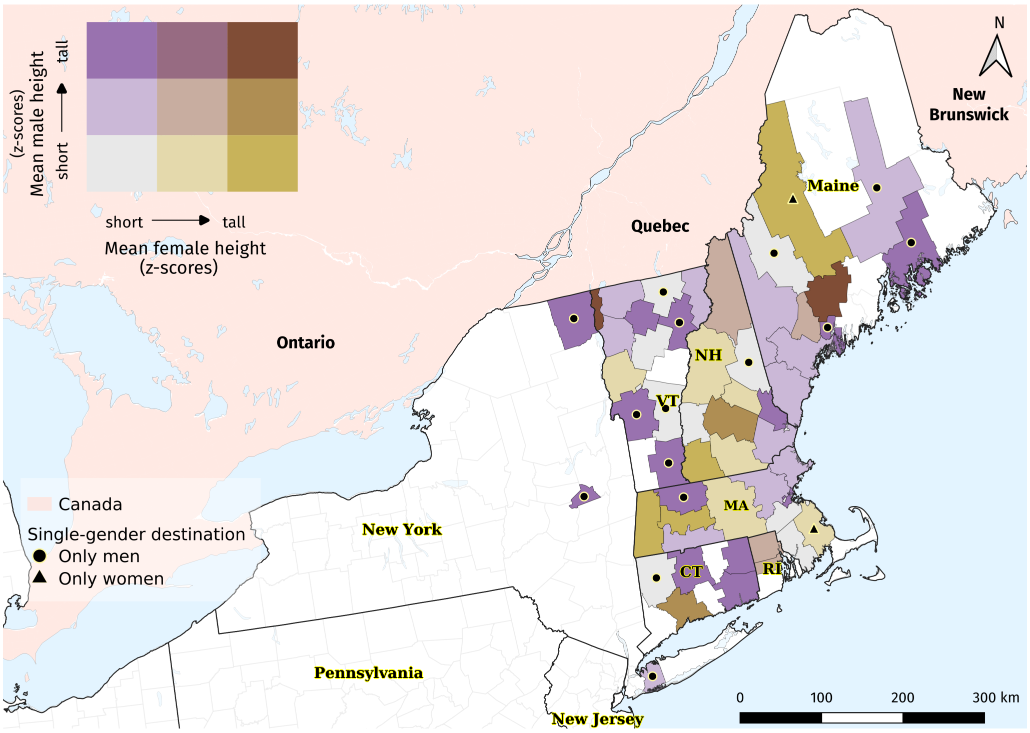
Figures and Tables

Figure 1: Origins and Destinations of French Canadian Migrants, 1905-1925



Source: St. Albans Lists, Publication number M1462.
Notes: The map displays migration corridors: origin (Canadian region)-destination (US county) pairs. Each line represents an individual. Overlapping lines capture the intensity of migration by adding pixel values—brighter lines represent more intense migration flows. *Avg height*: weighted mean height across corridors. The number of corridors for single male migrants is 213 with 1,285 individuals. For single female migrants, the numbers are 162 and 712, respectively. State abbreviations: Connecticut (CT), Massachusetts (MA), New Hampshire (NH), New Jersey (NJ), Pennsylvania (PA), Rhode Island (RI), and Vermont (VT).

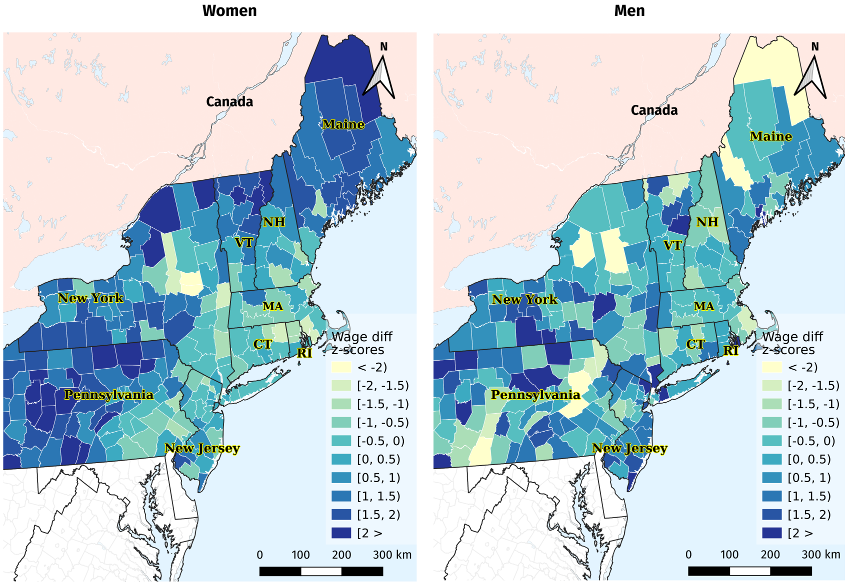
Figure 2: Sorting of Single French Canadian Migrants by Height



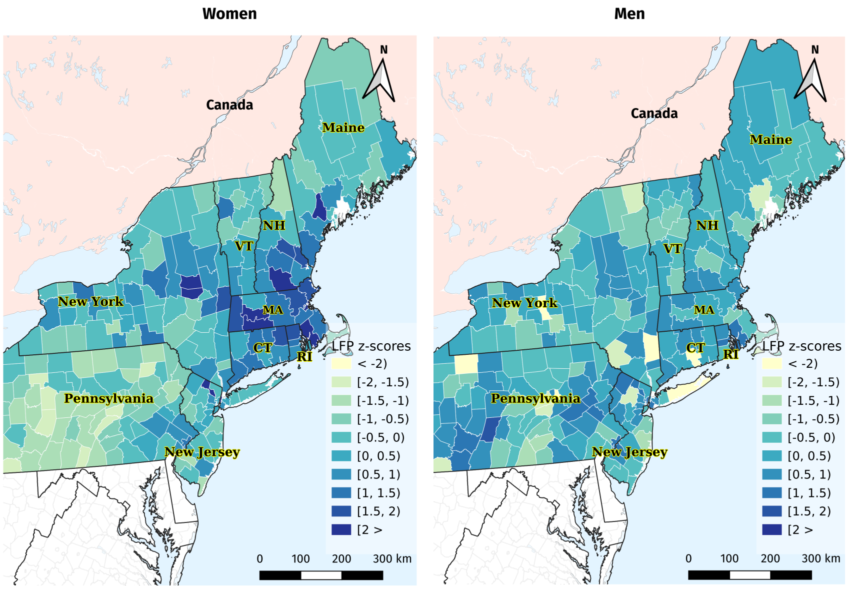
Source: St. Albans Lists, Publication number M1462.
Notes: The map displays mean height by county, standardized separately for men and women. State abbreviations: Connecticut (CT), Massachusetts (MA), New Hampshire (NH), New Jersey (NJ), Pennsylvania (PA), Rhode Island (RI), and Vermont (VT).

Figure 3: Labor Market Characteristics at the Destination

Panel A. Returns to Skill



Panel B. Labor Force Participation



Source: 1920 US full count census (Ruggles et al., 2021).
Notes: The maps display the spatial distribution of standardized absolute returns to skill and labor force participation rates. We standardize the values for men and women separately using the distribution parameters of all counties belonging to New England, Pennsylvania, New York, and New Jersey. State abbreviations: Connecticut (CT), Massachusetts (MA), New Hampshire (NH), New Jersey (NJ), Pennsylvania (PA), Rhode Island (RI), and Vermont (VT).

Table 1: Sorting of French Canadian Migrants

	(1)	(2)	(3)	(4)	(5)
Panel A: Men					
	Single			Single and Married	
				All corridors	Consistent corridors
Returns to skill	0.148*	0.215*	0.196*	0.178**	0.175*
	(0.084)	(0.114)	(0.109)	(0.081)	(0.091)
Employment probabilities	-0.196	-0.235*	-0.212	-0.130	-0.155
	(0.141)	(0.139)	(0.131)	(0.086)	(0.094)
Female-intensive industries		0.181	0.230	0.215*	0.234*
		(0.168)	(0.154)	(0.119)	(0.128)
Migration corridors	202	202	202	254	202
No. of migrants	1,201	1,201	1,201	2,033	1,963
R-squared	0.137	0.140	0.142	0.146	0.171
Panel B: Women					
	Single			Single and Married	
				All corridors	Consistent corridors
Returns to skill	0.939**	0.819*	0.716	0.186	0.747*
	(0.437)	(0.448)	(0.517)	(0.250)	(0.381)
Employment probabilities	1.457**	1.619***	1.515**	0.391	1.125**
	(0.540)	(0.583)	(0.643)	(0.354)	(0.482)
Female-intensive industries		-0.496***	-0.405**	0.013	-0.140
		(0.165)	(0.196)	(0.153)	(0.157)
Migration corridors	159	159	159	202	159
No. of Migrants	670	670	670	1,191	1,131
R-squared	0.279	0.292	0.295	0.297	0.338
Origin FE - region	Yes	Yes	Yes	Yes	Yes
Migration corridor controls	Yes	Yes	Yes	Yes	Yes
Destination county controls	Yes	Yes	Yes	Yes	Yes
Female-intense industries	No	Yes	Yes	Yes	Yes
French Canadian enclave	No	No	Yes	Yes	Yes

Source: St. Albans Lists, Publication number M1462.

Notes: *Returns to skill* is the difference in occupational income score between the 80th and 20th percentiles. *Employment probabilities* is the working-age (16-65) population share that is employed. *Female-intensive* industries is the labor force share in female-intensive industries (retail trade, telecommunications, manufacturing of non-durable goods, and personal services). Migration corridor controls include the linear and quadratic origin-destination distance and an indicator for origin and destination locations on the border. Destination county controls include the population share living in cities (25K+ inhabitants), population share living in rural areas, mortality rate, population share of illiterate (10+ years old), population share attending school (6-14 years old), share of non-farm homes owned free, share of farm homes owned free, average value of owner-operated farms, female-to-male sex ratio, religious population share of Catholics, and population share of blacks. French Canadian enclave is the population share of working-age individuals that reported French as native language and Canada as birthplace. All continues explanatory variables are standardized. The unit of observation is migration corridor. All regressions are weighted by corridor size (number of male or female migrants). Robust standard errors clustered by destination county. * = Significant at 10% level; ** = Significant at 5% level; *** = Significant at 1% level.

Online Appendix for

Explaining gender differences in migrant sorting: Evidence from Canada-US migration

December 4, 2024

A. A Conceptual Framework of Migrant Sorting by Gender

We adapt the Grogger and Hanson (2011) model of international migration to illustrate how gender-specific factors can influence the destination choice of both single and married (tied) migrants. The model uses an income maximization framework to generate predictions on the scale of migration, the selection of migrants, and the sorting of migrants by skill across destinations. To explain sorting patterns, the model focuses on absolute earnings differences between skill groups in the destination while accounting for skill-related migration costs in the manner of Chiquiar and Hanson (2005) or McKenzie and Rapoport (2010).

Single Migrants

The model assumes that individuals with different skills consider wages w and migration costs c in their migration decision. Migration costs consist of a fixed component f and a skill-varying component g such that

$$c^j_{ish} = f_{sh} + g^j_{sh}, \tag{4}$$

where c^j_{ish} is the cost of migrating from source s to destination h for individual i belonging to skill group j. We consider two skill groups for simplicity: 1 (unskilled) and 2 (skilled). Assuming that the utility associated with migrating from s to h is a linear function of the difference between wages and migration costs, we can write a utility function for an individual as

$$U^j_{ish} = \alpha\,(w^j_{ih} - c^j_{ish}) + \varepsilon^j_{ish}, \tag{5}$$

where $\alpha > 0$ is the marginal utility of income and ε^j_{ish} is an idiosyncratic error term. The log odds of migrating to h versus staying in s for skill group j can be written as

$$ln\,\frac{E^j_{sh}}{E^j_s} = \alpha\,(w^j_h - w^j_s) - \alpha\,f_{sh} - \alpha\,g^j_{sh}, \tag{6}$$

where E^j_{sh} is the share of skill group j that migrates from s to h, and E^j_s is the share that remains in the source location. Taking differences between skilled and unskilled individuals from the above equation yields predictions about migrant selection:

$$ln\,\frac{E^2_{sh}}{E^1_{sh}} - ln\,\frac{E^2_s}{E^1_s} = \alpha\,[(w^2_h - w^2_s - g^2_{sh}) - (w^1_h - w^1_s - g^1_{sh})]. \tag{7}$$

The left-hand side compares the skill mix of migrants to that of non-migrants. The right-hand side shows that selection (the sign of the left-hand side) depends on the magnitude of the wage difference between the source and destination faced by each skill group and the size of skill-varying migration costs. Rearranging the above equation yields

$$ln\,\frac{E_{sh}^{2}}{E_{sh}^{1}} = \alpha\,(w_{h}^{2} - w_{h}^{1}) - \alpha\,(g_{sh}^{2} - g_{sh}^{1}) + ln\,\frac{E_{s}^{2}}{E_{s}^{1}} - \alpha\,(w_{s}^{2} - w_{s}^{1}), \tag{8}$$

where the first two terms of the right-hand side capture the rewards to skill (net of migration costs) that explain the intensity of sorting. Destinations offering higher net rewards to skill should receive a higher-skilled mix of migrants from source s. To observe the model's implication on migrant sorting by sex, we can index Equation 8 by gender with $k = \{\mathrm{m = men, f = women}\}$, and take the difference between men and women to compare sorting

$$ln\,\frac{E_{sh}^{2,m}}{E_{sh}^{1,m}} - ln\,\frac{E_{sh}^{2,f}}{E_{sh}^{1,f}} = \alpha\left[(w_{h}^{2,m} - w_{h}^{1,m}) - (w_{h}^{2,f} - w_{h}^{1,f})\right] + \alpha\left[(g_{h}^{2,m} - g_{h}^{1,m}) - (g_{h}^{2,f} - g_{h}^{1,f})\right] + (\tau_{s}^{m} - \tau_{s}^{f}) \tag{9}$$

where $\tau_{s}^{k} = ln\,(E_{s}^{2,k}/E_{s}^{1,k}) - \alpha\,(w_{s}^{2,k} - w_{s}^{1,k})$. Note that for simplicity, we assume that men and women face the same marginal utility of income. Equation 9 states that under complete gender parity the skill mix of migrants should be gender-balanced across destinations given we hold source county s characteristics fixed, as men and women face the same rewards to skill $(w_{h}^{2,m} - w_{h}^{1,m}) - (w_{h}^{2,f} - w_{h}^{1,f}) = 0$ and skill-related migration costs $(g_{sh}^{2,m} - g_{sh}^{1,m}) - (g_{sh}^{2,f} - g_{sh}^{1,f}) = 0$. Any gender-specific factor affecting the benefits or costs associated with migration would generate distinctive sorting patterns by gender.

Married Migrants

To study the sorting of married (tied) migrants across destinations, we augment Equation 5 by introducing spousal returns to migration to the individual income maximization problem:

$$U_{ish}^{j} = \alpha\,(w_{ih}^{j} - c_{ish}^{j} + \theta_{ish}^{j\,spouse}) + \varepsilon_{ish}^{j}, \tag{10}$$

where $\theta_{ish}^{j\,spouse} = w_{ih}^{j\,spouse} - c_{ish}^{j\,spouse}$ is the spouse's net income faced by the individual i. The individual i will base their understanding of the net income of their spouse on observables, and thus will not account for the idiosyncratic error term $\varepsilon_{ish}^{j\,spouse}$ that would enter the spouse's own utility function. Accordingly, Equation 10 treats spouse net income as an endowment from the perspective of an individual migration decision. This approach is consistent with the migrant data we possess. Outside a small subsample, we do not observe spousal pairs but individuals and their marital status, which does not allow us to model sequential migration decisions of spouses. To see how the spouse net income influences the scale of tied migration, we extend Equation 6 as

$$ln\frac{E_{sh}^{j}}{E_{s}^{j}} = \alpha\,(w_{h}^{j} - w_{s}^{j} - g_{sh}^{j} - f_{sh}) + \alpha\,(w_{h}^{j_{spouse}} - w_{s}^{j_{spouse}} - g_{sh}^{j_{spouse}} - f_{sh}). \tag{11}$$

Equation 11 shows that the log odds of migrating for skill group j depend positively on the spouse's skill-group-specific difference in wages between destination h and source s net of migration costs. Note that for simplicity we assume that both spouses face the same marginal utility of income, α, and fixed migration costs, f_{sh}. In this sense, fixed costs paid by spouses can deter tied migration if they are sufficiently large.

To see the implications on the skill-mix of migrants, we can take differences between skilled (2) and unskilled (1) individuals from the above equation to yield

$$\begin{aligned} ln\frac{E_{sh}^{2}}{E_{sh}^{1}} - ln\frac{E_{s}^{2}}{E_{s}^{1}} = \alpha\,[(w_{h}^{2} - w_{s}^{2} - g_{sh}^{2}) - (w_{h}^{1} - w_{s}^{1} - g_{sh}^{1})] + \alpha[(w_{h}^{2_{spouse}} - w_{s}^{2_{spouse}} - g_{sh}^{2_{spouse}}) \\ - (w_{h}^{1_{spouse}} - w_{s}^{1_{spouse}} - g_{sh}^{1_{spouse}})]. \end{aligned} \tag{12}$$

Equation 12 indicates that the selection of migrant couples depends on two factors. The first is the degree of assortative matching in marriage captured by the correlation between j and j_{spouse}. A low correlation between j and j_{spouse} implies that on average, spousal benefits and costs of migration are similar for skilled and unskilled potential migrants. In this case, the skill mix of migrants is relatively unaffected by marriage partners. Although we cannot comprehensively test this interdependency given the lack of information on spouses for most of our migrants, we were able to obtain data about the spousal characteristics, including height, for about 200 married migrants. Figure B.6 shows a clear positive correlation in height between spouse pairs, suggesting that assortative matching on physical stature was a feature of French Canadian marriage.

The second factor is skill-specific wage differences between the destination and source, net of skill varying migration costs. If j and j_{spouse} are highly correlated, the returns to skill for men and women are positively correlated at the destination, and the wage difference between h and s is greater for skilled workers, then migrant couples should be positively selected. Rearranging Equation 12 yields

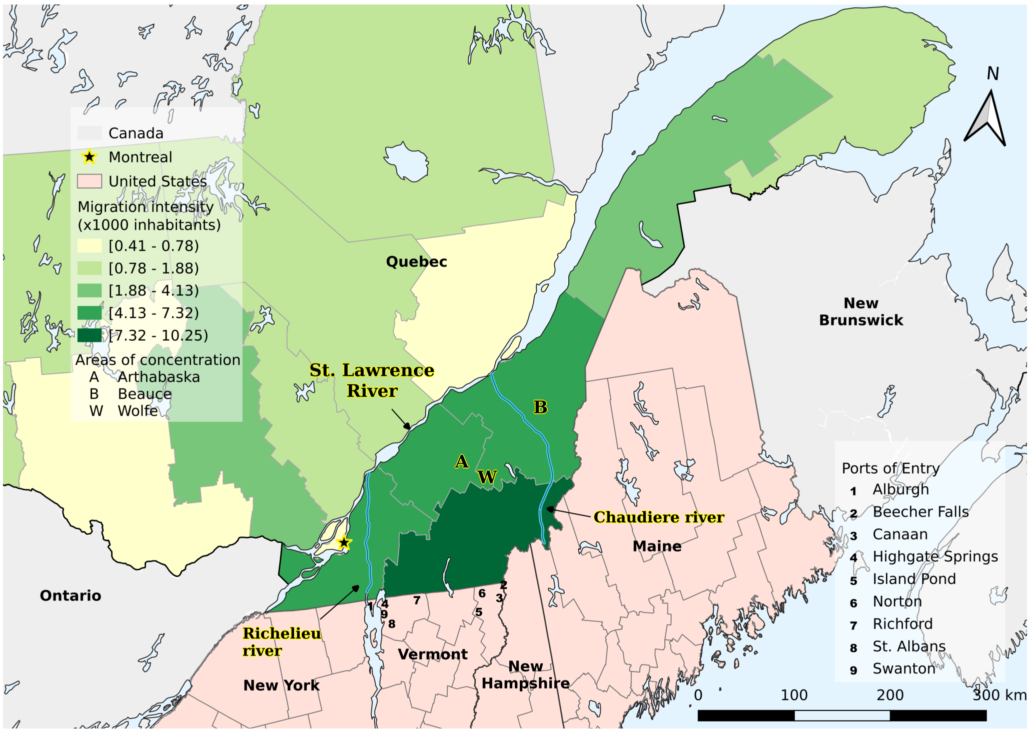
$$ln\frac{E_{sh}^{2}}{E_{sh}^{1}} = \alpha\,[(w_{h}^{2} - w_{h}^{1}) - (g_{sh}^{2} - g_{sh}^{1})] + \alpha\,[(w_{h}^{2_{spouse}} - w_{h}^{1_{spouse}}) - (g_{sh}^{2_{spouse}} - g_{sh}^{1_{spouse}})] + \tau_{s}, \tag{13}$$

where $\tau_s = ln\,(E_s^2/E_s^1) - \alpha\,(w_s^2 - w_s^1) - \alpha\,(w_s^{2_{spouse}} - w_s^{1_{spouse}})$. Equation 13 implies that destinations offering higher effective rewards to skill to both spouses should receive a higher-skilled mix of migrant couples.

Our framework considers that either partner may arrive in the destination as a tied mover, who may be better suited to an alternative location. In practice, differences between men and women in their labor market positions, particularly after marriage, would imply that married women are mostly tied migrants with limited agency in the migration decision (Mincer, 1978; Borjas and Bronars, 1991). This pattern is exactly what is observed in the early 20th century North America, where less than a quarter of adult women in Canada and the United States were employed and less than 10 percent of married women held a job (Goldin, 1986, p. 560). In such a setting, the location decision of migrant spouses would depend disproportionately on the returns faced by married men, which could significantly attenuate the correlation between observable skills of married women and destination characteristics.

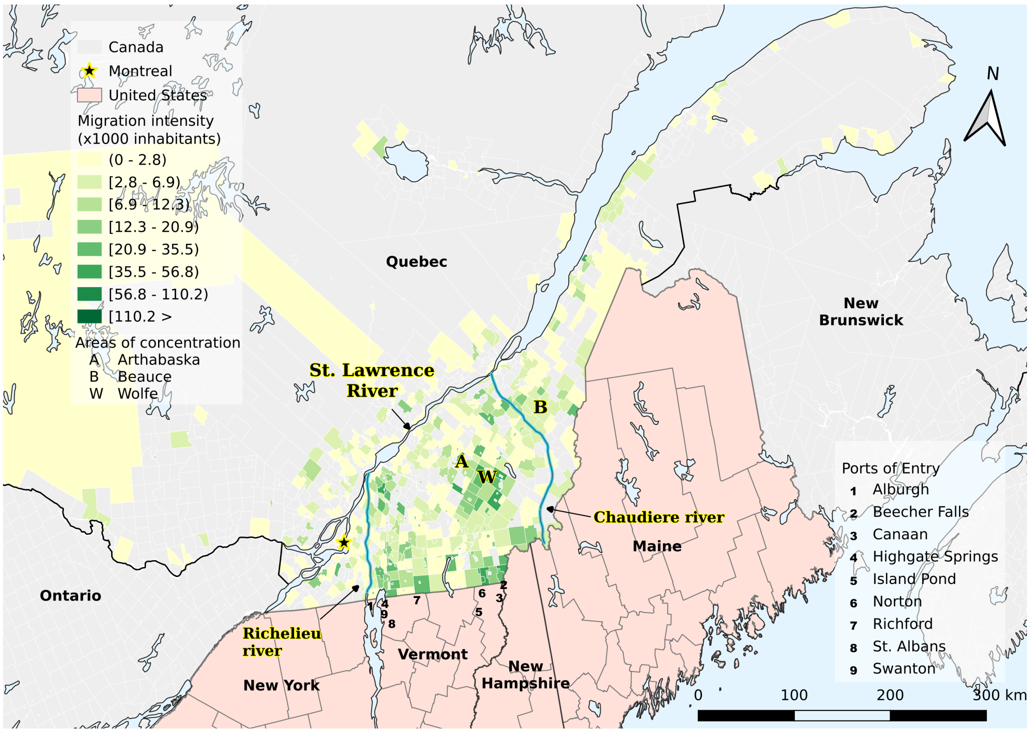
B. Additional Figures

Figure B.1: Canadian Migration to the United States, 1905-1925



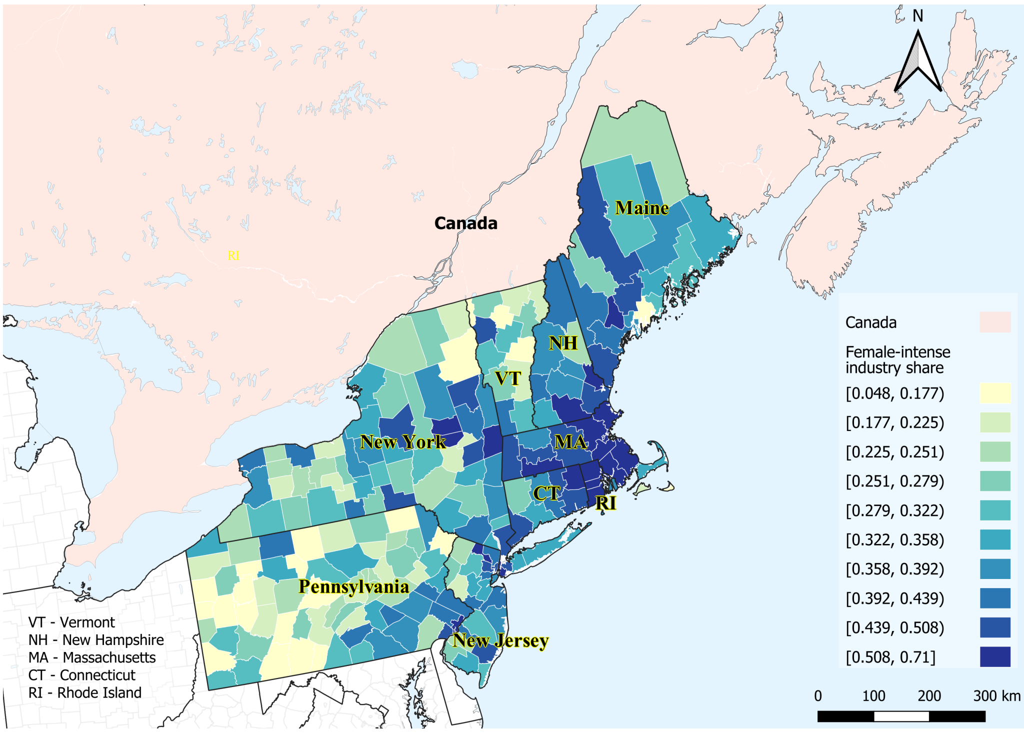
Source: St. Albans Lists, Publication number M1462.
Notes: The polygons display emigration rates per 1000 inhabitants by Quebec regions (classes determined using Jenks Natural Breaks method). We use information of sub-districts with outmigration to calculate migration intensity by region (Figure B.2).

Figure B.2: Canadian Migration to the United States, 1905-1925



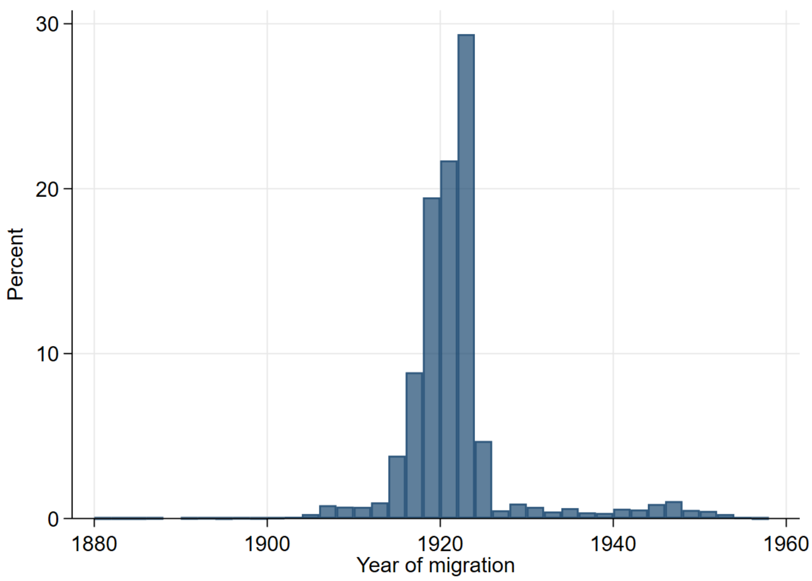
Source: St. Albans Lists, Publication number M1462.
Notes: The polygons display emigration rates per 1000 inhabitants by census sub-districts (classes determined using Jenks Natural Breaks method).

Figure B.3: Distribution of Female-Intensive Industries, 1910



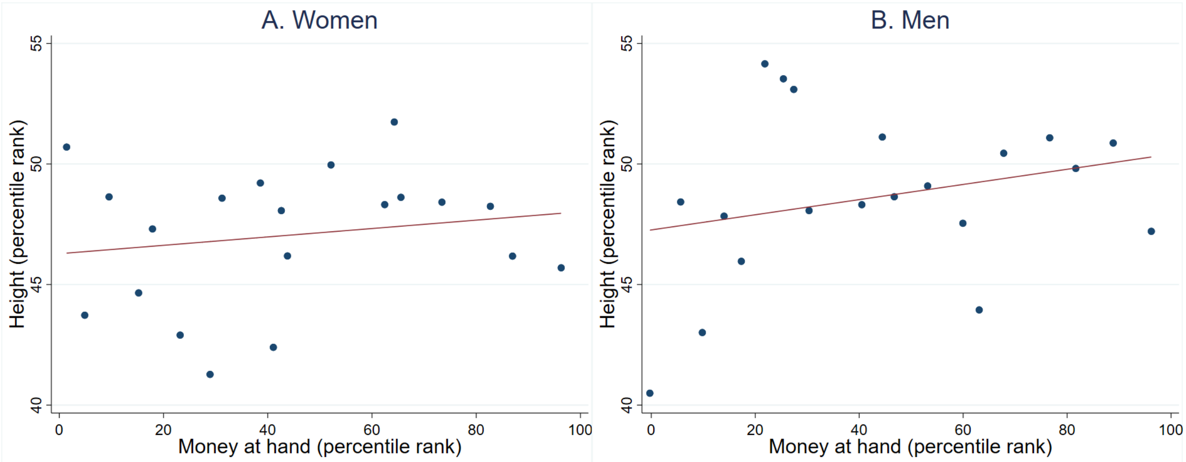
Source: Full-count 1910 US Census.
Notes: The map displays the share of female-intensive industries in the overall employment by county. Female-intensive industries are those employing more than 75 percent of the female labor force in 1910: manufacturing of non-durable goods, retail trade, personal services, and telecommunications.

Figure B.4: Distribution of Border Crossings by Birth Cohort



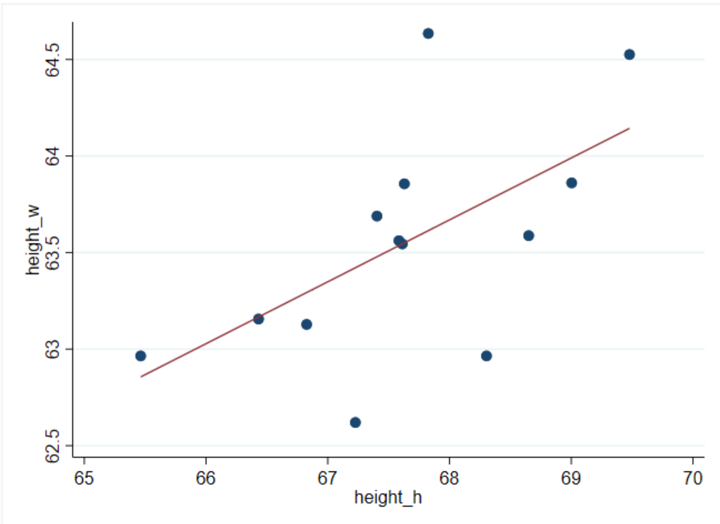
Source: St. Albans Lists, Publication number M1462.
Notes: The figure shows the distribution of the border crossing records (before refinements) over time.

Figure B.5: Height and Money in Hand



Source: St. Albans Lists, Publication number M1462.
Notes: The binned scatter plots show the correlation between height and the amount of cash held at the time of crossing, which is an indicator of savings capacity and wealth prior to emigration. Both variables are percentile ranked. The unit of analysis is individual migrants. The regressions include year-of-birth and district-of-birth fixed effects.

Figure B.6: Assortative Matching in Marriage



Source: St. Albans Lists, Publication numbers M1462, M1509, M2042, A3402, A3451, T625, T626, T627, T840; World War I Draft Registration Cards, 1917-1918; and Records of the Selective Service System, 1926–1975.
Notes: The figure shows the relationship between the height of spouses. Assortative matching on height was a feature of French Canadian married migrants. That is, relatively tall men married relatively tall women. The unit of analysis is individual migrants. The regression model includes year-of-birth and district-of-birth fixed effects. The Online Appendix (Section G) describes the process we follow to match married migrants to their spouses.

C. Additional Tables

Table C.1: Skill Premia in Canada and the United States in the Early 20th Century

	United States		Canada			
			Ontario		Quebec	
Year	*Men*	*Women*	*Men*	*Women*	*Men*	*Women*
1909	1.7	2.0	–	–	–	–
1911	–	–	1.2	1.4	1.2	1.3
1914	1.7	2.1	–	–	–	–
1919	1.2	1.7	–	–	–	–
1921	–	–	–	1.2	1.0	–

Source: Clerical/production earnings ratios for the United States adapted from Goldin and Katz (1999). Clerical/operative earnings ratios for Canadian men adapted from Inwood, MacKinnon and Minns (2010). Clerical/operative earnings ratios for Canadian women are authors' estimates from the 1911 and 1921 Census.

Table C.2: Summary Statistics, Immigrant Sample

	All Canadians		French Canadians		British Canadians	
	Men	*Women*	*Men*	*Women*	*Men*	*Women*
Panel A: Immigrants						
Age (years)	30.9	30.7	30.5	30.5	32.8	31.9
Literate (%)	95.5	96.0	94.3	96.3	98.1	97.5
Money (median, US dollars)	58.0	32.5	56.2	32.5	60.0	30.0
Occupation (%)						
professional	6.8	9.0	6.2	7.8	12.1	19.1
skilled	16.6	3.5	15.5	3.3	18.6	2.8
lower skilled	28.1	24.9	29.4	23.9	24.7	25.9
unskilled	27.6	1.4	29.4	1.5	16.0	0.2
farmers	12.0	0.1	12.8	0.1	11.2	0.1
none	8.3	60.3	6.3	62.3	16.1	51.8
Marital status (%)						
single	52.1	49.0	52.0	48.7	55.0	53.8
married	43.9	42.3	44.2	43.3	42.4	33.2
other	4.0	8.7	3.8	8.0	2.7	13.0
Networked (%)	95.2	95.0	95.7	95.0	93.6	93.4
US before (%)	63.2	61.5	60.8	59.8	75.5	63.6
Internal migration (%)	47.3	50.5	44.6	51.7	65.9	57.7
Distance (100s of km)	3.7	3.7	3.7	3.7	3.1	2.9
Height (in) by age cohort						
Average (full sample)	67.0	63.8	66.9	63.8	67.6	63.9
35-44	.	59.0	.	59.0	.	.
45-54	66.4	63.3	66.3	63.1	67.8	64.0
55-64	67.2	63.8	67.1	63.8	67.6	64.0
Observations (corridors)	274.0	219.0	254.0	202.0	90.0	68.0
Panel B: Residents						
Canadian Bulletin of Nutrition						
Height (in) by age cohort						
35-44	67.5	62.4	65.3	61.8	67.9	62.7
45-54	66.9	61.8	65.5	61.2	67.2	62.1
55-64	66.0	61.3	64.6	60.5	66.4	61.6

Source: St. Albans Lists, Publication number M1462; and Canadian Bulletin of Nutrition (Pett and Ogilvie, 1957).
Notes: Panel A reports migrant characteristics (means or shares) across corridors. Migration corridors are identified using a sample of 3,928 migrants. Panel B reports mean heights for Canadian residents (non-migrants), which we use to infer selection into migration by age-cohort.

Table C.3: Summary Statistics, Top 10 Migration Corridors

Corridor	Migrants	Employment Males	Employment Females	R Skill Males	R skill Females	Female-intensive Industry
Average corridor	10.68	0.936	0.327	0.675	0.783	0.457
	(26.495)	(0.011)	(0.070)	(0.120)	(0.372)	(0.138)
Estrie - Coos (NH)	280	0.930	0.186	0.536	1.185	0.434
Chaudiere Appalaches - Coos (NH)	157	0.930	0.186	0.536	1.185	0.434
Chaudiere Appalaches - Androscoggin (ME)	153	0.935	0.407	0.636	0.323	0.646
Chaudiere Appalaches - York (ME)	143	0.938	0.334	0.817	0.617	0.491
Centre du Quebec - York (ME)	123	0.938	0.334	0.817	0.617	0.491
Estrie - York (ME)	104	0.938	0.334	0.817	0.617	0.491
Estrie - Androscoggin (ME)	102	0.935	0.407	0.636	0.323	0.646
Estrie - Franklin (VT)	98	0.934	0.248	0.904	1.265	0.256
Monteregie - Providence (RI)	98	0.947	0.386	0.678	0.315	0.542
Bas St Laurent - Androscoggin (ME)	66	0.935	0.407	0.636	0.323	0.646

Source: St. Albans Lists, Publication number M1462; and Full-count 1910 US Census (Ruggles et al., 2021).
Notes: *R Skill* is the difference between the 80th and the 20th percentiles of the occupational income scores. *Employment* is the share of employed working-age population (16-65). *Female-intensive industries* is the labor force share in female-intensive industries (retail trade, telecommunications, manufacturing of non-durable goods, and personal services). Standard deviations in parenthesis.

Table C.4: Sorting of French Canadian Migrants - Gendered Enclaves

	(1)	(2)	(3)	(4)	(5)
Panel A: Men					
		Single		Single and Married	
				All corridors	Consistent corridors
Returns to skill	0.148*	0.215*	0.198*	0.175**	0.172*
	(0.084)	(0.114)	(0.109)	(0.080)	(0.089)
Employment probabilities	-0.196	-0.235*	-0.213	-0.130	-0.155
	(0.141)	(0.139)	(0.135)	(0.086)	(0.094)
Female-intensive industries		0.181	0.221	0.211*	0.234*
		(0.168)	(0.156)	(0.118)	(0.126)
Migration corridors	202	202	202	254	202
No. of Migrants	1,201	1,201	1,201	2,033	1,963
R-squared	0.137	0.140	0.142	0.145	0.171
Panel B: Women					
		Single		Single and Married	
				All corridors	Consistent corridors
Returns to skill	0.939**	0.819*	0.735	0.219	0.788**
	(0.437)	(0.448)	(0.520)	(0.258)	(0.386)
Employment probabilities	1.457**	1.619***	1.527**	0.408	1.160**
	(0.540)	(0.583)	(0.654)	(0.366)	(0.496)
Female-intensive industries		-0.496***	-0.428**	-0.013	-0.170
		(0.165)	(0.192)	(0.157)	(0.161)
Migration corridors	159	159	159	202	159
No. of Migrants	670	670	670	1,191	1,131
R-squared	0.279	0.292	0.294	0.293	0.335
Origin FE - region	Yes	Yes	Yes	Yes	Yes
Migration corridor controls	Yes	Yes	Yes	Yes	Yes
Destination county controls	Yes	Yes	Yes	Yes	Yes
French Canadian enclave (gendered)	No	No	Yes	Yes	Yes

Source: St. Albans Lists, Publication number M1462.
Notes: The table reports regression results controlling for gendered enclaves. That is, the enclave variable is constructed as the share of male (female) French-Canadians over all men (women) in the county. *Returns to skill* is the difference in occupational income score between the 80th and 20th percentiles. *Employment probabilities* is the working-age (16-65) population share that is employed. *Female-intensive* industries is the labor force share in female-intensive industries (retail trade, telecommunications, manufacturing of non-durable goods, and personal services). Migration corridor controls include the linear and quadratic origin-destination distance and an indicator for origin and destination locations on the border. Destination county controls include the population share living in cities (25K+ inhabitants), population share living in rural areas, mortality rate, population share of illiterate (10+ years old), population share attending school (6-14 years old), share of non-farm homes owned free, share of farm homes owned free, average value of owner-operated farms, female-to-male sex ratio, religious population share of Catholics, and population share of blacks. All continuous explanatory variables are standardized. The unit of observation is migration corridor. All regressions are weighted by corridor size (number of male or female migrants). Robust standard errors clustered by destination county. * = Significant at 10% level; ** = Significant at 5% level; *** = Significant at 1% level.

Table C.5: Sorting of French Canadian Migrants - SSIV Estimates

	(1)	(2)	(3)	(4)	(5)
Panel A: Men					
	Single			Single and Married	
				All corridors	Consistent corridors
Employment probabilities (OLS)	-0.196	-0.235*	-0.212	-0.130	-0.155
	(0.141)	(0.139)	(0.131)	(0.086)	(0.094)
Employment probabilities (SSIV)	-0.115	-0.086	0.019	-0.166	-0.161
	(0.417)	(0.356)	(0.338)	(0.258)	(0.247)
Adao-SE	[2.196]	[1.83]	[2.041]	[1.376]	[1.25]
SSIV (Reduced Form)	-0.410	-0.323	0.072	-0.637	-0.631
	(1.504)	(1.390)	(1.284)	(1.026)	(0.996)
Migration corridors	202	202	202	254	202
No. of Migrants	1,201	1,201	1,201	2,033	1,963
KP F-stat (1st stage)	5.683	8.177	8.970	8.909	9.246
Panel B: Women					
	Single			Single and Married	
				All corridors	Consistent corridors
Employment probabilities (OLS)	1.457**	1.619***	1.515**	0.391	1.125**
	(0.540)	(0.583)	(0.643)	(0.354)	(0.482)
Employment probabilities (SSIV)	1.838**	2.602***	2.668***	2.399***	2.779***
	(0.872)	(0.962)	(0.895)	(0.848)	(0.876)
Adao-SE	[1.044]*	[1.041]**	[1.131]**	[1.511]	[1.205]**
SSIV (Reduced Form)	6.489**	9.511***	10.571***	9.586***	11.205***
	(3.137)	(3.105)	(3.046)	(2.387)	(2.433)
Migration corridors	159	159	159	202	159
No. of Migrants	670	670	670	1,191	1,131
KP F-stat (1st stage)	15.77	14.66	17.65	20.82	17.68
Origin FE - region	Yes	Yes	Yes	Yes	Yes
Migration corridor controls	Yes	Yes	Yes	Yes	Yes
Destination county controls	Yes	Yes	Yes	Yes	Yes
Female-intensive industries	No	Yes	Yes	Yes	Yes
French Canadian enclave	No	No	Yes	Yes	Yes

Source: St. Albans Lists, Publication number M1462.
Notes: The table reports IV estimates. We instrument employment probabilities with the sum of workers in 1900 and the SSIV-predicted change from 1900 to 1910, divided by the working-age population in 1900. *Employment probabilities* is the working-age (16-65) population share that is employed. Migration corridor controls include the linear and quadratic origin-destination distance and an indicator for origin and destination locations on the border. Destination county controls include returns to skill, population share living in cities (25K+ inhabitants), population share living in rural areas, mortality rate, population share of illiterate (10+ years old), population share attending school (6-14 years old), share of non-farm homes owned free, share of farm homes owned free, average value of owner-operated farms, female-to-male sex ratio, religious population share of Catholics, and population share of blacks. Female-intensive industries is the labor force share in female-intensive industries (retail trade, telecommunications, manufacturing of non-durable goods, and personal services). French Canadian enclave is the population share of working-age individuals that reported French as native language and Canada as birthplace. All continuous explanatory variables are standardized. The unit of observation is migration corridor. All regressions are weighted by corridor size (number of male or female migrants). We also report the coefficient of a separate regression of the shift-share instrument on average migrant height, i.e. the reduced form. Robust standard errors clustered by destination county in parentheses. Standard errors robust to overrejection bias in brackets (Adao, Kolesár and Morales, 2019). * = Significant at 10% level; ** = Significant at 5% level; *** = Significant at 1% level.

D. Immigration Data: The St. Albans Lists

The recording of immigrants entering the United States through Canada started in 1895 and was formalized under the Immigration Act of 1903.[31] The US Immigration and Naturalization Service (INS) used immigration cards and manifests to record immigrant arrivals at the Canadian border. These documents are popularly known as the "St. Albans Lists" and were the main administrative tool to quantify the flow of immigrants from and via Canada (Ramirez and Otis, 2001, p. 190). The National Archives and Records Administration (NARA) catalogues these documents in publication numbers covering the period ca. 1895-ca. 1954.

In this paper, we use individual-level data from the publication number M1462 (6 microfilm reels), containing 41,679 immigration cards.[32] The cards are arranged alphabetically and by border posts located in Vermont, New England.[33] The immigration cards record rich demographic (age, height, literacy, marital status, nationality, occupation, race, and sex) and geographic (locality of birth, last permanent residence, and final destination) data. Immigration data include intended time to remain, previous immigration experiences, and if any, the contact of a friend or relative in the United States. These data allow us to differentiate between permanent and temporary immigrants and to identify those individuals with access to immigrant networks. We draw a random sample of the cards in each reel by selecting every fifth card to digitize.[34] The complete digitized sample comprises 8,336 individual border crossings (20% of the total cards) spanning from 1880 to 1954.

Our sample consists of border crossings at official border posts and will exclude undocumented migration. However, to track and control immigration, from 1894 the INS in agreement with Canadian railroad companies recorded all passengers destined to the United States (Smith, 2000).[35] Moreover, after 1906 immigration certificates—as proof of entry—became a requirement for all foreign-born residents applying for US naturalization, and individuals without a certificate were required to exit the United States and register at the border. Therefore, there was little reason for Canadian immigrant to avoid border posts where immigrant registration took place. Another feature of the data is that some records consist of registry cards that provide immigration information retrospectively, which

[31] The 1903 Immigration Act instructed the inspection of aliens along the borders of Canada and Mexico (US Congress, 1903, p. 1221).

[32] Publication M1462: "Alphabetical Index to Canadian Border Entries through Small Ports in Vermont, 1895-1924." The data contain records from years other than this period.

[33] The border posts by reel are Norton and Island Pond; Beecher Falls; Highgate Springs, Swanton, Alburg, and Richford; and St. Albans and Canaan. Other border posts are reported in the data, but they represent less than 1% of the sample.

[34] The starting point for the transcription was determined randomly.

[35] At train stations in Canada, INS immigrant inspectors issued certificates of admission that were required for boarding US-bound trains. The certificates were collected by another inspector at the border ports, where immigrants were registered using manifest list or immigration cards.

can be inaccurate if the registration occurred long after the arrival. However, previous research shows that yearly immigration fluctuations captured by the St. Albans Lists present a close correspondence with official US Bureau of Census data, suggesting that undocumented flows or errors in retrospective information were negligible (Ramirez and Otis, 2001, p. 192).

E. Determining race using family names

We classify all migrants as Canadian or non-Canadian based on birthplace and nationality. Additionally, we implement some refinements based on ethnicity, which allow us to identify French- or British-Canadian immigrants. Many immigrants reported their race, in which case the classification is straightforward: French corresponds to French-Canadian, and British, English, Irish, Scottish, or Welsh corresponds to British-Canadian. For Canadians not reporting their race, we use their family name to infer their ethnicity following a two-step process.

In the first step, we classify individuals with missing ethnicity as French (British) if their family name matches the family names of immigrants identifying themselves as French (British) Canadian. For observations that we are unable to classify, we infer their ethnicity using the geographic distribution of each family name, which we obtain from Geneanet Family Name Origin Database (Geneanet, 2024).

F. Classification of reported locations of birth and last residence into administrative areas

Many of the reported locations consist of two words that capture the town (area) and state of birth and last residence. We classify these locations into well-defined administrative areas as follows. First, we use the Canadian Century Research Infrastructure (CCRI) list of census subdistricts, with which we are able to classify 60% and 65% of the towns of birth and towns of last residence, respectively. Second, we use the Canadian Geographical Names Database,[36] the Quebec Toponymy Commission Name Bank,[37] and the New Brunswick Provincial Archives Database[38] to classify the remaining locations. We then compare the classified location to an overlay of CCRI historical census subdistrict polygons on OpenStreetMap, and classify the town into the subdistrict it is located in. This process increases the classified locations up to 88% for towns of last residence and up to 86% for towns of birth. Finally, we use historical (sub)district shapefiles and modern region shapefiles to classify the immigrant sources into present-day administrative regions.

This fine-grained geographic classification allows us to do three main things. First, it enables us to discriminate between migrants, return migrants, and visitors. Second, it allows us to identify migration corridors (Quebec region-US county pairs)—our unit of analysis—and control for local-level factors that may have influenced stature. Third, it enables us to estimate the distance from each individual's locality of origin to the nearest border post and to their intended destination. These distance estimates serve as a proxy for out-of-pocket transportation costs.

[36] http://www4.nrcan.gc.ca/search-place-names/search

[37] https://toponymie.gouv.qc.ca/ct/ToposWeb/recherche.aspx?avancer=oui

[38] Place Names of New Brunswick: Where is Home? New Brunswick Communities Past and Present https://archives.gnb.ca/exhibits/communities/home.aspx?culture=en-CA

G. Matching married women to their husbands

We use a 20% random sample (228 out of 1,134 observations) of married women from our immigration data. We implement the matching process as follows. First, we recover the wife's immigration card and take note of potential identifying information on the husband. Second, if the husband's first name is mentioned, we recover his immigration card based on first and last names and immigration date. In some cases, the immigration date diverges by some months, but we consider a match if the locations of last residence and destination, or the same reported contact match. Third, if the husband's first name is not mentioned, we look through all immigration cards with the same last name. We declare a match if any identifying information on the wife is found, or if the ages, dates of immigration, and locations match up. Fourth, for the remaining unmatched observations, we use Ancestry Historical Records Database to search for the husband based on the wife's reported name and geographic information. After implementing this process, we obtain 113 matches (couples), which we use to test whether assortative matching by height was a feature of French-Canadian immigrants in the early 20th century.

Oxford Economic and Social History Working Papers

is edited by

Victoria Gierok,
Nuffield College, Oxford, OX1 1NF