Explaining gender differences in the selection and sorting of migrants: Evidence from Canada-US migration

David Escamilla-Guerrero*   Miko Lepistö†   Chris Minns‡

University of St Andrews   University of Helsinki   London School of Economics

December 30, 2021

Abstract

This paper uses newly digitized border crossing records from 1906 to 1954 and US census data to study the selection and sorting of migrants from Canada to the United States. We exploit evidence on physical stature (height) to assess selection patterns in the absence of wage data. We find that Canadian women and men were positively selected on height. Our results show that sorting across destinations was notably gender-specific. Among single women, career possibilities, enclave effects, and destination marriage markets were important determinants of migrant sorting. These factors were less important for married women and unimportant for men.

Keywords: Migrant self-selection, sorting, gender, Canada, United States
JEL Codes: J61, N31, N32

---

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

†Faculty of Science, University of Helsinki, P.O. Box 64 (Gustaf Hällströmin katu 2), 00014 Helsinki, Finland. miko.lepisto@helsinki.fi.

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

1. Introduction

Despite accounting for about half of all international migration flows, few studies have assessed the migration decisions of women (Cobb-Clark, 1993; Docquier, Lowell and Marfouk, 2009; Docquier et al., 2012). How women select into migration and sort across destinations has profound implications for immigrant integration, the effectiveness of immigration policies, and the labour market outcomes of the first generation born in the destination economy through the connection between maternal skill (education) and the outcomes of their children (Rosenzweig and Wolpin, 1994; Black, Devereux and Salvanes, 2005). However, existing theoretical tools may require modification to understand the selection and sorting among female migrants. Models of migrant self-selection focus mostly on differences in the wage distributions in the source and destination economies (Borjas, 1987, 1991), which may be partially mediated by migration costs and the presence of migrant networks (Chiquiar and Hanson, 2005; McKenzie and Rapoport, 2010; Grogger and Hanson, 2011). Lower rates of labour market participation among women may imply that wage-related selection dynamics may not operate as they do for men. Barriers to employment and entry into high-skilled careers— especially in historical settings—impact labour market opportunities and migration decisions of women in ways not fully captured by measures of earnings dispersion or skill premiums (Kanaiaupuni, 2000; Goldin and Olivetti, 2013; Bertrand, Kamenica and Pan, 2015; Goldin, 2021). Particularly, gender-specific factors such as sex discrimination, marital utility, social norms, and cultural constraints can have a significant impact on the incentives to migrate and the scope for women to integrate into destination societies (Ruyssen and Salomone, 2018). Are these forces also relevant in explaining the selection and sorting of women migrants across multiple destinations?

In this paper, we combine a sample of recently digitized Canada-Vermont border crossing records from 1896 to 1924 with complete count US census data from 1920 to assess selection of Canadian migrants and how they sorted across destinations in the United States.[1] Women were a major contributor to arrivals in the United States, with almost 6 million arriving between 1900 and 1924 (Willcox, 1929, p. 396). Flows from Canada were large, with about 1.1 million Canadians present in the United States in 1920, similar to the number of immigrants from Ireland or Poland. Entry to the United States for Canadians was largely unrestricted, even after the introduction of national origin quotas for European migrants in 1921. As a result, the characteristics of migrants observed reflect self-selection due to personal characteristics

---

[1] As discussed later in the paper, we also use the 1910 complete count files for robustness checks.

and the attractiveness of different destinations rather than policy selection.[2] Flows from Canada to the United States were relatively gender balanced: 51 percent of Canadian migrants in the US 1920 Census were female, as compared to 33 percent of overseas immigrant flows between 1900 and 1924. (Truesdell, 1943, p. 49; Willcox, 1929, p. 396). There were two distinct streams of Canadian migrants to the United States. French Canadians, mostly from the province of Quebec, had high emigration rates, particularly until 1900, and mostly settled in urban areas in New England. English-speaking Canadians, who will denote as British Canadians in this paper, were drawn from all parts of Canada–mostly not Quebec, but with a significant minority from the province—and dispersed more widely across US destinations. The Vermont's crossings are dominated by French Canadian immigrants, for whom language, religion, and cultural constraints in the United States may have shaped their destination choices differently to native-English speakers migrating from other parts of Canada. Franco-American communities in New England were prominent in the early 20th century, with "Petit Canadas" replicating aspects of French Canadian life at home through church and parochial schools.

We use anthropometric evidence on stature (height) as a measure of selection. Physical stature is a function of childhood conditions (nutrition less disease environment and work assignments) that are highly correlated with human capital and earnings potential (Borrescio-Higa, Bozzoli and Droller, 2019; Komlos and Baten, 2004; Komlos and Meermann, 2007; Schultz, 2002), and has been used to study migrant selection in other contexts (Juif and Quiroga, 2019; Kosack and Ward, 2014; Spitzer and Zimran, 2018; Stolz and Baten, 2012).[3] A major advantage of using stature rather than conventional evidence such as wages or occupation rankings is that it permits us to characterise the selection patterns among all women and men, not just those who were active participants in the labour market. This is particularly important in the period we study, when few married women were in the labour force in the United States or Canada (Goldin, 1990, 2021). In addition, information on individual wages in the United States prior to the 1940s are mostly limited to a small number of state census or state labour force surveys that often exclude women or do not identify pay by gender.

Our analysis of the border crossing records reveals that Canadian women who moved to the United States were strongly positively selected. To explain sorting across destinations, we introduce a conceptual framework that extends conventional approaches used to explain selection among male migrants. Our

---

[2]Border crossing records were used to track entry for future citizenship and naturalization requests but were not used to reject prospective migrants from Canada

[3]A higher average height indicates that a population faced better conditions of health care, nutrition, and disease environment some 10 to 50 years before being observed (Schneider and Ogasawara, 2018).

framework introduces the possibility of gender-specific barriers to migration benefits through restricted employment opportunities, as well as gender-specific migration costs through social norms and marriage. We find strong evidence that these gender-specific factors can account for sorting patterns of women across destination counties in New England, New York, New Jersey, and Pennsylvania. The selection of single women was responsive to labour market possibilities and the presence of ethnic enclaves, where French Canadian norms were imported from Quebec. These destination features were only weakly connected to sorting patterns for married women, and appear unimportant for men.[4]

We contribute to the literature on immigrant self-selection by providing a more comprehensive picture of the factors that can influence the selection and sorting of women migrants. We marry this with a conceptual framework that goes beyond the most widely used theories of immigrant self-selection that seek to explain outcomes based on earnings-dispersion differences between the destination and source. Finally, since we identify the contribution of economic and social barriers faced by women in the early 20th century as important forces in shaping the sorting of female migrants across destinations in the United States, our findings can be connected to recent research exploring the relationship between migration and gender discrimination (Ruyssen and Salomone, 2018).

2.   Emigration from Canada in the Age of Mass Migration

About 30 million immigrants moved to the United States during the Age of Mass Migration. Although most research during the Age of Mass Migration has focused on trans-Atlantic flows, significant migrations also took place across US land borders (Abramitzky and Boustan, 2017; Hatton and Williamson, 1998). Canada was one of the leading immigrant source countries, with over one million Canadians present in the United States in 1920.[5] Both British and French Canadians crossed the border in large numbers, with French Canadians accounting for approximately 30 percent of all Canadian immigrants during the early twentieth century (Ramirez and Otis, 2001). High emigration rates in Quebec meant that about 20 percent of all French Canadians resided in the United States at the beginning of the 20th century. 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 French or British ancestry.[6]

---

[4]In related work we find strong evidence of assortative matching on height for married French Canadian spouses. These results are available on request

[5]Mexican immigration to the United States also increased gradually from the 1880s (Gratton and Merchant, 2015). During the 1910s, about 350 to 700 thousand Mexicans fled across the border due to the Mexican Revolution, and 80 thousand more crossed the border under the First Bracero Program (1917-1921) (Escamilla-Guerrero, Kosack and Ward, 2021).

[6]At least until 1900, Canadian gross out-migration to the United States completely offset European gross immigration to Canada (McInnis, 1994).

French Canadians were on average poorer and less well educated than their Anglophone counterparts. Quebecois francophones in the 1911 census were about 30 percent less likely to hold a professional or clerical occupation than Ontario Anglophones, and almost 70 percent more likely to be labourers. British Canadians in Ontario also had average earnings 23 percent higher than Quebec francophones (Inwood, MacKinnon and Minns, 2010). Evidence on physical stature shows that these differences in earnings were accompanied by noticeable differences in height between British and French Canadians (Cranfield and Inwood, 2007; Morin, Geloso and Kufenko, 2017). Intergenerational mobility, at least for those who remained in Canada, was also lower for French Canadians (Antonie et al., 2021). Lower literacy rates and less fluency in English also made migration to industrial towns and cities in New England particularly attractive for French Canadians relative to alternative destinations, as French-speaking enclaves were well-established in this region by 1900 (Green, MacKinnon and Minns, 2005).

The conventional view among Canadian economic historians is that persistent emigration from Canada to the United States reflected differences in economic opportunities between the two countries. This argument is based mostly on labour market data for men in both countries. Income per capita was significantly higher in the United States than in Canada in the early 20th century, and in most occupations real wages for men were 5 to 20 percent higher in the United States (Green, MacKinnon and Minns, 2002; Bolt and van Zanden, 2020). Comparisons of men's wages between the two countries suggest that skill premiums were also larger in the United States. Table 1 reports the earnings ratio of clerical to production workers as best available proxy for earnings dispersion by skill.[7] The figures show that skill premia in the United States were higher than in Ontario or Quebec in the first decades of the twentieth century, with gaps closing only after the First World War. Although 51 percent of Canadian immigrants in the US 1920 Census were female—as compared to 33 percent of overseas immigrant flows between 1900 and 1924—the factors that influenced the emigration of Canadian women have been mostly ignored by historians (Truesdell, 1943, p. 49; Willcox, 1929, p. 396). Wages appear to have been higher in the United States for women, with a skill premium larger to that enjoyed by men in the United States (see Table 1), data on labour force participation show that opportunities for women were more constrained in Canada, with participation rates of adult women about 6 percentage points lower than in the United States (17.7 v. 23.7). Research on French-Canadian emigration to the United States has emphasized the role of secondary earners and family migration in cross-border flows, as opportunities for women to work in New England offered insurance to households in case of unemployment of male heads (Ramirez, 1986).

---

[7]The US censuses did not include income information until 1940, and there are relatively few sources of alternative earnings or wage data disaggregated by occupation or gender.

Emigration data, however, show that the number of unaccompanied women was rising through the first half of the 20th century, as opportunities in Canada for independent women were even fewer than in other parts of North America (Ramirez and Otis, 2001; Waldron, 2005).

3.   A conceptual framework for migrant sorting by gender

Models of immigrant self-selection connect the differences in the types of immigrants arriving in destinations to how well those economies reward skill. The classic example is the Borjas (1987) model under which differences in earnings dispersion between source and destination economies determine migrant selection. Later studies emphasize the role of migration costs that may differ by skill type in complicating patterns of selection (Chiquiar and Hanson, 2005; McKenzie and Rapoport, 2010). In addition, Grogger and Hanson (2011) show how the sorting of immigrants across destinations may be better predicted through income maximization—and therefore the absolute gains from migration— rather than the relative returns to skill afforded by different locations. To motivate our analysis we use a simplified version of the Grogger and Hanson (2011) model to illustrate how gender-specific factors can matter for selection and sorting.

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

$$c_{ish}^{j} = f_{sh} + g_{sh}^{j},\tag{1}$$

where c_{ish}^{j} is the cost of migrating from source s to destination h for individual i with skill level $j = \{1\,(unskilled), 2\,(skilled)\}$.[8] 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

$$U_{ish}^{j} = \alpha\,(w_{ih}^{j} - c_{ish}^{j}) + \epsilon_{ish}^{j},\tag{2}$$

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

$$ln\,\frac{E_{sh}^{j}}{E_{s}^{j}} = \alpha\,(w_{h}^{j} - w_{s}^{j}) - \alpha\,f_{sh} - \alpha\,g_{sh}^{j},\tag{3}$$

---

[8]We assume 2 skill groups for simplicity.

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

$$ln\,\frac{E_{sh}^{2}}{E_{sh}^{1}} - ln\,\frac{E_{s}^{2}}{E_{s}^{1}} = \alpha\left[(w_h^2 - w_s^2 - g_{sh}^2) - (w_h^1 - w_s^1 - g_{sh}^1)\right]. \tag{4}$$

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 wage differences for the two skill groups and the size of skill-varying migration costs. Rearranging the above equation yields

$$ln\,\frac{E_{sh}^{2}}{E_{sh}^{1}} = \alpha\left(w_h^2 - w_h^1\right) - \alpha\left(g_{sh}^2 - g_{sh}^1\right) + ln\,\frac{E_{s}^{2}}{E_{s}^{1}} - \alpha\left(w_s^2 - w_s^1\right), \tag{5}$$

where the first two terms of the right-hand side capture the rewards to skill (net of migration cost) that explain the intensity of sorting. Hence, destinations offering higher net rewards to skill should receive a higher-skilled mix of emigrants from source s. To identify the role of gender-specific factors that affect the net rewards to skill, Equation 5 can be written as

$$ln\,\frac{E_{sh}^{2,k}}{E_{sh}^{1,k}} = \alpha\left(w_h^{2,k} - w_h^{1,k}\right) - \alpha\left(g_{sh}^{2,k} - g_{sh}^{1,k}\right) + ln\,\frac{E_{s}^{2,k}}{E_{s}^{1,k}} - \alpha\left(w_s^{2,k} - w_s^{1,k}\right), \tag{6}$$

where $k = \{1\,(female), 2\,(male)\}$. The implication of Equation 6 is that under complete gender parity the skill mix of emigrants received by the different destinations should be gender balanced, as men and women should face the same skill-based returns to migration $(w_h^{2,1} - w_h^{1,1}) - (w_h^{2,2} - w_h^{1,2}) = 0$ and skill-related migration costs $(g_{sh}^{2,1} - g_{sh}^{1,1}) - (g_{sh}^{2,2} - g_{sh}^{1,2}) = 0$. A priori there are plausible reasons why both conditions may not be true. Next, we address three gender-specific factors that can induce differences in the sorting of men and women migrants.

3.1 Social capital: immigrant networks and ethnic enclaves

The notion that networks reduce liquidity constraints for future migrants has a long lineage in research on the economics of migration (McKenzie and Rapoport, 2007, 2010). Although all migrants draw upon networks in the destination to reduce the costs of integrating into their new place of residence and to make migration feasible in the first place, female migrants have more to gain from social capital due to differences in the ability to save between men and women (Curran and Rivero-Fuentes, 2003; Munshi, 2003). Similarly, the access to ethnic enclaves is particularly attractive to the less-skilled, as ethnic

connections reduce information and job search costs for newly arrived immigrants (Damm, 2009; Edin, Fredriksson and Åslund, 2003). Contemporary research suggests that social capital may have gendered impacts, with female networks strongly influencing the destination choice of women migrants (Davis and Winters, 2001).

In the context of emigration from early twentieth-century French Canada, language played a clear role in generating skill-biased migration costs. More skilled migrants with more education and better childhood conditions were more likely to speak English (and to speak English better) than their less-skilled counterparts. Hence, the impact of French-Canadian enclaves in mitigating migration costs may have been greater for women than for men, as migration costs for independent women were much higher if they were unable to reduce them either through skill or language, or through migration to a French Canadian enclave where the co-ethnic community provided a structure that eased integration. Table 2 shows an important variation in the share of French Canadians in the destination counties, with a range from less than 5 percent (Essex County MA) to 19 percent (Androscoggin County ME). If the short-run returns to moving to an enclave increase with enclave quality (size)—as suggested by contemporary research (Damm, 2009)—larger French-Canadian communities should have attracted a higher share of unskilled, single, French-Canadian women.

3.2   Employment potentials and career opportunities

Women's ability to work and to work in a full range of occupations plays a large role in shaping the net gains associated with migration. These extensive margins have been identified as an important feature for the migration decisions of contemporary women, as low female labor force participation rates in the destination can depress the expected returns to women's human capital (Aksoy and Poutvaara, 2021; Bertoli, Moraga and Ortega, 2013; Sandell, 1977). A similar mechanism should apply to North American labour markets in the early 20th century, due to the spatial variation in the demand for skilled women, the application of barriers to employment such as marriage bars, and social norms about the role of women in the workplace. During this period, less than a quarter of adult women in Canada and the United States were employed, with less than 10 percent of married women holding a job (Goldin, 1986, p. 560). Marriage bars and internal labour market policies among employers limited promotion and access to skilled positions for women, while industries varied in the extent to which women could access to skilled and semi-skilled positions on the factory floor.[9] The variation in employment potential for women is

---

[9]Marriage bars were policies adopted by firms and local school boards in the United States to fire single women when they married and not to hire married women (Goldin, 1988).

evident if we look at the leading destination counties for French Canadian migrants identified in the border crossing records, with a range from less than 20 percent (Coos County NH) to almost 45 percent (Hillsborough County NH) (see Table 2).

3.3 Marriage

Another dimension that can influence destination choices differently by gender relates to marriage. Men and women who move to destinations with unfavourable gender ratios will experience an increase in the cost of finding a spouse due to the worsening of their bargaining position in the marriage market (Bhaskar, 2019). If less skilled individuals are less attractive in marriage markets due to lower earnings, lower social status, or less wealthy parents, they may experience a disproportionate increase in the cost of finding a spouse when partners are scarce. If marriage related costs are relevant to the sorting of single migrants, we would expect symmetrical effects for both genders, with unskilled single men (women) choosing locations where men (women) are scarce and their cost for finding a partner is low. Predictions on the sorting of single women migrants across destinations will depend on how marital utility varies across skill groups. If marital utility decreases with human capital, destinations with environments favoring the employment and careers of women should receive a higher-skilled mix of single, female migrants.

In the early 20th century marriage was a crucial decision for many women. Social norms that limited women's participation in the labor force and the lack of highly effective, female-controlled contraception pushed women to choose between employment and family (Bailey, 2006; Goldin, 2021). In this context single, French-Canadian women with high employment or career prospects had the opportunity to postpone or forgo marriage in favor of financial independence and greater freedoms in the United States (Waldron, 2005). It is also important to note that marriage may play a significant role in dampening the forces driving selection and sorting among women if a large share of female migrants were "tied movers" for whom location decisions mostly reflected the returns experienced by male partners attached to the labour market (Mincer, 1978; Borjas and Bronars, 1991).

4. Data

The recording of immigrants entering the United States through Canada started in 1895 and was formalized under the Immigration Act of 1903.[10] The US Immigration and Naturalization Service (INS) used

---

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

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. In this paper, we use individual-level data from the publication number M1462 (6 microfilm reels), containing 41,679 immigration cards.[11] The cards are arranged alphabetically and by border posts located in Vermont, New England.[12] 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.[13] The complete digitized sample comprises 8,336 individual border crossings (20 percent of the total cards) spanning from 1880 to 1954.

Our migrant sample consists of border crossings at official border posts and does not capture 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).[14] 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 a 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 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).

---

[11]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.

[12]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.

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

[14]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.

In the analysis that follows, we limit the migrant sample to border crossings that took place after 1906, as only non-Canadian immigrants were supposed to be recorded at the border previously (Smith, 2000). We also restrict both samples to individuals reporting complete geographic data, as these data allow us to discriminate between immigrants, return immigrants, and visitors. We then retain migrants reporting their complete name (given name and family name) and race. Finally, we limit the sample to individuals who had passed their pubertal growth spurt before being observed (crossing the border or being enlisted): males aged 16-65 years and females aged 14-65 years. This refinement avoids capturing growing and shrinkage effects, which can distort human health capital estimates based on human stature (Spitzer and Zimran, 2018). Following these restrictions, our final sample contains 6,644 immigrants (2,661 females and 3,983 males). To analyse the role of destination characteristics in selection and sorting, we focus on counties in New England, New York, and Pennsylvania.

Most of the border crossing records include a reported "race" for each individual, which we use to classify migrants as either French or British Canadian using the *Dictionary of Races or Peoples* (Folkmar, 1911). Note that we classify British Canadians as those indicating ethnic origins that were English, Irish, Scotch, or Welsh. For records where the race entry is incomplete, we use the family name to assign ethnicity. In the Online Appendix, we provide a full description of the methodology that we follow to identify French and British surnames in the migrant sample. We also classify the localities of birth and last residence into census districts and sub-districts following St-Hilaire et al. (2007). This fine-grained geographic classification will allow us to control for unobserved factors that influenced height at the regional and local level. By classifying the geographic data into sub-districts, we are also able to estimate the distance from the localities of origin to the nearest border post and to the intended destination for each individual.[15] These estimates proxy for out-of-pocket transportation costs, which were fairly small relative to income and increased only modestly with distance within Canada.[16]

4.1 Summary Statistics

In Figure 1 we map the origin of Canadian emigrants by 1921 census sub-district. The map reveals some clear emigration patterns. First, distance matters, with sub-districts close to the border—and especially south of the St Lawrence River—having higher emigration rates. Second, within Southern Quebec clusters are visible along the Vermont–Quebec border, north of Montreal on the eastern side of

---

[15]Note that the migrant's final destination is uncertain.

[16]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 labourer at the time Green, MacKinnon and Minns (2005).

the Richelieu River, in Arthabaska and Wolfe, and in Beauce near the Chaudière River. These clusters are either close to the border, or near waterways and railroads that provided access to Vermont. Arthabaska and Beauce were areas of intermediate population density, with many small towns and villages that may have had surplus labour but were also not well connected to Montreal, the economic centre of Quebec and Canada at the time. We also classify the place of origin into French or British sub-districts depending on which ethnicity was most prominent. This analysis shows that French Canadians were more clustered in a small number of sub-districts in Southern Quebec, whereas British Canadians were more dispersed, with many migrants arriving from places in Ontario and New Brunswick (see Figure 2).[17] In Figure 3 we trace migrants from their Canadian origins to their destinations. The figure highlights the relative concentration of destinations in New England (Maine, Vermont, New Hampshire, Massachusetts, Connecticut, Rhode Island) with a few migrants heading further afield to Pennsylvania, New York, and New Jersey. Figure 3 also makes clear that many of the migrants in our sample settled in locations relatively close to the Canada-US border, where migration costs were lower and regular information was likely available on the Canadian side.

In Table 2 we report the main characteristics for the top 10 French Canadian destinations in the United States. We observe substantial variation among leading destination characteristics, with French Canadian population shares (our measure of enclave size) ranging from five to almost twenty percent. While the vast majority of adult men (92 percent) were employed in all locations, the women's labor force participation rate was on average 28 percent. Female participation rates also varied across destination choices, with some counties having more than 40 percent of their women residents employed. In contrast, there was modest variation in single adult sex ratios across counties, where Coos County NH stands out as a major outlier. On average, the share of skilled occupations in the destination (our proxy for career opportunities) was 18 percent, less than half of the unskilled occupations. Note that the figures combine men and women to compute the skilled and unskilled shares. In Table A.1 and Table A.2 we present separate ratios by sex and for women by marital status, respectively. Interestingly, the share of skilled occupations held by women is roughly twice that of men. Comparing the shares of skilled and unskilled occupations across genders, reveals that women moved to locations with more even occupational distributions. Similar patterns are observed between single and married women, as destination choices of the latter mostly reflect choices of males partners as argued before.

---

[17]We also present the spatial distribution emigration rates for men and women in the Online Appendix.

The border crossings data provide a host of socioeconomic characteristics that may impact the decision to migrate. Panel A of Table 3 shows that migrants were relatively young when observed (age 30-33). However, more than 60% have had other immigration experiences in the United States previously, suggesting that first-time immigrants were younger. On average, Canadians crossing the border for the first time were 26 years old, the same as the average age of overseas immigrant arrivals to the United States between 1900 and 1910 (Hatton and Ward, 2018). The vast majority of migrants were literate, as one would expect given well-established primary schooling in Canada from the 19th century. The median amount of cash carried by both British and French Canadian immigrants was between 30 and 50 US dollars, which is equivalent to two to three times the typical weekly earnings of labourers in 1921.[18] However, migration costs varied substantially between ethnic groups: British Canadians were on average twice as far from the closest Vermont border post than their Francophone counterparts (340km v. 170km).

Male migrants covered the breadth of the occupational spectrum, with relatively few farmers compared to European source countries. The occupational mix among British Canadians was more heterogeneous, while most French-Canadian migrants had semi-skilled occupations or were unskilled labourers. The majority of migrant women did not report an occupation, but the share who did (about 35%) is considerably higher than female labour force participation rates in Canada or the United States prior to 1940 (Goldin, 1990). Both male and female migrants were fairly well balanced between married and single individuals. Most migrants (more than 80%) reported having a contact in the United States and about 60% had previous migration experience, as one would expect in a population with relatively fluid access to migrant networks in the destination. About 42% of migrants with both their birth and residence places in Canada reported a town of residence differing from their birthplace, implying they had moved at least once within Canada before crossing the border.

How did migrants compared with those left behind? The bottom of Table 3 summarizes the stature of migrants, in comparison to cohort-specific mean heights for men and women observed in 1953 reported in the Canadian Bulletin of Nutrition (Pett and Ogilvie, 1957). These estimates are based on a sample of 22 thousand Canadians, who were examined by trained enumerators of the Department of National Health and Welfare. One caveat is that the data is reported for two population cohorts: residents within Quebec and residents outside Quebec. This division may underestimate height differences between French and British Canadians due to the presence of English-speaking Canadians in Quebec and French-speaking

---

[18]The median annual labour wage in 1921 was $850 in Ontario and $800 in Quebec. See Table 2 in MacKinnon (1996) for weekly earnings across occupations. Only 2% of migrants reported carrying no cash.

Canadians outside Quebec. However, it is the only data source that allow us to make inferences about the self-selection on height of both men and women migrants.

In Figure 4 we plot adjusted cohort heights for migrants and mean heights for residents reported in the CBN. We estimate the adjusted values regressing height on a full factorial structure for ethnicity, year-of-birth cohort, and sex. The model also includes an indicator variable for individuals who migrated after the enactment of the 1921 Emergency Quota Act, an interaction between this variable and sex, and district-of-birth fixed effects. We cluster standard errors at the year-of-birth level. This procedure reveals differences across cohorts and gender. For male migrants, we find positive self-selection for cohorts born in the late 19th century: a 1 inch height gap in favor of male migrants born in 1899-1908 (aged 45-54). However, we find evidence of neutral selection for younger cohorts of male migrants. For female migrants, we find evidence of greater positive selection than for males for the same cohort: a 2 inches height gap in favor of women migrants aged 45-54. In contrast to men, later cohorts of women migrants were positively selected, though estimates for these groups are less precise due to sample size.

In the Appendix we present selection estimates for male migrants using residents micro data: military records of volunteer soldiers who enlisted in the Canadian Expeditionary Force (CEF) during the First World War (see Table A.4).[19] Height differences between migrants and residents are very similar to those that can be inferred using the CBN, suggesting that inferences on the selection of female migrants are also informative.[20] Overall, these preliminary results show that the Canadian migrants were positively selected, but despite limited interaction with labour markets women were more positively selected than men.

5. Empirical approach

Our empirical approach evaluates how Canadian migrants sorted across US destinations, and whether gender-specific factors at the destination can account for differences in the sorting of women migrants. We focus on the three factors identified in our theoretical framework: the effects of employment or career possibilities for women in destination labour markets, the presence of enclaves that provide ethnic networks and environments with group-specific social norms, and the structure of marriage markets that

---

[19]The data come from Cranfield and Inwood (2015).

[20]The occupational distribution of male migrants also suggest an intermediate self-selection on skills, as the share of lower skilled and skilled migrants is about 10 percentage points higher than that of military soldiers, while the share of professional workers and farmers is higher among enlistees (see Table A.3).

affect partner allocations and marital utility. Our estimating equations take the form:

$$h_{iydc} = \alpha + \eta_y + \theta_d + \gamma_1 \, enclave_{ic} + \mathbf{X}'_{iydc}\mathbf{B} + \mathbf{Z}'_{ic}\mathbf{\Gamma} + e_{iydc}, \tag{7}$$

$$h_{iydc} = \alpha + \eta_y + \theta_d + \delta_1 \, career_{ic} + \delta_2 \, employment_{ic} + \mathbf{X}'_{iydc}\mathbf{B} + \mathbf{Z}'_{ic}\mathbf{\Gamma} + e_{iydc}, \tag{8}$$

$$h_{iydc} = \alpha + \eta_y + \theta_d + \phi_1 \, marriage_{ic} + \mathbf{X}'_{iydc}\mathbf{B} + \mathbf{Z}'_{ic}\mathbf{\Gamma} + e_{iydc}, \tag{9}$$

$$h_{iydc} = \alpha + \eta_y + \theta_d + \lambda_1 \, enclave_{ic} + \lambda_2 \, career_{ic} + \lambda_3 \, marriage_{ic} + \mathbf{X}'_{iydc}\mathbf{B}$$
$$+ \mathbf{Z}'_{ic}\mathbf{\Gamma} + e_{iydc}. \tag{10}$$

In all models h_{iydc} is the height of migrant i born in year y in Canadian census district d and resident after migration in US county c. The models include a common set of control variables, $\mathbf{X}'_{iydc}$, derived from the border crossings data. These variables are individual characteristics that are plausibly relevant for the migration decision through their connection to the costs and benefits of migration. These consist of indicator variables for whether the migrant has a personal contact in the United States, whether the migrant moved within Canada prior to entering the United States (last place of residence not equal to place of birth), whether the migrant had previous immigration experience in the United States, and whether the migrant entered the United States after the introduction of the national origin quotas in 1921.[21]

We also include a set of control variables, $\mathbf{Z}'_{ic}$, that vary by the migrant's destination choice. These variables are the source-to-destination linear distance, which proxies for out-of-pocket transportation costs, and the distance squared, which captures nonlinear effects of out-of-pocket costs on the destination choice (Hatton and Williamson, 1998). Although Table 3 shows that both men and women migrants travelled very similar distances, if out-of-pocket transportation costs rise with distance, differences in capital and ability to save between men and women may lead to different sorting patterns across destinations. In Table A.5 we present correlation coefficients between these controls and our outcome variable.[22] In addition, we control for year-of-birth, η_y, and district-of-birth, θ_d, fixed effects. They

---

[21]While Canadians were not directly affected by the 1921 Emergency Quota Act, they had a substantial impact on immigration flows from Europe, and led to changes in the demand for Canadian labour in the United States (Abramitzky et al., 2019).

[22]We estimate these correlation coefficients using different distance calculations. The estimates do not change materially in terms of magnitude or significance.

capture the effect of any shock affecting the height of specific age cohorts, or of any time-invariant, district characteristic that may explain height differences across geographic areas.

The key variables of interest in our study are county-specific measures for US destinations constructed from 1920 full count census data, which we summarize in Table 2. These location characteristics are factors we identify as having potential direct effects on migrant sorting across destinations. In Equation 7 we model the role of ethnic enclaves with a variable that measures the share of French Canadians by destination county. If ethnic enclaves reduce information and search costs for migrants, then locations with a larger share of French Canadians would be more attractive to relatively disadvantaged migrants. This pattern has been identified in studies examining the effects of enclaves in both historical and contemporary settings (Damm, 2009; Eriksson, 2020; Edin, Fredriksson and Åslund, 2003).[23] However, we know little about how sorting into enclaves varies by gender. In our preferred specification $enclave_{ic}$ is a categorical variable capturing the migrant's destination choice by French-Canadian enclave size. We define French-Canadian enclaves as counties with at least 5% of French-Canadians in 1920: individuals reporting French as mother tongue and Canada as birthplace. Our coefficient of interest, δ_1, captures differences in height between immigrants who moved to an enclave and those who settled in a county with relatively small concentrations of French-Canadian immigrants.[24]

In Equation 8 we include two measures of destination labour market conditions that capture the extent of opportunities for women. First, we calculate labour force participation rates as a share of the adult working age population who were employed, $employment_{ic}$. Due to the lower participation rates of women in the labor markets, destinations with more available jobs for women may have attracted women from all segments of the skill distribution, as for most women their main priority was to secure a job. Hence, predictions for the sorting of women migrants are ambiguous when considering employment opportunities only. Our second measure, the availability of higher skill positions, $career_{ic}$, provides a clearer prediction, as such opportunities were available only to women from the top end of the distribution.

In Equation 9 we model gender imbalance in the destination as a factor influencing sorting across destinations. This is aligned with the notion that marriage markets differed across possible destinations. As discussed earlier, a model of bargaining in marriage implies that in locations where men were relatively abundant, women enjoyed better bargaining positions in finding a spouse. As a result, marriage costs

---

[23]The effect of enclaves on immigrant outcomes, however, is less clear, with negative effects and positive effects varying across periods.

[24]The size of immigrant enclaves varied by source country, but in 1920 the average county had 12% foreign born residents (Eriksson, 2020).

for women were lower in such locations, shifting the threshold at which marriage occurs and enabling the migration of the less positively selected if prospective marriage was desirable. The coefficient ϕ_1 provides the association between gender imbalance and migrant height. In our preferred specifications we use full count data from the 1920 census to compute county-level female to male sex ratios for single women and men aged 16 to 40 years.

In Equation 10 we combine all of our county-level factors determining the sorting of migrants in a single model. All regression models are estimated using ordinary least squares (OLS) and clustering standard errors at the province level. While our focus is on understanding selection and sorting among women, to properly evaluate whether there were distinctive gender patterns we estimate all models for both women and men. In these models we control for the individual's ethnicity (British or French) and marital status (single or married). We also estimate models for all Canadians and French Canadians only and for single French Canadian men and women separately. Comparing French Canadians to all Canadians gives a sense of whether sorting patterns were different for populations that had a more distinctive profile in terms of language, religion, and culture. Comparing married and single female migrants speaks to whether sorting intensity was constrained for women who were "tied movers" attached to a male spouse who was in almost all cases the family's primary earner.

6. Results

In Table 4 we present estimates capturing the importance of French Canadian enclaves in shaping the sorting of women migrants. We explore several possible cutoffs for enclave size, as informed by the current literature and French Canadian residential patterns in New England, New York, and Pennsylvania when the migration we observe took place. Columns 1 to 3 show results for all Canadian women, French Canadian women only, and single French Canadian women, respectively. The results reveal a strong negative height sorting into enclaves that increases with enclave size. Francophone women moving to locations where at least 15 percent of the population were French Canadian were 0.8 to 1.1 inches shorter than women in locations with few co-ethnics (columns 2 and 3). If we include all Canadian women (column 1), the effect of French Canadian enclaves falls to 0.6 inches. This is consistent with co-residence mattering for French Canadians but not British Canadians, as one would expect (see Figure 5). In column 4 we provide comparable estimates for Canadian men, then limit our migrant sample to French Canadians in column 5. We find little evidence of height sorting into enclaves for male migrants. All estimated coefficients are close to zero and mostly statistically insignificant. While 1920 is the year that falls closely

to the median year of observation (1921) in the migrants sample, we observe migrants arriving before 1920 when conditions may have been different, and their presence in 1920 may contribute to the enclave measure itself. In alternative specifications we use French Canadian shares calculated from the 1910 full count census, which precedes the arrival of the vast bulk of the migrants. As we show in Panel A of Table A.6 this has little effect on our results. However, when we use a broader definition of enclave, which includes second-generation French-Canadian immigrants, our results lose statistical significance. This suggests that only direct co-ethnic networks may have influenced women's sorting into enclaves.

What do the striking differences between women and men tell about about the gendered nature of migrant sorting? One view of enclaves is that they are locations where co-ethnic networks provide information about employment, access to social capital, and opportunities for migrants without host-country specific human capital (such as language ability). That enclave effects are gendered suggests that social norms related to migration and independence constrained the choices of most French Canadian women. While men across the skill distribution, as implied by height, were relatively unconstrained by the absence of Francophone communities in the destination, only the tallest women, who experienced the best childhood conditions, likely had more educated, prosperous (and perhaps enlightened parents) were able to settle away from Franco American communities.

In Table 5 we present estimates of the effects of labour market opportunities on the sorting or migrants. As mentioned before, we use the share of employment in high skilled occupations as a measure of the abundance of opportunities for women seeking a career not just a job in the United States, and who were drawn from the upper tail of the skill distribution. The results in columns 1 to 4 show that greater availability of skilled occupations drew more positively selected women to destinations: positive height sorting into career potential. As with evidence on enclaves, the results for men show that this is a gender-specific effect, with no apparent connection between the labour force participation and skill shares in destinations for men and sorting on height. Participation rates at the destination appear to have only a limited association with height sorting for either men or women. In Panel B we present results using occupational shares computed from the 1910 full count US census. The coefficients are very similar in terms of magnitude and statistical significance.

In Table 6 we present estimates of the effect gender imbalance in the destination on sorting patterns. We compute female to male population ratios for different age ranges (16-65 and 16-40 years), as well as those aged 16 to 40 who were single: our preferred indicator of the state of marriage markets with respect

to gender imbalance. In Panel A we show how gender imbalance in 1920 relates to the height of migrants. The main finding is that the effects of marriage markets appear to be statistically significant but close to zero. The sign on the gender imbalance variable for single French-Canadian women is also inconsistent with our theoretical framework (column 4), with less skilled women moving to locations where females are more abundant relative to males.For men the sign is consistent with those predictions, but in any case both coefficients are too small to be of economic significance. In Panel B we present results using gender imbalance figures computed from the 1910 full count US census. The coefficients are very similar in terms of magnitude but present less statistical significance.

Table 7 brings together the three strands of our theoretical framework into a single econometric model. The results confirm that the forces we identify behind the sorting of female migrants operated simultaneously, with signs, magnitudes, and significance broadly similar to the estimates presented above. For women the large effects of enclave addressed before are present, with a similar gradient for both married and single French Canadians. A greater gender imbalance (more women than men) has a clearer association with sorting of positively selected women—as predicted by our theoretical framework—once we adjust for other country-level determinants of sorting. Participation opportunities were attractive for skilled women only among those who were single and therefore likely to engage in paid work. Comparing columns 4 and 5 more generally highlights the differences between single women—for whom sorting appears highly responsive to marriage markets, career opportunities, enclaves effects—and married women for whom the correlations are typically smaller in magnitude and less precisely estimated. Comparing women (and especially single women) to men confirms that the forces we observe shaping sorting for women had little impact among men, for whom there was also much less of a divide implied between single men and all men (columns 9 and 10).

7.   Conclusion

Our assessment of Canadian immigration to the United States in the early 20th century reveals gender-specific differences in the factors that determined the sorting of migrants across destinations. The results show that single French Canadian women were highly responsive to the determinants of sorting identified in our theoretical framework, with both work and social factors important for the location choice. The weaker response for married women to these forces is consistent with these women being "tied movers" for whom selection was dependent on spousal decisions. In other work we have found a strong positive correlation in stature between married men and women that we are able to match as spouses in the border

crossing records. This suggests that among married women, individual selection was partly determined by these ties, and selection patterns among married men may account for the sorting of their spouses across destinations.

What accounts for the selection patterns among men is unobserved, as the forces driving sorting among single women are remarkably absent among male migrants. Marriage market effects are of the opposite sign and similar magnitude to those seen for single women, but enclaves and employment opportunities appear to have little weight in shaping sorting into destinations. That participation rates are unimportant comes as little surprise given near uniform work among prime age adult men in the early 20th century United States. That the share of skilled employment is similarly unimportant suggests that what may have mattered most for men were (currently unobserved) differences in skill premiums across destinations, rather than migration costs or local differences in job availability.

References

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

Abramitzky, Ran, Philipp Ager, Leah Platt Boustan, Elior Cohen, and Casper W Hansen. 2019. "The Effects of Immigration on the Economy: Lessons from the 1920s Border Closure." *National Bureau of Economic Research Working Paper 26536*.

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

Antonie, Luiza, Kris Inwood, Chris Minns, Fraser Summerfield, et al. 2021. "Intergenerational mobility in a mid-Atlantic economy: Canada, 1871-1901." London School of Economics and Political Science, LSE Library.

Bailey, Martha J. 2006. "More power to the pill: The impact of contraceptive freedom on women's life cycle labor supply." *The Quarterly Journal of Economics*, 121(1): 289–320.

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

Bertrand, Marianne, Emir Kamenica, and Jessica Pan. 2015. "Gender identity and relative income within households." *The Quarterly Journal of Economics*, 130(2): 571–614.

Bhaskar, Venkataraman. 2019. "The demographic transition and the position of women: A marriage market perspective." *The Economic Journal*, 129(624): 2999–3024.

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.

Bolt, Jutta, and Jan Luiten van Zanden. 2020. *Maddison Project Database, version 2020.* Maddison Project Database.

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

Borjas, George J. 1991. "Immigration and self-selection." In *Immigration, trade, and the labor market*. 29–76. University of Chicago Press.

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.

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.

Cranfield, John, and Kris Inwood. 2007. "The great transformation: A long-run perspective on physical well-being in Canada." *Economics & Human Biology*, 5(2): 204–228.

Cranfield, John, and Kris Inwood. 2015. "A Tale of Two Armies: Comparative Growth in the Mirror of the First World War: A tale of two armies." *Australian Economic History Review*, 55(2): 212–233.

Curran, Sara R, and Estela Rivero-Fuentes. 2003. "Engendering migrant networks: The case of Mexican migration." *Demography*, 40(2): 289–307.

Damm, Anna Piil. 2009. "Ethnic enclaves and immigrant labor market outcomes: Quasi-experimental evidence." *Journal of Labor Economics*, 27(2): 281–314.

Davis, Benjamin, and Paul Winters. 2001. "Gender, Networks and Mexico-US Migration." *Journal of Development Studies*, 38(2): 1–26.

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.

Edin, Per-Anders, Peter Fredriksson, and Olof Åslund. 2003. "Ethnic enclaves and the economic success of immigrants—Evidence from a natural experiment." *The quarterly journal of economics*, 118(1): 329–357.

Eriksson, Katherine. 2020. "Ethnic enclaves and immigrant outcomes: Norwegian immigrants during the Age of Mass Migration." *European Review of Economic History*, 24(3): 427–446.

Escamilla-Guerrero, David, Edward Kosack, and Zachary Ward. 2021. "Life after crossing the border: Assimilation during the first Mexican mass migration." *Explorations in Economic History*, 101403.

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.

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. 2021. *Career and Family: Women's Century-Long Journey toward Equity.* Princeton University Press.

Goldin, Claudia, and Claudia Olivetti. 2013. "Shocking labor supply: A reassessment of the role of

World War II on women's labor supply." *American Economic Review*, 103(3): 257–62.

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.

Gratton, Brian, and Emily Merchant. 2015. "An Immigrant's Tale: The Mexican American Southwest 1850 to 1950." *Social Sciences 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.

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., and Zachary Ward. 2018. "International Migration in the Atlantic Economy 1850–1940." In *Handbook of Cliometrics*. , ed. Claude Diebolt and Michael Haupert, 1–29. Springer.

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

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.

Juif, Dácil, and Gloria Quiroga. 2019. "Do you have to be tall and educated to be a migrant? Evidence from Spanish recruitment records, 1890–1950." *Economics & Human Biology*, 34: 115–124.

Kanaiaupuni, Shawn Malia. 2000. "Reframing the migration question: An analysis of men, women, and gender in Mexico." *Social Forces*, 78(4): 1311–1347.

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.

Komlos, John, and Lukas Meermann. 2007. "The introduction of anthropometrics into development and economics." *Historical Social Research/Historische Sozialforschung*, 260–270.

Kosack, Edward, and Zachary Ward. 2014. "Who Crossed the Border? Self-Selection of Mexican Migrants in the Early Twentieth Century." *The Journal of Economic History*, 74(4): 1015–1044.

MacKinnon, Mary. 1996. "New evidence on Canadian wage rates, 1900-1930." *Canadian Journal of Economics*, 114–131.

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. 2007. "Network effects and the dynamics of migration and inequality: theory and evidence from Mexico." *Journal of Development Economics*, 84(1): 1–24.

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.

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

Morin, Alex Arsenault, Vincent Geloso, and Vadim Kufenko. 2017. "The heights of French-Canadian convicts, 1780s–1820s." *Economics & Human Biology*, 26: 126–136.

Munshi, Kaivan. 2003. "Networks in the modern economy: Mexican migrants in the US labor market." *The Quarterly Journal of Economics*, 118(2): 549–599.

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.

Ruyssen, Ilse, and Sara Salomone. 2018. "Female migration: A way out of discrimination?" *Journal of Development Economics*, 130: 224–241.

Sandell, Steven H. 1977. "Women and the economics of family migration." *The Review of Economics and Statistics*, 406–414.

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.

Stolz, Yvonne, and Joerg Baten. 2012. "Brain drain in the age of mass migration: Does relative inequality explain migrant selectivity?" *Explorations in Economic History*, 49(2): 205–220.

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.

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

Waldron, Florence Mae. 2005. "'I've never dreamed it was necessary to marry!': Women and Work in New England French Canadian Communities, 1870-1930." *Journal of American Ethnic History*, 24(2): 34–64.

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

Figures and Tables

Table 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 worker ratios for the United States adapted from Goldin and Katz (1999). For Canadian men, clerical/operative ratios adapted from Inwood, MacKinnon and Minns (2010), and authors' calculations from the 1921 Census. For Canadian women, authors' calculation from the 1911 Census.

Table 2: Summary Statistics, Top 10 Destination Choices

State	County	French Canadians / Residents	LFP Males	LFP Females	Females / Males	Skilled / Employed	Unskilled / Employed
Average County		0.044	0.922	0.283	0.981	0.182	0.417
		(0.048)	(0.048)	(0.073)	(0.076)	(0.052)	(0.055)
NH	Coos	0.181	0.934	0.197	0.791	0.150	0.568
ME	Androscoggin	0.192	0.930	0.404	1.067	0.161	0.313
ME	York	0.148	0.942	0.353	0.987	0.139	0.344
MA	Hampden	0.066	0.936	0.382	1.031	0.202	0.394
VT	Franklin	0.101	0.903	0.203	0.984	0.180	0.448
RI	Providence	0.070	0.946	0.391	1.060	0.202	0.354
NH	Hillsborough	0.181	0.941	0.442	1.065	0.152	0.372
MA	Essex	0.049	0.946	0.400	1.030	0.181	0.357
MA	Worcester	0.062	0.933	0.351	0.989	0.182	0.374
MA	Bristol	0.094	0.802	0.364	1.068	0.149	0.290

Source: St. Albans Lists. Publication number M1462. US 1920 Full count census.
Notes: We define French Canadians as those reporting their native language as French and Canada as their birthplace in the 1920 US census. Labor force participation is the share of the relevant group aged 16-65 years that is employed. Sex ratio is the female to male ratio of single individuals aged 16-65 years. Standard deviations in parenthesis.

Table 3: Summary Statistics, Immigrant Sample

	All Canadians		French Canadians		British Canadians	
	Men	*Women*	*Men*	*Women*	*Men*	*Women*
Panel A: Immigrants						
Age (years)	30.7	30.4	30.1	30.1	33.1	32.3
Literate (%)	94.6	96.1	93.9	95.7	98.1	97.9
Money (median, US dollars)	50	30	50	30	50	30
Occupation (%)						
professional	7.8	9.3	5.2	6.7	19.2	21.4
skilled	14.5	2.6	15.9	4.0	17.4	1.9
lower skilled	27.9	21.2	28.8	22.2	21.8	17.7
unskilled	27.8	2.0	28.6	1.5	17.0	1.1
farmers	12.3	0.2	12.8	0.1	11.6	0.4
none	9.6	64.8	8.7	65.5	13.0	57.5
Marital status (%)						
single	53.2	48.0	53.6	47.3	50.0	50.7
married	43.3	45.0	43.0	46.0	45.8	40.4
other	3.5	7.0	3.4	6.7	4.2	8.9
Networked (%)	87.0	81.6	87.2	83.0	86.0	82.1
US before (%)	62.8	56.9	61.3	56.8	67.6	61.3
Internal migration (%)	53.5	55.1	50.1	53.5	66.7	63.9
Distance (100s of km)	3.5	3.6	3.1	3.3	5.0	4.9
Height (in) by age cohort						
Average (full sample)	67.0	63.7	66.9	63.6	67.8	64.4
30-34	67.4	62.5	66.9	61	72	64.5
35-44	67.5	63.6	67.4	63.0	68.8	65.1
45-54	66.3	63.3	66.4	63.2	67.6	63.9
55-64	66.9	63.5	67.0	63.7	67.7	64.3
Observations	3,951	2,657	3,208	2,114	630	476
Sample share (%)	59.8	40.2	60.3	39.7	57.0	43.0
Panel B: Residents						
Canadian Bulletin of Nutrition						
Height (in) by age cohort						
30-34	68.0	62.8	65.7	61.6	68.4	63.2
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: British and French migrants have a similar profile in terms of age, access to migrant networks, literacy, marital status, immigration experience, and money at hand. However, relevant measures of selection such as height and occupation class reveal important differences across ethnic groups.

Table 4: French-Canadian enclaves and migrant sorting
Dependent variable: height (inches)

	1	2	3	4	5
		Women		Men	
	Full sample	*French*	*French single*	*Full sample*	*French*
French-Canadian share (x) of county's population					
$0.05 < x \leq 0.10$	0.1338**	0.0479	-0.0803	0.0704	0.0570
	(0.0526)	(0.0746)	(0.1044)	(0.0477)	(0.0366)
$0.10 < x \leq 0.15$	0.0473	-0.1573**	-0.2474**	0.0087	0.0361
	(0.0654)	(0.0441)	(0.0570)	(0.0657)	(0.0349)
$0.15 < x$	-0.6395***	-0.8310***	-1.1046***	0.0932*	0.0631
	(0.0541)	(0.0418)	(0.0620)	(0.0418)	(0.0432)
Observations	1,689	1,438	806	2,674	2,361
R-squared	0.1485	0.1401	0.1986	0.1448	0.1153
controls	yes	yes	yes	yes	yes
year-birth FE	yes	yes	yes	yes	yes
district-birth FE	yes	yes	yes	yes	yes

Source: St. Albans Lists Publication No. M1462.

Notes: The omitted category is $x \leq 0.05$ (no enclave). French-Canadian enclaves are defined as US counties where individuals reporting French as language and birthplace in Canada represent at least 5% of the population. We use the 1920 full-count US census to identify French-Canadian immigrants. All models control for out-of-pocket transportation costs, access to immigrant networks, internal migration in Canada, previous US immigration experience, and immigration after the Emergency Quota Act in 1921. We also control for ethnicity and marital status in models 1 and 3. Estimates in column 5 are for single, divorce, and widow women. Standard errors (in parenthesis) are clustered at the province level. * = Significant at 10% level; ** = Significant at 5% level; *** = Significant at 1% level.

Table 5: Occupational structure and migrant sorting
Dependent variable: height (inches)

	1	2	3	4	5	6	7	8
	Women				**Men**			
	Full sample	French	French single	French single	Full sample	French	French single	French single
Panel A: 1920								
Share of skilled workers	0.0715	0.1647***	0.1481***	0.0864**	-0.0371	-0.0011	0.0284	0.0276
	(0.0609)	(0.0286)	(0.0235)	(0.0277)	(0.0423)	(0.0221)	(0.0379)	(0.0376)
Labor force participation				-0.0837***				0.0102
				(0.0121)				(0.0074)
Observations	1,689	1,438	806	806	2,673	2,361	1,340	1,340
R-squared	0.1361	0.1267	0.1805	0.1807	0.1452	0.1153	0.1516	0.1516
Panel B: 1910								
Share of skilled workers	0.0612	0.1625***	0.1315***	0.1263**	-0.0379	-0.0007	0.0298	0.0189
	(0.0671)	(0.0338)	(0.0280)	(0.0360)	(0.0390)	(0.0167)	(0.0285)	(0.0278)
Labor force participation				-0.0069				-0.1446***
				(0.0193)				(0.0113)
Observations	1,689	1,438	806	806	2,673	2,361	1,340	1,340
R-squared	0.1360	0.1264	0.1800	0.1800	0.1453	0.1153	0.1516	0.1521
controls	yes	yes	yes	yes	yes	yes	yes	yes
year-birth FE	yes	yes	yes	yes	yes	yes	yes	yes
district-birth FE	yes	yes	yes	yes	yes	yes	yes	yes

Source: St. Albans Lists Publication No. M1462.

Notes: The variables of interest are standardized (z-scores) skill class to employed population ratios. Labor force participation is the share of females or males aged 16-65 years that are employed. Standard errors (in parentheses) are clustered at the province level. * = Significant at 10% level; ** = Significant at 5% level; *** = Significant at 1% level.

Table 6: Marriage markets and migrant sorting
Dependent variable: height (inches)

	1	2	3	4	5	6	7	8
	Women				**Men**			
	full sample	full sample	single	single french	full sample	full sample	single	single french
Panel A: 1920								
16-65 years	0.0542* (0.0249)				-0.0700*** (0.0147)			
16-40 years		0.0693*** (0.0164)				-0.0285** (0.0094)		
16-40 years single			0.0375** (0.0133)	-0.0275* (0.0108)			-0.0296** (0.0125)	-0.0146** (0.0048)
Observations	1,689	965	965	806	2,674	1,516	1,516	1,340
R-squared	0.1361	0.1897	0.1893	0.1789	0.1456	0.1898	0.1898	0.1516
Panel B: 1910								
16-65 years	0.0176 (0.0243)				-0.0808*** (0.0170)			
16-40 years		-0.0183 (0.0415)				-0.0399*** (0.0102)		
16-40 years single			0.0260 (0.0212)	-0.0318 (0.0171)			-0.0392** (0.0122)	-0.0158** (0.0042)
Observations	1,689	965	965	806	2,674	1,516	1,516	1,340
R-squared	0.1357	0.1892	0.1892	0.1789	0.1453	0.1897	0.1898	0.1516
controls	yes	yes	yes	yes	yes	yes	yes	yes
year-birth FE	yes	yes	yes	yes	yes	yes	yes	yes
district-birth FE	yes	yes	yes	yes	yes	yes	yes	yes

Source: St. Albans Lists Publication No. M1462.
Notes: The variables of interest are standardized (z-scores) women to men ratios for different population groups. Standard errors (in parentheses) are clustered at the province level. * = Significant at 10% level; ** = Significant at 5% level; *** = Significant at 1% level.

Table 7: Factors influencing the sorting of immigrants into US counties by gender
Dependent variable: height (inches)

	1	2	3	4	5	6	7	8	9	10
			Women					**Men**		
	Full sample	Full sample	Full sample	French	Single French	Full sample	Full sample	Full sample	French	Single French
Panel A: 1920										
Sex ratio	0.0664***	0.1732***	0.1745***	0.1129***	0.0760**	-0.0700***	-0.0723***	-0.0890***	-0.0889***	-0.0289***
	(0.0193)	(0.0410)	(0.0319)	(0.0134)	(0.0204)	(0.0170)	(0.0165)	(0.0106)	(0.0014)	(0.0057)
Labor force participation		-0.1630*	-0.1674**	-0.0758**	0.1001**		-0.0158	-0.0238	-0.0439***	-0.0012
		(0.0802)	(0.0607)	(0.0266)	(0.0324)		(0.0132)	(0.0176)	(0.0044)	(0.0057)
Share of skilled workers			-0.0043	0.1089**	0.1675***			0.0589	0.0960***	0.0630
			(0.0625)	(0.0418)	(0.0325)			(0.0433)	(0.0244)	(0.0445)
>15% of French Canadian	-0.7245***	-0.6626***	-0.6619***	-0.7279***	-0.9941***	0.0408***	0.0424***	0.0518***	0.0366	0.0503**
	(0.0219)	(0.0459)	(0.0435)	(0.0248)	(0.0253)	(0.0106)	(0.0100)	(0.0087)	(0.0215)	(0.0140)
Observations	1,689	1,689	1,689	1,438	806	2,674	2,673	2,673	2,361	1,340
R-squared	0.1488	0.1496	0.1496	0.1409	0.1995	0.1457	0.1463	0.1464	0.1164	0.1517
controls	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
year-birth FE	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
district-birth FE	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes

Source: St. Albans Lists Publication No. M1462.

Notes: The variables of interest are standardized (z-scores) women to men ratios for different population groups. Sex ratio is the female to male ratio of single individuals aged 16-40 years. Labor force participation is the share of females or males aged 16-65 years that are employed. We estimate the share of skilled workers separately for females or males. We also include an indicator variable for individuals moving to a county with French Canadians representing at least 15% of the local population. All variables are standardized except the indicator for enclaves. Standard errors in parentheses are clustered at the province level. * = Significant at 10% level; ** = Significant at 5% level; *** = Significant at 1% level.

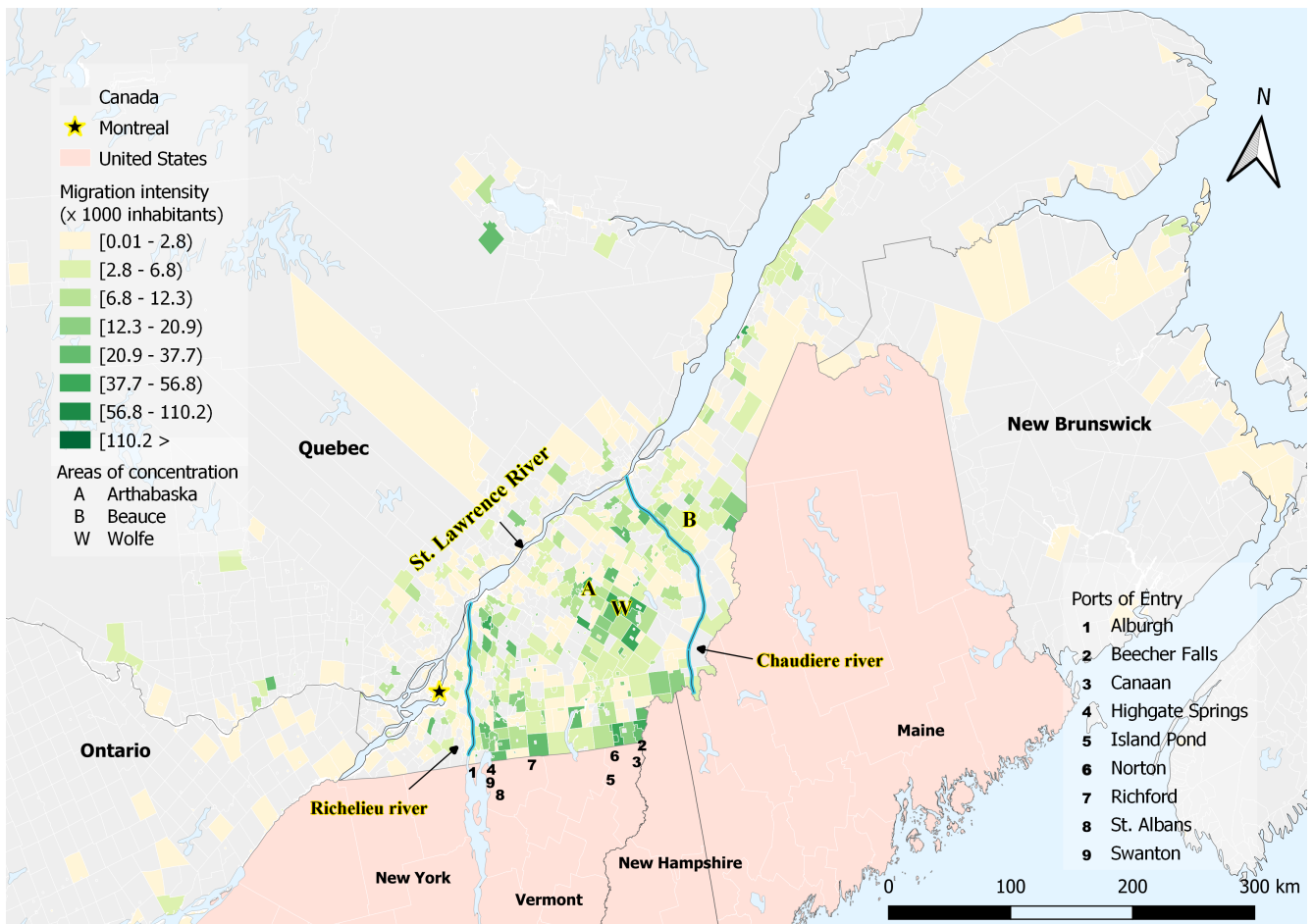*Figure 1: Canadian migration to the United States, 1906-1954*

Source: St. Albans Lists. Publication number M1462.

Notes: The polygons display emigration rates per 1000 inhabitants at the sub-district level (classes determined using Jenks Natural Breaks method).

Figure 2: Canadian migration to the United States by ethnic group, 1906-1954
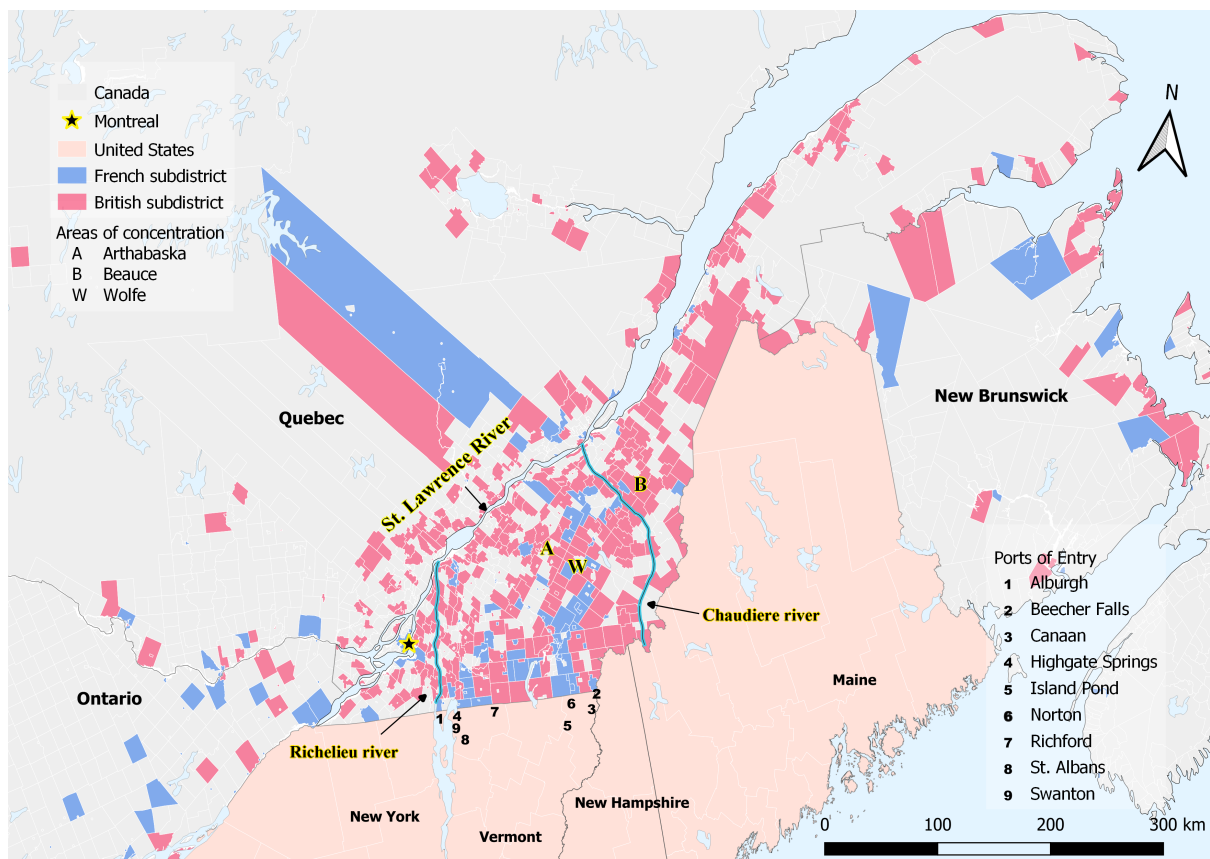
Source: St. Albans Lists. Publication number M1462.
Notes: The polygons display sub-districts classified as French or British according to the higher number of migrants from each sub-district recording migration.

Figure 3: Source and destination choices of Canadian migrants, 1906-1954
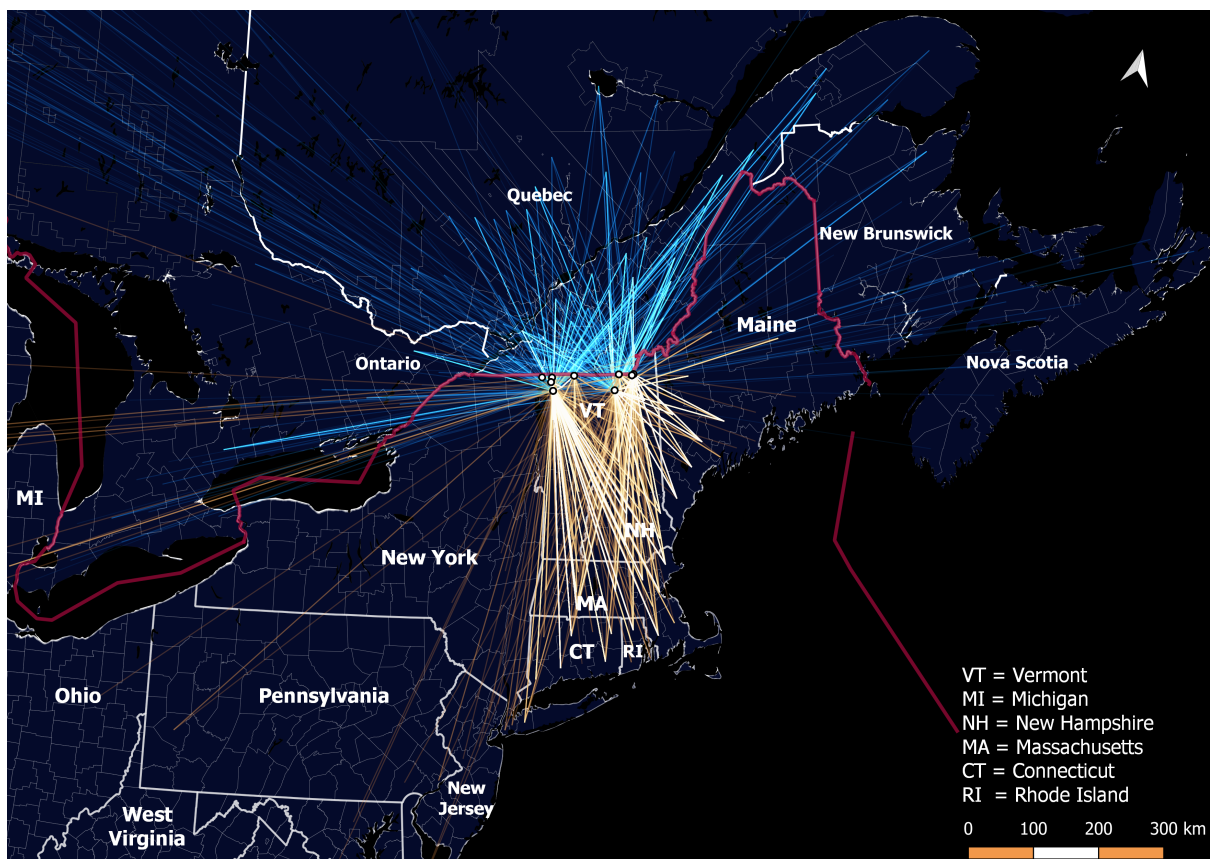
Source: St. Albans Lists. Publication number M1462.
Notes: Each line represents an individual. Overlapping lines capture the intensity of a source-destination pair by adding pixel values. Hence, brighter lines represent more intense migration flows.

Figure 4: Migrant self-selection on height by gender and year of birth

Notes: We estimate the adjusted values regressing height on a full factorial structure for ethnicity, year-of-birth cohort, and sex. The model also includes an indicator variable for individuals who migrated after the enactment of the 1921 Emergency Quota Act, an interaction between this variable and sex, and district-of-birth fixed effects. We cluster standard errors at the age-cohort level. Whiskers show 95% confidence interval. Men and women immigrants were clearly positively self-selected on height by the end of the 19th century. Self-selection patterns start to change in the early 20th century, with men and women gradually becoming less positively selected. By the 1920s, both male and female Canadian immigrants were neutrally self-selected. The degree of selection varied across genders, with females being more positively self-selected than males. The figure also shows that living standards in Canada improved significantly for men and women from the late 19th century.

Figure 5: Sorting of women migrants by enclave size
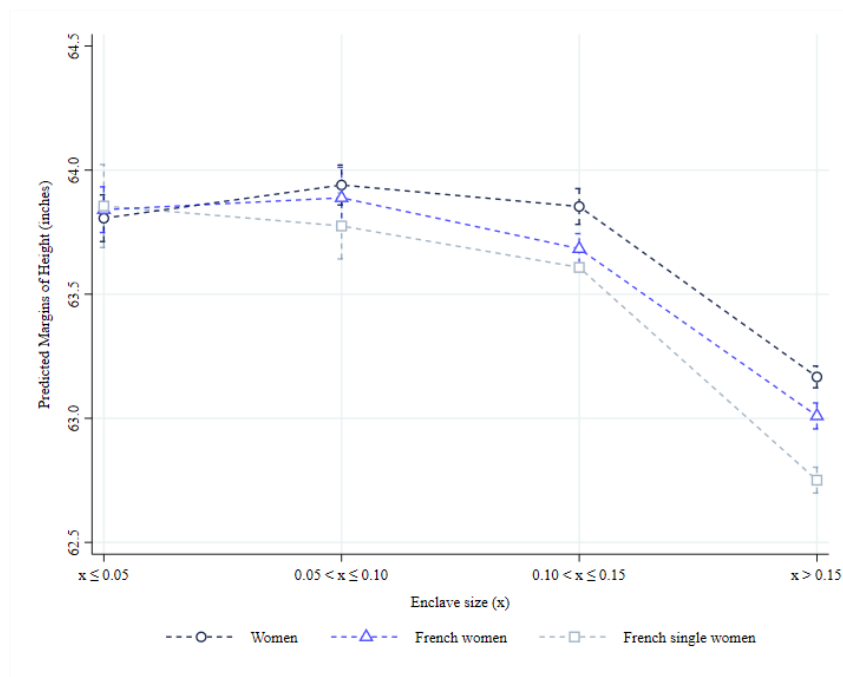
Source: St. Albans Lists. Publication number M1462.

Notes: We estimate adjusted heights controlling for ethnicity, marital status, out-of-pocket transportation costs, access to immigrant networks, internal migration in Canada, previous US immigration experience, and immigration after the Emergency Quota Act in 1921. We also include a full set of year-of-birth and district-of-birth fixed effects. The figure shows that relatively tall, single, French-Canadian women migrants were less likely to settle in large French-Canadian enclaves in the United States.

Online Appendix

Table A.1: Summary Statistics, Top 10 Destination Choices for men and women

State	County	French Canadians / Residents	LFP	Females / Males	Skilled / Employed	Unskilled / Employed
Panel A: Women						
Average County		0.054	0.291	0.989	0.298	0.411
		(0.05)	(0.077)	(0.073)	(0.076)	(0.081)
ME	Androscoggin	0.192	0.404	1.067	0.204	0.263
ME	York	0.148	0.353	0.987	0.200	0.280
NH	Coos	0.181	0.197	0.791	0.374	0.510
MA	Hampden	0.066	0.382	1.031	0.278	0.358
VT	Franklin	0.101	0.203	0.984	0.380	0.473
NH	Hillsborough	0.181	0.442	1.065	0.189	0.330
RI	Providence	0.070	0.391	1.060	0.259	0.315
MA	Essex	0.049	0.400	1.030	0.242	0.308
ME	Cumberland	0.032	0.293	1.051	0.404	0.401
MA	Bristol	0.094	0.364	1.068	0.176	0.210
Panel A: Men						
Average County		0.046	0.921	0.979	0.146	0.42
		(0.048)	(0.049)	(0.076)	(0.048)	(0.055)
NH	Coos	0.181	0.934	0.791	0.113	0.578
ME	Androscoggin	0.192	0.930	1.067	0.140	0.336
ME	York	0.148	0.942	0.987	0.117	0.367
MA	Hampden	0.066	0.936	1.031	0.170	0.409
VT	Franklin	0.101	0.903	0.984	0.136	0.442
RI	Providence	0.070	0.946	1.060	0.177	0.372
NH	Hillsborough	0.181	0.941	1.065	0.133	0.393
VT	Essex	0.049	0.946	1.030	0.155	0.378
MA	Worcester	0.062	0.933	0.989	0.149	0.378
MA	Essex	0.120	0.935	0.838	0.098	0.487

Source: St. Albans Lists. Publication number M1462. US 1920 Full count census.

Notes: Table for women and men separately. We define French Canadians as those reporting their native language as French and Canada as their birthplace in the 1920 US census. Labor force participation is the share of the relevant group aged 16-65 years that is employed. Sex ratio is the female to male ratio of single individuals aged 16-65 years. We estimate the share of skilled and unskilled workers separately for each group. Standard deviations in parenthesis.

Table A.2: Summary Statistics, Top 10 Destination Choices for women by marital status

State	County	French Canadians / Residents	LFP	Females / Males	Skilled / Employed	Unskilled / Employed
Panel A: Single Women						
Average County		0.062	0.685	1.005	0.348	0.321
		(0.051)	(0.135)	(0.074)	(0.086)	(0.074)
ME	Androscoggin	0.192	0.743	1.067	0.267	0.242
ME	York	0.148	0.699	0.987	0.269	0.250
NH	Coos	0.181	0.529	0.791	0.473	0.451
MA	Hampden	0.066	0.774	1.031	0.345	0.340
VT	Franklin	0.101	0.490	0.984	0.503	0.377
NH	Hillsborough	0.181	0.785	1.065	0.250	0.311
RI	Providence	0.070	0.802	1.060	0.310	0.295
MA	Essex	0.049	0.771	1.030	0.320	0.297
MA	Suffolk	0.003	0.778	1.029	0.447	0.372
MA	Cumberland	0.032	0.644	1.051	0.492	0.336
Panel B: Married Women						
Average County		0.056	0.116	0.991	0.177	0.440
		(0.051)	(0.047)	(0.075)	(0.072)	(0.115)
ME	Androscoggin	0.192	0.202	1.067	0.115	0.242
ME	York	0.148	0.177	0.987	0.106	0.268
NH	Coos	0.181	0.052	0.791	0.226	0.592
MA	Hampden	0.066	0.158	1.031	0.146	0.347
VT	Franklin	0.101	0.066	0.984	0.199	0.616
RI	Providence	0.070	0.140	1.060	0.135	0.306
MA	Essex	0.049	0.187	1.030	0.111	0.283
MA	Worcester	0.062	0.118	0.989	0.126	0.382
NH	Hillsborough	0.181	0.228	1.065	0.096	0.317
MA	Bristol	0.094	0.182	1.068	0.070	0.162

Source: St. Albans Lists. Publication number M1462. US 1920 Full count census.

Notes: Table for single and married women separately. We define French Canadians as those reporting their native language as French and Canada as their birthplace in the 1920 US census. Labor force participation is the share of the relevant group aged 16-65 years that is employed. Sex ratio is the female to male ratio of single individuals aged 16-65 years. We estimate the share of skilled and unskilled workers separately for each group. Standard deviations in parenthesis.

Table A.3: Male Canadian Enlistees. Summary Statistics

	Canadians	French	British
Male Canadian Residents			
Age (years)	23.4	23.6	23.3
Occupation (%)			
professional	24.2	15.6	26.2
skilled	17.1	17.5	17.0
lower skilled	16.5	14.3	17.1
unskilled	9.3	13.8	8.2
farmers	28.8	34.8	27.3
none	4.2	3.8	4.3
Marital status (%)			
single	88.7	90.0	88.4
married	11.0	9.7	11.2
other	0.3	0.3	0.4
Height (in) by age cohort			
Average (full sample)	67.1	65.8	67.4
30-34	–	–	–
35-44	–	–	–
45-54	–	–	–
55-64	66.9	65.7	67.3
Observations	7,610	1,513	6,097
Sample share (%)	100.0	19.9	80.1

Source: Data are from Cranfield and Inwood (2015). The data consists of military records of volunteer soldiers who enlisted in the Canadian Expeditionary Force (CEF) during the First World War. Due to the lowered opportunity costs of enlistment caused by the 1913 recession and widespread patriotism, lower, middle, and upper classes alike enlisted in the CEF, making these records the best representative sample for young and middle-aged Canadian men at the time. To our knowledge, this is the only data source reporting medical information, including height, at the individual level for Canadian males.

Table A.4: Self-selection of male Canadian migrants
Dependent variable: height (inches)

	1	2	3	4	5
French Males					
migrant	1.083***	1.201***	1.157***	1.256***	1.377***
	(0.130)	(0.020)	(0.016)	(0.049)	(0.053)
Constant	66.917***	68.330***	66.404***	66.371***	66.534***
	(0.130)	(0.062)	(0.648)	(0.065)	(0.034)
Observations	4,434	4,434	4,168	4,156	4,128
R-squared	0.069	0.083	0.121	0.263	0.275
British Males					
migrant	0.522***	0.822***	0.739**	0.621**	0.600**
	(0.193)	(0.206)	(0.273)	(0.196)	(0.216)
Constant	67.478***	65.414***	67.170***	67.432***	67.741***
	(0.193)	(0.331)	(0.965)	(0.180)	(0.294)
Observations	6,525	6,525	6,128	6,005	5,914
R-squared	0.019	0.041	0.066	0.269	0.284
year-birth FE	yes	yes	yes	yes	yes
state-birth FE	no	yes	no	no	no
district-birth FE	no	no	yes	no	no
subdistrict-birth FE	no	no	no	yes	yes
skill class	no	no	no	no	yes

Source: St. Albans Lists Publication No. M1462.

Notes: Standard errors (in parentheses) are clustered at the province level. * = Significant at 10% level; ** = Significant at 5% level; *** = Significant at 1% level.

Table A.5: Control variables
Dependent variable: height (inches)

	1	2	3	4	5	6
	Women			**Men**		
Distance	0.1024	0.0789	0.0815	0.0039	0.0024	-0.0967**
	(0.0558)	(0.0554)	(0.0482)	(0.0082)	(0.0104)	(0.0397)
Distance^2	-0.0025	-0.0017	-0.0011	0.0001	0.0002	0.0008
	(0.0016)	(0.0016)	(0.0016)	(0.0006)	(0.0004)	(0.0011)
Networks	-0.3192**	-0.3175**	-0.3614***	-0.1983**	-0.1978**	-0.2231*
	(0.1154)	(0.1140)	(0.0975)	(0.0815)	(0.0781)	(0.1001)
Internal migration	-0.2579***	-0.2576***	-0.2389***	-0.1766***	-0.1768***	-0.1356***
	(0.0463)	(0.0461)	(0.0346)	(0.0133)	(0.0136)	(0.0177)
US before	-0.2630***	-0.2647***	-0.2898***	-0.0231	-0.0231	-0.0471
	(0.0525)	(0.0549)	(0.0438)	(0.0696)	(0.0683)	(0.0632)
After 1921 quotas	0.0644	0.0648	0.0037	-0.1714**	-0.1714**	-0.1422**
	(0.0509)	(0.0517)	(0.0470)	(0.0588)	(0.0594)	(0.0565)
French Canadian	-0.5049**	-0.5008**	-0.5901**	-1.4069***	-1.4060***	-1.4358***
	(0.1879)	(0.1882)	(0.1770)	(0.0773)	(0.0775)	(0.0823)
Single	-0.0823*	-0.0792*	-0.0960**	-0.0880**	-0.0883**	-0.1161**
	(0.0369)	(0.0373)	(0.0334)	(0.0364)	(0.0354)	(0.0358)
Observations	1,694	1,694	1,782	2,681	2,681	2,853
R-squared	0.1356	0.1351	0.1277	0.1428	0.1428	0.1397
year-birth FE	yes	yes	yes	yes	yes	yes
district-birth FE	yes	yes	yes	yes	yes	yes

Source: St. Albans Lists Publication No. M1462.

Notes: In columns 1 and 4 distance is the linear distance from the immigrant's last residence to the US county of intended destination. In columns 2 and 5 distance is the sum of the linear distance from the immigrant's last residence to the border post of registration and the linear distance from the border post of registration to the US county of destination. In columns 3 and 6 distance is the linear distance from the immigrant's birthplace to the US county of intended destination. Standard errors clustered at the province level in parenthesis. * = Significant at 10% level; ** = Significant at 5% level; *** = Significant at 1% level.

Table A.6: Sorting of women migrants into French-Canadian enclaves in 1910
Dependent variable: height (inches)

	1	2	3	4	5
	Men		Women		
	Full sample	*French*	*Full sample*	*French*	*French single*
Panel A: Standard enclave definition					
Enclave size (x)					
$0.02 < x \leq 0.04$	0.0457	-0.0535	0.2718***	0.1355	0.3117**
	(0.1157)	(0.0940)	(0.0680)	(0.0904)	(0.0841)
$0.04 < x \leq 0.06$	0.1895	0.1027	0.1784*	0.0934	-0.0156
	(0.1065)	(0.0981)	(0.0884)	(0.0871)	(0.0961)
$0.06 < x$	0.1348	0.0200	-0.3479**	-0.5784***	-0.7074***
	(0.0887)	(0.0970)	(0.1079)	(0.0761)	(0.0936)
Observations	2,674	2,361	1,689	1,438	806
R-squared	0.1446	0.1148	0.1435	0.1363	0.1953
Panel B: Alternative enclave definition					
Enclave size (x)					
$0.02 < x \leq 0.04$	-0.0079	-0.4029***	0.2045	-0.3318	-0.6413
	(0.3381)	(0.0845)	(0.5223)	(0.5980)	(0.7582)
$0.04 < x \leq 0.06$	-0.0333	-0.4506***	0.4518	-0.1378	-0.2994
	(0.2481)	(0.0452)	(0.4155)	(0.4920)	(0.6420)
$0.06 < x$	0.1089	-0.3120***	-0.1154	-0.7533	-1.1631
	(0.2876)	(0.0508)	(0.4676)	(0.5481)	(0.6679)
Observations	2,674	2,361	1,689	1,438	806
R-squared	0.1447	0.1153	0.1417	0.1324	0.1913
controls	yes	yes	yes	yes	yes
year-birth FE	yes	yes	yes	yes	yes
district-birth FE	yes	yes	yes	yes	yes

Source: St. Albans Lists Publication No. M1462.
Notes: The omitted category is $x \leq 0.02$ (no enclave). In Panel A French-Canadian enclaves are defined as US counties where individuals reporting French as language and birthplace in Canada represent at least 2% of the population. In Panel B French-Canadian individuals report French as language or their birthplace in "French Canada". The second panel also includes the second-generation extensions of these definitions: Francophones with parents born in Canada, or any individuals with parents born in "French Canada". Enclaves are defined as counties where individuals corresponding to the extended definition represent at least 2% of the county population. We use the 1910 full-count US census to identify French-Canadian immigrants. The census undercounts the number of French speakers, and thus French-Canadians, because the enumerators were to enter a language other than English only if the individual in question did not speak English, as compared to the 1920 census where they simply recorded the language spoken at home. All models control for out-of-pocket transportation costs, access to immigrant networks, internal migration in Canada, previous US immigration experience, and immigration after the Emergency Quota Act in 1921. We also control for ethnicity and marital status in models 1 and 3. Estimates in column 5 are for single, divorce, and widow women. Standard errors (in parenthesis) are clustered at the province level. * = Significant at 10% level; ** = Significant at 5% level; *** = Significant at 1% level.