

**UCLA**

**Program on International Migration**

**Title**

The cream of the crop? Geography, networks and Irish migrant selection in the Age of Mass Migration

**Permalink**

<https://escholarship.org/uc/item/3n5876cb>

**Author**

Connor, Dylan

**Publication Date**

2018-08-17

Peer reviewed

# The cream of the crop? Geography, networks and Irish migrant selection in the Age of Mass Migration

Dylan Shane Connor  
[dylan.connor@colorado.edu](mailto:dylan.connor@colorado.edu)

*[An edited version of this article is forthcoming at the Journal of Economic History]*

## **ABSTRACT**

With over 30 million people moving to North America during the Age of Mass Migration (1850-1913), governments feared that Europe was losing its most talented workers. Using new data from Ireland in the early twentieth century, I provide evidence to the contrary, showing that the sons of farmers and illiterate men were more likely to emigrate than their literate and skilled counterparts. Emigration rates were highest in poorer farming communities with stronger migrant networks. I constructed these data using new name-based techniques to follow people over time and to measure chain migration from origin communities to the United States.

## **Acknowledgements**

For their guidance and assistance with this work, I want to express my most sincere thanks to Ann Carlos and Bill Collins (editors), two anonymous reviewers, Leah Boustan, Cormac Ó Gráda, Dora Costa, Jamie Goodwin-White, David Rigby, Michael Storper Alan Fernihough, Roger Waldinger, Peter Catron, Joe Ferrie, Francisco Haimovich, Katherine Eriksson, Gerald Mills, Niamh Moore, Eoin McLoughlin, Enda Delaney and participants at the 2015 annual meetings of the Population Association of America, the Social Science History Association and the Economic History Association, the All-UC Graduate Student Workshop in Economic History, the UCLA Proseminar in Economic History and the UCLA Center for the Study of International Migration. This research was supported by a Travelling Studentship from the National University of Ireland.

## INTRODUCTION

During the Age of Mass Migration (1850-1913), over four million people moved from Ireland to the United States. This period of intense emigration was sparked by the Great Famine in the 1840s and persisted for over half a century. Although the Irish emigrated in force to England, Canada and Australia, the United States became the primary destination for Irish settlement overseas. Emigration contributed to the decline of the Irish population from over eight million in 1840 to less than four and a half by 1913 and was precipitated by sluggish domestic economic development and expanding opportunity in the United States. The stagnation of opportunity in Ireland not only exacerbated emigration but was thought to have made matters worse by pushing Ireland's "best men" to leave (Royal Commission on Labour 1893, p. 49).

This article provides new insight on Irish migration by examining who exactly left Ireland for the United States. Despite the profound importance of Irish emigration to the economic development and social history of Ireland, Britain and the United States, we know surprisingly little about how the people who emigrated compared to those they left behind ("migrant selection"). Using new longitudinal data linking migrants and non-migrants back to their childhood homes in Ireland at the turn of twentieth century, I shed new light on who self-selected into emigration.

Whether or not emigration hindered Irish economic development is hotly debated. Published reports over 50 years claimed that Irish agricultural and industrial growth was stymied by the emigration of Ireland's most capable people (Carter et al. 1956; Royal Commission on Labour 1893). These claims have been rejected by economic historians who argue that emigration actually relieved pressure on Irish labor markets and permitted otherwise unattainable growth in wages (Begley, Geary, and Stark 2014; Boyer, Hatton, and O'Rourke 1994). Knowing who exactly left Ireland can provide new insight on this issue.

It is also probable that migrants' characteristics affected economic development in the United States. Immigration benefits the receiving economy when the skills of immigrants compare favorably to natives, or when migrants change institutions for the better (Borjas 2014). Evidence suggests that immigration from Europe benefited the US economy, and highly skilled

Irish immigration may have contributed to this (Ager and Brückner 2013; Rodríguez-Pose and von Berlepsch 2014). It is also claimed that migrant selection has direct implications for the economic trajectories of immigrants and their children in the United States (Borjas, 1994). Few studies of immigrant economic mobility, however, reliably observe immigrants' socioeconomic backgrounds in the country of origin (see Connor, 2017a for exception).

To measure migrant selection, I assemble new data on Irish-born males who decided to stay in Ireland or move to the United States in the early twentieth century. I assembled this dataset by modifying the methods pioneered by (Ferrie 1996) to link 66,000 people between the 1901 and 1911 censuses of Ireland or to the 1910 decennial census of the United States.<sup>1</sup> By integrating new measures of the migrant networks and agricultural characteristics of 158 source communities, I examine whether migration decisions vary with the characteristics of origin places.

Contrary to official reports, I find that Ireland's highly skilled and wealthier sons were *underrepresented* in the flow across the Atlantic. Irish emigrants were more likely to be the children of illiterate men and farmers in poorer communities. I estimate that the sons of farmers were 2 percentage points more likely to emigrate than their more highly skilled counterparts, and the sons of literate men were 1.2 percentage points less likely to emigrate than the sons of illiterate men (from a base probability of 4.3%). My findings are, therefore, at odds with claims that Ireland was primarily losing highly skilled workers to emigration. The fact that emigrants, on average, appear to have been particularly low skilled compared to migrants *within* Ireland suggests that the United States provided opportunities for people whose prospects were otherwise bleak.

Origin communities may have played a key role in determining who left Ireland. Emigrants were more likely to leave rural than urban areas, and farmers' sons emigrated at higher rates than the sons of agricultural and non-agricultural laborers. I show the high rate of emigration from households headed by farmers and illiterate men to be closely linked to their

---

<sup>1</sup> In Section 6 of the Online Appendix, I show that the results derived from these data are robust to individuals erroneously linked between censuses, often referred to as "false positives" (Bailey et al. 2017).

disproportionate residence in less productive agricultural communities. Further, by interacting household characteristics with a new name-based measure of historical community emigration, I show that migrant networks were particularly important in facilitating emigration for the poor and lower skilled. This likely reflects the influence of migrant networks in ameliorating poverty constraints on emigration (Spitzer and Zimran 2018).

These findings contribute to a growing literature on the economics of migrant selection (Borjas 2014; Grogger and Hanson 2011). The influence of international politics and borders in shaping contemporary migration flows makes it difficult to study who self-selects into emigration today (Boustan and Abramitzky 2016). Thus, recent studies from the Age of Mass Migration – a largely unregulated era of transatlantic migration – yield valuable new insights (Abramitzky, Boustan, and Eriksson 2012, 2013; Kosack and Ward 2014).<sup>2</sup> This article contributes to this emerging historical work by examining selection from the country with the highest emigration rate during the Age of Mass Migration, and by highlighting that community conditions can shape national-scale selection patterns.<sup>3</sup> Although this study does not directly address female migration or migrant selection to Britain, these topics would be valuable future areas for study.<sup>4</sup>

### ***Irish emigration in the Age of Mass Migration***

There are three overlapping explanations for why so many people were leaving Ireland at the turn of the twentieth century. These interpretations highlight how poverty and colonial oppression pushed people to emigrate (Miller 1988), the overarching pull of economic

---

<sup>2</sup> In addition to studies of historical international migrant selection, a growing body of research is also examining migration within the United States, particularly during the Great Migration from the South (Boustan 2016; Collins and Wanamaker 2014, 2015).

<sup>3</sup> The annual emigration for Ireland was up to 13 per 1,000 people during the Age of Mass Migration (Hatton and Williamson 1993).

<sup>4</sup> A comparison of the 1901 Census of England and Wales and the 1900 Census of the United States shows four times as many Irish born people living in the United States than in England and Wales. Further, around half the Irish emigration flow to the United States in the early twentieth century was female. In Section 12 of the Online Appendix, I use age-heaping techniques to show that Irish-born male and females in the United States were likely less numerate than Irish-born people in Ireland, England and Wales.

opportunity in the United States, and the more intermediate role of migrant networks in creating a path dependency of emigration (O'Rourke 1995). Although these interpretations suggest that emigration was higher among the poor, they disagree on the role of individual and community characteristics in influencing who decided to leave Ireland.

Historians typically emphasize the role of poverty and diminishing opportunity in Ireland as the key driver of emigration. The decline of labor-intensive farming and the consolidation of farm holdings, for example, limited land ownership and agricultural laboring opportunities (Fitzpatrick 1980). These adjustments affected the availability of economic opportunity, and in turn, demographic outcomes such as marriage and emigration (Guinnane 1997). In these accounts, people were effectively pushed off the land, with particularly acute effects on the emigration of laborers, peasant farmers, Catholics, and Irish-speakers in poorer agricultural communities (Miller 1988).

Access to overseas opportunity also affected migration decisions. Time-series analyses for Irish counties from 1877-1913 suggests that emigration decisions were made through an evaluation of economic opportunity in the United States relative to origin communities (Hatton and Williamson 1993). The correlation of emigration rates with rising overseas wages could imply that the pull of opportunity was a more important driver of emigration than push factors in Ireland (O'Rourke 1995). If this is the case, individuals with personal characteristics that were relatively better rewarded in the United States may have been particularly attracted to emigration (Abramitzky, Boustan, and Eriksson 2012).

Migrant networks also likely played a significant role in eliminating economic constraints on migration. Although travel costs declined substantially over time, the price of passage to the United States at the end of the nineteenth century could be similar to the annual rent of a farm in a poorer Irish county, or the weekly wage of an unskilled worker in the United States (Fitzpatrick 1984; Guinnane 1997). Given these sizeable costs, it may not be surprising that from 1904-1910, at least 35 percent of Irish steerage passengers crossed the Atlantic on tickets paid for by somebody known to them in the USA (Harkness 1931). Having access to migrant networks likely depended on whether one grew-up in a community highly exposed to emigration, which

may in turn have affected who decided to leave Ireland.<sup>5</sup>

Although some earlier literature suggests that Irish emigrants came from poorer backgrounds (Miller 1988), there has been no direct examination of migrant self-selection in the later decades of the Age of Mass Migration. Evidence from ship manifests from the Great Famine period (1845-1852) shows emigrants to be poorer and less educated relative to the Irish population (Anbinder and McCaffrey, 2015; Mokyr and Ó Gráda, 1982). Emigration in this period was, however, unusual in both volume and the motivations of migrants. This study is, therefore, the first micro-analysis of Irish migrant selection during the more typical early twentieth century period of emigration.<sup>6</sup> I specifically examine who moved, and how community and individual characteristics interacted to influence the composition of the emigration flow.

## **DATA AND ANALYTICAL STRATEGY**

To study migrant selection from Ireland, I used record linkage techniques to construct samples of Irish-born males whom I could observe in their homes in Ireland as children (“Child Sample”) or as young working men (“Adult Sample”). By linking these males to later censuses, I can ascertain whether they decided to stay in Ireland or move to the United States. Although the Adult Sample allows me to assess how migrants were selected from the Irish labor force as of 1901, the Child Sample provides the cleanest estimates of selection as fathers’ occupation is not likely to have been affected by their sons’ childhood expectations regarding migration.<sup>7</sup>

The Child Sample is comprised of sons aged 6-18 in 1901. Because the occupations of adults may have been jointly determined by their expectations regarding emigration, I prefer the Child Sample for comparing emigrants to non-migrants by their economic backgrounds. That is,

---

<sup>5</sup> The high rate of savings that would be required by emigrants to remit substantial sums of money to Ireland have been noted as far back as the Great Famine (Wegge, Anbinder, and Ó Gráda 2017).

<sup>6</sup> This study is a novel contribution to a rich literature on the history of Irish migration, which has traditionally relied on qualitative sources and aggregate statistics (see Kenny 2014; Casey and Lee 2007; Meagher 2009; Fitzpatrick 1994).

<sup>7</sup> If a male aged 15-18 held an occupation of their own and also lived with their father, they could be included in both the Adult and the Child Sample.

to avoid the issue of joint determination, I focus on the occupations of co-resident fathers while their sons were in childhood. Focusing on fathers' characteristics while the sons are young also helps avoid reverse causality in fathers' wealth through the earlier migration of their children.

The Adult Sample is comprised of men aged 15-25 who reported holding an occupation of their own in Ireland in 1901. It is important that I observe these men in the Irish census prior to migration, as immigrants may hold occupations below their true level of skill in the new country.<sup>8</sup> While I cannot overcome the fact that the occupations of these men in 1901 may have been affected by expectations regarding subsequent emigration or that these cohorts may have been partially depleted by prior emigration, the migration decisions of these men provide new insight on which segments of the working-age population decided to leave Ireland.

### ***Background to record linkage***

The last ten years has seen fast progress in the use of automated record linkage to construct longitudinal samples. One of the main challenges in following individuals across censuses is distinguishing between individuals with similar names, years of birth and birthplaces. In pioneering work, Ferrie (1996) manually linked 5,000 individuals from the Public Use Micro Sample for 1850 census of the United States to the 1860 federal census manuscript schedules. He did this by developing a set of decision rules to minimize the possibility of incorrectly linking different individuals with similar names and ages. While this linkage approach provides some (pre-defined) leeway in the reporting of years of birth and names to adjust for transcription and reporting error, individuals with more closely matching years of birth are preferred as "successful links." By exploiting advances in computing power and data availability, economic historians have continued to build on the logic of Ferrie's (1996) approach to construct larger and more diverse longitudinal samples (e.g. Abramitzky, Boustan and Eriksson, 2012; Abramitzky, Mill, and Pérez, 2018; Feigenbaum, 2016a; Kosack and Ward, 2014).

Recent evidence suggests that the quality of linked samples varies by their linkage criteria:

---

<sup>8</sup> It is also preferable to measure occupations prior to the immediate migration period (as in departure records), as transitory income or employment shocks can induce emigration leading to the underestimation of migrants' permanent income or occupational status (Fernández-Huertas Moraga 2011).



the phonetic standardization of names and low quality data are particularly likely to generate incorrect matches or “false positives” (Bailey et al. 2017). This is concerning as high rates of false positives can bias estimates from linked samples (Massey, 2017; Ruggles, Fitch, and Roberts, 2017). Although the evidence is thin on whether false positives have led to distorted research conclusions, these findings demand new vigilance when constructing linked samples.

Such care is particularly important when linking migrants across international borders. Because censuses typically only report the country of birth of immigrants (e.g. “Ireland”) rather than their county or state of birth, we have less information to distinguish between individuals with similar names and birth years. Losing detailed birthplace information can lower sample sizes and increase false positives. False positives are of particular concern here, as they inflate the error in observed origin and destination locations, which could bias measures of selection and weaken the estimated effect of origin community characteristics on migration outcomes.

To improve the quality of my linked sample from Ireland to the United States, I develop a new linkage procedure (the “Geographical Name Match”). I augment the linkage approaches of Ferrie (1996) and Abramitzky, Boustan and Eriksson (2012) in two ways.<sup>9</sup> First, based on recent findings that name standardization or excessive cleaning can increase false positives, I rely mainly on names as they were reported in the census. Second, and of great novelty, the Geographical Name Match exploits the fact that many Irish last names are strongly tied to counties and provide meaningful information for analyzing migration (Matheson 1901).

For a sizeable share of the Irish-born population in Ireland in 1901, their birth county could be inferred with reasonable accuracy from their last name. I used the 1901 Census of Ireland to calculate how concentrated last names were to specific counties of birth and then limit the original census records to the 15 percent of men with county-specific last names. A county-specific last name is one that is 60 percent confined to a single county of birth. I chose this cut-off as it reflects the point at which, after imputation, four out of five men have an identical county

---

<sup>9</sup> The basic ABE matching codes can be downloaded directly from <https://people.stanford.edu/ranabr/matching-codes>. The Online Appendix provides a detailed discussion of my record linkage approach and comparisons to other methods.

of birth reported in the census as that implied by their last name.<sup>10</sup> As the county of birth is imputed for individuals in each of the data sources, even an incorrect imputation of birthplace will not preclude a successful or correct link from the 1901 Census of Ireland to later census years.

Having limited the sample to men with geographical last names, I linked individuals aged from 6 to 25 in the 1901 Census of Ireland to either the 1911 Census of Ireland or the 1910 Census of the United States based on their full name and approximate age.<sup>11</sup> I accounted for age misreporting and differences in the enumeration dates of the censuses by using the iterative age matching procedure employed by Abramitzky, Boustan and Eriksson (2012).<sup>12</sup> In using this approach, I first searched for individuals from the 1901 census of Ireland with an identical name and an exactly matching year of birth.<sup>13</sup> If no match could be found, I iteratively broadened the search criteria to people that are up to two years younger or older. Cases matched to multiple individuals by name and age are ineligible for inclusion and dropped from the sample.

In total, there are 125,317 men with a name above the 60 percent threshold in the 1901 census. I successfully linked 36,060 or 29 percent to an Irish-born person with the same name and approximate age in Ireland or the United States in 1910/1911. Of these linked men, 1,953 lived in the United States in 1910, corresponding to an annual emigration rate of around 6 per thousand for males aged 6-25 in 1901.<sup>14</sup> Overall, for the Geographical Name Match the “forward

---

<sup>10</sup> It would be possible to use this information to impute a county of origin for Irish-born people in the USA. However, the occupational backgrounds of individuals with geographically concentrated last names are not randomly distributed. Thus, it is more sensible to use the same matching criteria for the migrant and non-migrant samples.

<sup>11</sup> The data for the 1901 Census of Ireland were prepared by Connor (2017b) and Connor, Mills, and Moore-Cherry (2011), and the data from the United States have been prepared by the IPUMS-USA group and their collaboration with Ancestry.com (Ruggles, Genadek, Goeken, Grover, and Sobek, 2017).

<sup>12</sup> The 1901 census of Ireland was enumerated on the 31<sup>st</sup> of March 1901. As the 1911 census was enumerated on April 2<sup>nd</sup>, 1911 and the 1910 decennial census of the United States on the 15<sup>th</sup> of April 1910, it is possible for an individual to have returned home between the two censuses. Such individuals were dropped from the analysis.

<sup>13</sup> The only major modification I make to names is to standardize for nicknames (e.g. making “Joe” and “Joseph” or “Abe” and “Abraham” equivalent). To simultaneously match individuals from the 1901 census of Ireland to the two later censuses, I inflate the ages of males in the 1910 US census by one year and merge it with the 1911 census.

<sup>14</sup> The average annual emigration rate from the official statistics is 7.9 per thousand over the 1901-1909 period and

linkage” rate from the 1901 census to the later censuses is 29 percent for males aged 6-25 in 1901.<sup>15</sup> Although some of the non-matched cases are due to linkage failure or name commonness, a non-negligible share of the missing individuals will have either died or moved to countries like England, Canada or Australia. While migration to these other destination or selective mortality are topics worthy of further study, they are beyond the scope of this paper.

Nonetheless, the Geographical Name Match appears to be a significant improvement on samples generated by other common approaches. Despite employing quite a stringent matching criteria (i.e. matching on full names), the linkage rate of 29% is high compared to previous studies. The estimates of inter-county migration generated by this sample are also much more plausible than those generated by other linkage approaches.<sup>16</sup> In the Online Appendix, I compare estimates generated from samples matched with different levels of stringency. These results suggest that my analyses are not being heavily biased by false positive matches.

Table 1 presents the basic summary statistics from the Geographical Name Match and the base census data from which they were produced. This table shows the population share for categorical variables, and the mean values with standard deviations in parentheses for continuous variables. The average son in the Child Sample is 11 years old in 1901 and is growing up in a farming household with more than four siblings. Although the characteristics of the linked

---

the correlation of the county level emigration rate from the linked data and the official statistics is +0.90.

<sup>15</sup>The “forward linkage” rate refers to the share of males successfully linked from the first year of observation (1901) to the later censuses and the “backward linkage” rate is the share of Irish males in the eligible age range (15-34 in 1910) in the United States in 1910 that are successfully linked back to Ireland. In this study, the forward linkage rate from the 1901 census is 29 percent and the backward linkage rate for Irish-born males aged 15-34 in the 1910 US census is 11 percent. Discrepancies in these rates have been observed elsewhere (Abramitzky, Boustan and Eriksson, 2012) and in the Irish case likely reflect prior migration to destinations such as England or Canada, or reporting and enumeration discrepancies between the Irish and US censuses.

<sup>16</sup>The intercountry migration rate in Ireland in the Geographical Name Match is 17 percent for the 1901-1911 period. This is quite similar to the rate of 13 percent estimated by Guinnane (1997, p. 122-123). As shown in Appendix Table 3 of the Online Appendix, however, when the sample is linked only by name and year of birth with no birthplace-name restriction, the intercountry migration rate is implausibly high at 34 percent.

sample generally resemble those for the base population, the linked sample contains lower shares of Catholics, rural dwellers and sons of illiterate men, suggesting that higher status households are slightly overrepresented in the linked sample.

TABLE 1. SUMMARY STATISTICS FOR LINKED SAMPLES AND THE 1901 CENSUS OF IRELAND

	Variable type	Child Sample		Adult Sample	
		Base Population	Linked Sample	Base Population	Linked Sample
Can read and write	Categorical	0.73	0.77	0.91	0.93
Occupation					
White-collar		0.08	0.08	0.08	0.09
Skilled worker		0.17	0.20	0.20	0.21
Urban laborer	Categorical	0.09	0.09	0.11	0.10
Farmer		0.56	0.54	0.41	0.43
Farm laborer		0.10	0.09	0.20	0.17
Age in 1901	Continuous	12 (3.7)	11 (3.7)	20 (3)	20 (3)
Catholic	Categorical	0.79	0.71	0.75	0.74
Speaks Irish	Categorical	0.11	0.13	0.15	0.19
Parents present	Categorical	-	-	0.49	0.58
Married	Categorical	-	-	0.04	0.04
Family size	Continuous	4.35 (2.64)	5.5 (2.13)	-	-
Birth order					
First son		0.28	0.28	-	-
Last son	Categorical	0.25	0.25	-	-
Middle son		0.37	0.37	-	-
Only son		0.10	0.10	-	-
Origin community					
Lived in urban area	Categorical	0.18	0.21	0.27	0.22
Network strength	Continuous	1.002 (0.002)	1.002 (0.001)	-	-
Land value (per acre)	Continuous	0.56 (0.29)	0.53 (0.29)	-	-
Inequality of landholding	Continuous	0.43 (0.10)	0.42 (0.10)	-	-
Observations	-	55,589	17,669	53,565	13,785
Match rate	-	-	32%	-	26%

Notes: The match rate for the Birthplace Sample is calculated from a base population with a last name above the 60 percent threshold. The standard deviations of continuous variables are shown in parentheses. The Child Sample is aged 6-18 in 1901 and live with their fathers in 1901. The Adult Sample is aged 15-25, report an occupation of their own and may or may not live with their parents. In the Child Sample, occupation and literacy are measured for the fathers (heads of household) and in the Adult Sample, these characteristics are measured for the men in the sample.

Comparisons of the Adult and Child Samples also indicate the presence of intergenerational change. The literacy rate – the share of individuals reporting an ability to read and write – is substantially higher for sons in the Adult Sample than for their fathers in the Child Samples. This difference between the two samples reflects rising literacy levels in Ireland in the

late nineteenth and earlier twentieth century. Although there is a much lower share of farmers in the Adult Sample than farmers' sons in the Child Sample, this partly reflects that men in the Adult Sample were still quite young (average age of 20) to have acquired land by 1901. This is suggested by the fact that around half the Adult Sample were still living with their parents in 1901. Thus, the analyses that follow include a control variable for whether individuals were co-resident with their parents.

### ***Estimating migrant selection***

To measure migrant selection, I estimate how personal and household characteristics are correlated with migration decisions. The selection variables of interest are measures of occupation, literacy, religion, Irish-speaking ability, family size and birth order.<sup>17</sup> As neither variable is recorded in the censuses of the United States, the religion and Irish-speaking variables have not typically been available for examination. Economic historians regard the individual-level reports of Irish-speaking ability as quite clean measures of whether or not an individual possessed Irish fluency (Ó Gráda 2006).<sup>18</sup> It is worth keeping in mind, however, that most young Irish-speakers would have spoken English. As the census did not ask about earnings, I used recent work from Fernihough, Ó Gráda, and Walsh (2015) to classify occupations according to a modified version of the Historical International Social Class Scheme (HISCLASS) (van Leeuwen and Maas 2011). This permits me to infer skill, class and land ownership from the occupational returns.

I estimate migrant selection for the Adult and Child Samples using linear probability models. These models predict whether migrants and non-migrants differed from each other with respect to the selection variables of interest. The basic regression equation is specified as:

$$Y(\text{Migrate}) = \beta_0 + \sum_{k=1}^K \beta_k X_k + u_j \quad (1)$$

where  $Y$  refers to whether an Irish male in 1901 moved to the United States by

---

<sup>17</sup> The data on birth order and total siblings were extracted directly from the 1901 census. Individuals were assigned a birth order of "First son", "Last son", "Middle son" or "Only son" based on their age relative to their siblings, as observed within their childhood household. As these variables will be mismeasured for families where older siblings have already left home, I control for fathers' age with a fixed effect when using this variable.

<sup>18</sup> Consult Wolf (2014) for discussion of the Irish-speaking landscape of nineteenth century Ireland.

1910/1911. Thus,  $\beta_k$  can be interpreted as the percentage point difference in the probability of an individual moving to the USA with a one-unit change in the  $k^{th}$  independent variable. Selection is inferred from the differences in the probability of migration by occupation and literacy, where the reference group are individuals still living in Ireland in 1911.

### ***Characterizing origin communities***

Rural Irish communities may have influenced migration decisions by providing economic opportunity or affecting access to migrant networks (Hatton and Williamson 1993). As such, I use Poor Law Unions (“PLU”) as my units to measure conditions in origin areas.<sup>19</sup> Poor Law Unions were local government units formed under the Irish Poor Law Act of 1838 and owing to their administrative function, a wide range of social and economic data were collected for these locales. As the digitized census records do not contain information on households’ PLU, I used the Topographical Index of Ireland to link the 3,000 enumeration districts from the 1901 census to one of the 158 PLUs listed in the land census (Matheson 1908). This PLU link allowed me to characterize communities according to inequality in landholding, the productivity of land, and the strength of emigrant networks. To derive comparable estimates, I standardize these measures so that a one-unit increase in any of these variables is associated with a one standard deviation change.

### ***Measuring migrant networks***

I use a new name-based approach to measure the strength of emigrant social networks in Irish PLUs. Extending on previous work using last names as proxies for chain migration (Wegge 1998, 2008), I measure community-based migrant networks by comparing how similar the last names of each PLU in 1901 are to the Irish-born population in the United States in 1900.<sup>20</sup> To measure name similarity, I calculate the average number of letter replacements that it would

---

<sup>19</sup> The Irish Poor Law Act emerged from the Royal Commission on the Poorer Classes in Ireland in 1833. PLUs were charged with the administration of poor relief which were based on local poor rates. PLUs were comprised of electoral divisions and existed as administrative units into the early twentieth century. (Crossman 2013) has undertaken a detailed analysis of the Poor Law Unions and Gregory (2008) has georeferenced their geographical boundaries (accessible through the British Data Archive website: [www.britishdataarchive.com](http://www.britishdataarchive.com)).

<sup>20</sup> See Section 2 of the Online Appendix for a fuller discussion of the migrant network measure.

take to transform one name into all others. The Jaro-Winkler string distance has become a widely used method for calculating the similarity of names based on letter replacements.<sup>21</sup>

I begin by computing the Jaro-Winkler distance between every pair of names observed in Ireland in 1901 and the United States in 1900. Using this matrix, I calculate two statistics for individuals. Statistic 1 is the average distance between each last name in Ireland in 1901 to all last names held by Irish-born people in the 1900 Census of the United States. This captures the average string distance of each last name in Ireland to the names held by Irish-born people in the United States. Statistic 2 is like Statistic 1 but is, instead, calculated relative to the last names held by all Irish-born people in the *combined* 1901 Irish Census and the 1900 US census. Thus, Statistic 1 represents the average similarity of each name in Ireland to the distribution of names in the United States, while Statistic 2 measures the similarity of each name to the complete distribution of names in Ireland and the United States.

I use Statistic 2 to adjust for baseline similarities between names. Take, for example, the names Connor, McCarthy and Duffy, which were held by similar numbers of people in the US in 1900. While the name Connor is similar to common names such as Connelly and O'Connor, and McCarthy resembles names like McGrath and McCauley, there are fewer names like Duffy. As a result, names like Connor and McGrath are mechanically more similar to emigrant names irrespective of actual migration patterns. I overcome this problem in Equation 2 by taking the average string distance of each individual's name  $i$  to the Irish-born population in the United States (Statistic 1) and dividing by the average distance to names in the combined population (Statistic 2). This produces a measure of how overrepresented each individual's name is in the United States relative to full Irish-born population in Ireland and the United States. I average these measures across every  $j$  PLU to measure community network strength.<sup>22</sup>

---

<sup>21</sup> I rely on the `jarowinkler` package in Stata for these operations (Feigenbaum 2016b). In this procedure, extra weight is placed on the similarity of the beginning of names, and the metric is scaled between 0 (not similar at all) and 1 (exact match).

<sup>22</sup> Although I could implement this measure at the level of the individual, my approach reduces noise among unusual last names and provides a comparable scale of measurement to other PLU characteristics.

$$\text{network strength}_j = \frac{\sum \left( \frac{\text{Name similarity to Irish population in USA ("Statistic 1")}_i}{\text{Name similarity to Irish population in Ireland or USA ("Statistic 2")}_i} \right)}{N_j} \quad (2)$$

This network strength measure can be interpreted as a measure of the social proximity of Irish communities to the Irish population in the United States. By comparing emigrant names to those of people still living in origin communities, my measure plausibly captures existing ties, reflecting the strength of a community's migrant network.<sup>23</sup> With respect to interpretation, it should be noted that this network measure partially captures the persisting local conditions that generated historical migration.

The performance of this variable in measuring migrant network strength is supported by its strong positive correlation (+0.61) with the official estimate of the county-level emigration rate from 1851-1902.<sup>24</sup> Figure 1 shows higher emigration rates and migrant network strength in the western Irish counties of Munster and Connacht including Cork, Kerry, Clare, Galway and Limerick. Emigration rates and network strength are lower in eastern counties in the provinces of Leinster and Ulster including Dublin and the Belfast counties of Antrim and Down. These strong correlations provide credence for the validity of the network measure, with its main strength being that it can be flexibly derived at different spatial scales (i.e. PLUs).

### ***Land structure and productivity***

The measures of the distribution and value of landholdings were transcribed from the Irish land census of 1901. These registers detail the valuation and number of families living on agricultural holdings within each PLU.<sup>25</sup> The census classified valuations into eleven categories ranging from £4 or less to greater than £300. I used this information to measure the average per

---

<sup>23</sup> Typically, networks are measured using the total number or the share of people leaving a community to measure the total or relative size of the migrant network (Hatton and Williamson 1993; Munshi 2003).

<sup>24</sup> These data are drawn from official county-level estimates of from the Registrar General which are available in the ("Emigration Statistics of Ireland for the Year 1902" 1903) and other similar reports. These statistics were collected at Irish ports and substantially underestimate the true emigration rate (consult (Ó Gráda 1975) for further discussion).

<sup>25</sup> A family was defined as a married couple with children (if any) or a collective of people who shared a house and boarded at the same table.



acreage value of landholdings and to construct a Gini index of inequality for each PLU. These data from the land census are only applicable to rural areas.

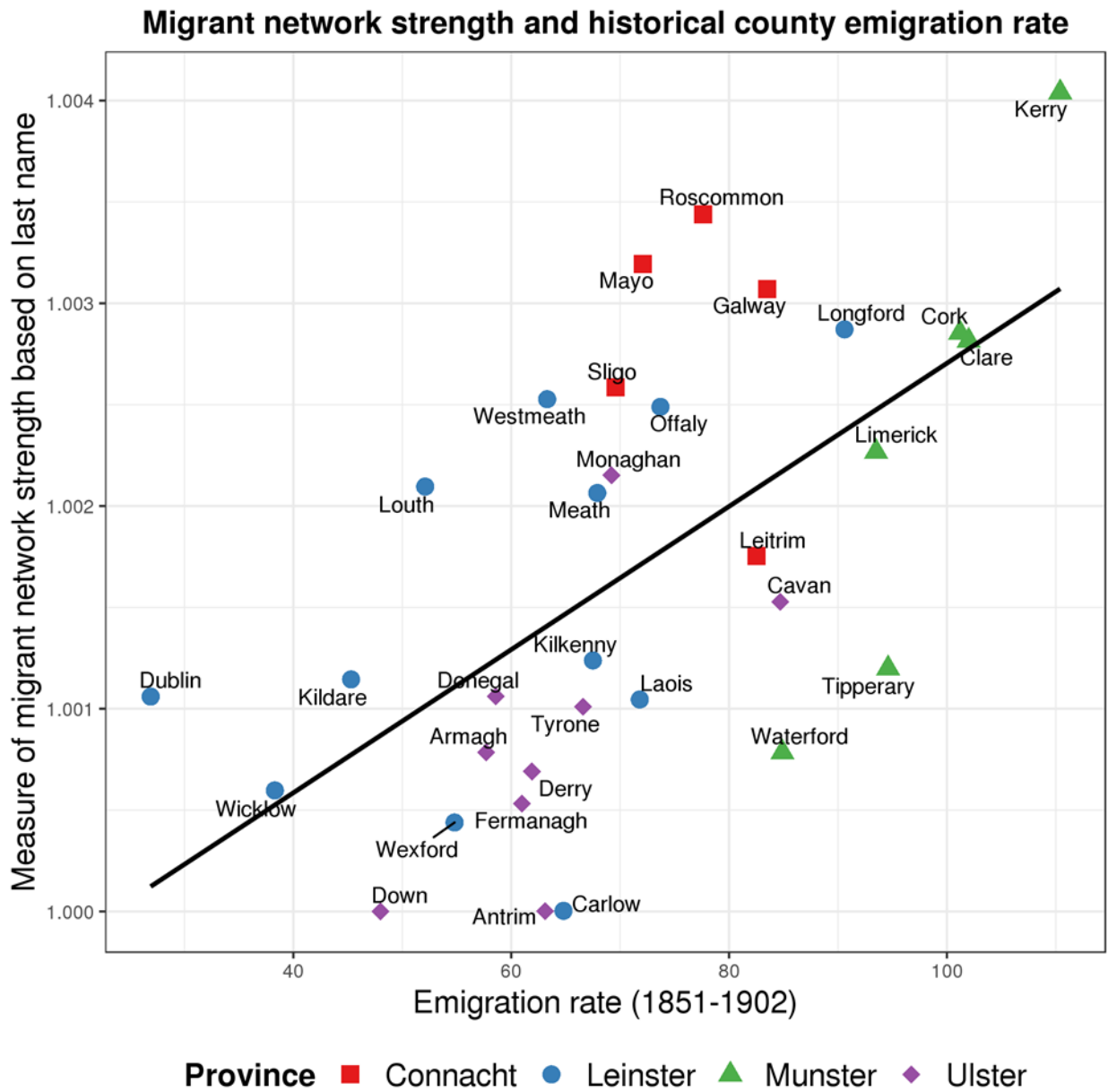


FIGURE 1. MIGRANT NETWORK MEASURE AND COUNTY EMIGRATION RATE 1851-1902

Notes: This graph plots the name based migrant network measure for the 1900/1901 census years for counties against the estimated county-level emigration rate reported in the official emigration reports. See “Emigration statistics of Ireland for the year 1902” in the references. The emigration rate corresponds to the number emigrating for every 100 of the average population over the full 1851-1902 period.

The value of landholding per acre may be best interpreted as a measure of productivity rather than wealth. Because the history of tenancy in Ireland makes it difficult to draw sharp distinctions between owners and occupiers (Turner 2002), the valuation of land may better reflect agricultural rents (Burt 1986; Featherstone and Baker 1987) and the capitalization of land use (Borchers, Ifft, and Kuethe 2014). Thus, this measure provides a general proxy for rural economic productivity. I also use these data indirectly to measure the economic position of farming households. To overcome the lack of information on household wealth in the indexed census, I use the per acreage land values of PLUs to classify farmers and laborers based on whether they lived in high, medium or low productivity areas.<sup>28</sup>

## **IRISH MIGRANT SELECTION FROM 1901-1911**

### ***Describing Irish emigration patterns***

I begin by using these data to describe the occupational backgrounds and origin locations of emigrants. Figure 2 decomposes the sample of 66,000 males in the linked dataset by whether they stayed in their origin county, moved county within Ireland or moved to the United States between 1901-1911. Panel A segments migration decisions by fathers' occupation (Child Sample) and Panel B presents differences by own occupation (Adult Sample). The shares presented in Figure 2 sum to 100% by the categories of migration decision.

It is evident from Figure 2 that migration decisions did, in fact, vary by occupational background. Panel A shows that although around half of non-migrants in the Child Sample are the sons of farmers, they are still highly overrepresented in the flow to the United States: farmers' sons account for 55% of the population staying in their home communities, 38% of migrants to Irish urban areas, but 67% of the emigration flow. Moreover, the sons of farmers are the only occupational group overrepresented among emigrants, highlighting the importance of farming origin in the story of Irish emigration.

Although the patterns based on own occupation (Panel B) mostly mirror those for fathers' occupation (Panel A) with respect to emigration, the main points of divergence are *within* the

---

<sup>28</sup> The classifications of high, medium and low were made by cutting the land value variable into terciles based on the farmer and farm laborer sample.

agricultural sector. Panel B shows that farmers themselves equally account for around 45% emigrants, inter-county movers and non-migrants. *Farm laborers*, in contrast, are overrepresented among inter-county movers and emigrants when compared to the non-migrant population. Thus, the Adult Sample appears to diverge from the Child Sample in that farmers (landholders) are relatively less likely to emigrate than farm laborers (landless), suggesting that landholding itself affected emigration decisions. Moreover, the similarity of migration shares by occupation in 1901 in the Adult Sample suggests emigration was widely spread, and selection by own occupation was relatively weak.

Individuals and families in rural Ireland not only differed by landholding status but also with respect to productivity. As such, the occupation of “farmer” masks much heterogeneity. By mapping the land value of PLUs (left) and their emigration rates (right), Figure 3 offers further insight on the backgrounds of migrants. Land values tended to be high in proximity to the major port cities of Dublin, Belfast and Cork, and in more productive landlocked agricultural regions across the south of the island. Land values were lower, in contrast, in less productive rural communities along Ireland’s western Atlantic coast.<sup>29</sup>

It is telling, therefore, that emigration was most pronounced from less productive rural communities. Figure 3 shows that up to 10 percent of the sample emigrated from some of these western communities: such emigration rates are more than ten times higher than those in the areas surrounding Dublin. Thus, the juxtaposition of these maps indicates that emigration may not only have been affected by landholding status but was also most acute in Ireland’s least productive rural communities. The following sections use regression frameworks to more directly measure and test the robustness of these patterns of migrant selection.

---

<sup>29</sup> In the early twentieth century, many of these areas were the focus of targeted development policies (Breathnach 2005).

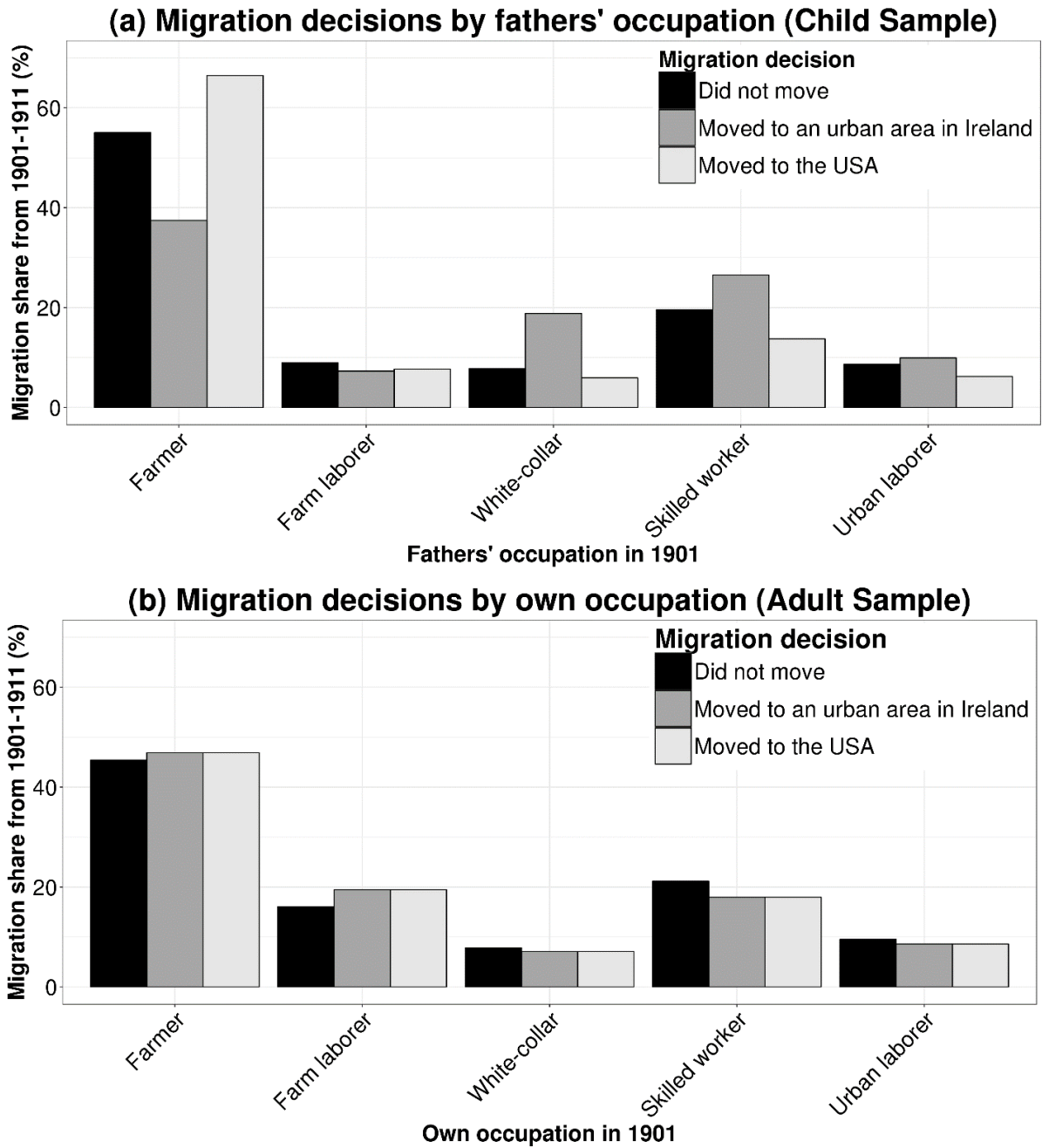


FIGURE 2. OCCUPATIONAL BACKGROUND IN IRELAND AND MIGRATION DECISIONS 1901-1911

Notes: These graphs reflect the share of males in the sample moving to the United States or moving to urban areas within Ireland. These statistics are based on the raw percentages observed in the Child Sample and the Adult Sample.

## Emigration rate and land value of Poor Law Unions

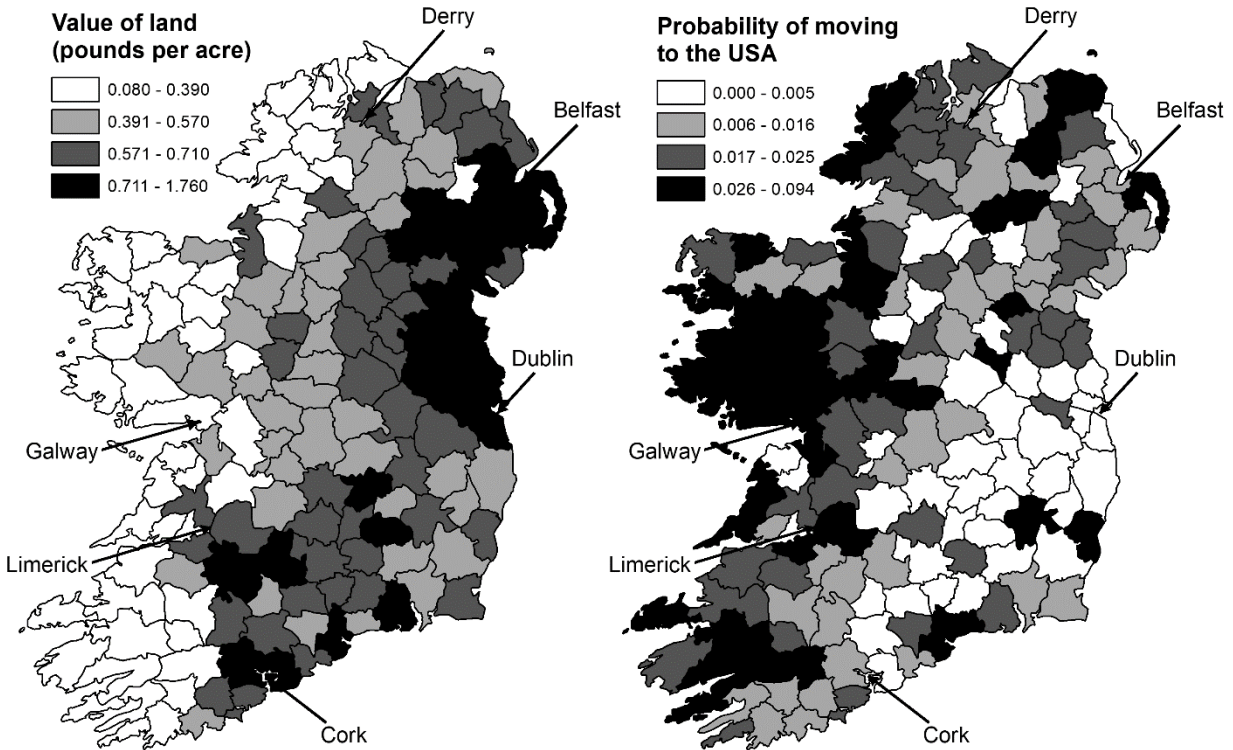


FIGURE 3. EMIGRATION TO THE USA AND LAND VALUES IN THE EARLY TWENTIETH CENTURY BY POOR LAW UNION

Notes: Darker colors are associated with higher probabilities of migration. These estimates are generated from two separate linear probability models predicting migration to the United States or to Ireland with no control. These estimates are based on the sample of 17,669 sons who were co-resident with their fathers in 1901 (Child Sample). The data are mapped with quantile breaks.

### ***Selection from the Adult and Child Samples***

How did emigrants compare to the people they left behind? This section addresses this question by estimating migrant selection using linear probability models (Equation 1). As fathers' characteristics provide the cleanest measures of selection, I begin by estimating the probability that sons left Ireland based on their fathers' occupation and literacy. As occupation and literacy are correlated with one another, this stage of the analysis focuses on their univariate effects on emigration decisions, which are conditional only on age (modelled with a fixed-effect).

Columns 1-3 of Table 2 present the models pertaining to emigrant selection within the Child Sample. The estimates in Column 1 show the sons of literate men to be 1.2 percentage points less likely to move to the United States than the sons of illiterate men (from a base probability of 4.3%). The estimates in Column 2 also confirm the finding of relatively higher emigration rates for farmers' sons: the sons of white-collar and skilled workers, and urban laborers, are around 1.8 to 2 percentage points less likely to emigrate than the reference group in the regressions (the sons of farmers). As farm laborers' sons are also less likely to emigrate than farmers' sons, these estimates do not just reflect emigration differences between urban and rural areas.

By decomposing farming and laboring households by local per capita land value, Column 3 provides evidence regarding the economic backgrounds of emigrants from agricultural areas. As a baseline, 6.5 percent of the sons of farmers in low productivity areas – the reference group in the regressions presented in Column 3 – moved to the United States. The estimates in Column 3 show farmers' sons in medium and high productivity areas were 1.1 and 3.5 percentage points less likely to emigrate, implying that the sons of poorer farmers were more likely to leave Ireland. In Column 3, there is also *no significant difference* in the probability of emigration for farmers' and laborers' sons within low productivity areas ( $p < 0.10$ ), suggesting stronger patterns of selection on origin location than occupation background. Further evidence for this claim is provided by Columns 3 and 6, which show that farm laborers and their sons were also relatively less likely to leave medium and high productivity areas than their farming and laboring counterparts in low productivity areas. These results suggest that the overrepresentation of farmers' sons in the emigration flow partly stemmed from their concentrated residence in less productive areas.

TABLE 2. OCCUPATION, LITERACY AND EMIGRATION TO THE UNITED STATES 1901-1911

Outcome:	Child Sample			Adult Sample		
	(1)	(2)	(3)	(4)	(5)	(6)
	Moved to USA	Moved to USA	Moved to USA	Moved to USA	Moved to USA	Moved to USA
Can read and write	-0.012*** (0.004)			-0.011 (0.009)		
Occupation (ref= Farmer):						
White-collar		-0.018*** (0.005)			-0.023*** (0.008)	
Skilled worker		-0.020*** (0.004)			-0.019*** (0.006)	
Urban laborer		-0.018*** (0.005)			-0.016* (0.008)	
Farm laborer		-0.013** (0.005)			0.006 (0.007)	
Occupation (ref= Farmer [low productivity area])						
Farmer [medium productivity area]			-0.011* (0.006)			-0.026*** (0.010)
Farmer [high productivity area]			-0.035*** (0.005)			-0.058*** (0.009)
Laborer [low productivity area]			0.011 (0.018)			0.035* (0.018)
Laborer [medium productivity area]			-0.045*** (0.007)			-0.036*** (0.013)
Laborer [high productivity area]			-0.025*** (0.008)			-0.039*** (0.011)
Constant	0.015*** (0.004)	-0.004 (0.007)	0.020*** (0.004)	0.115*** (0.014)	0.095*** (0.012)	0.136*** (0.015)
Observations	17669	17669	11196	13785	13785	8270
R Squared	0.036	0.038	0.048	0.009	0.010	0.017
Adjusted R Squared	0.036	0.037	0.047	0.008	0.009	0.015
Age FE	Yes	Yes	Yes	Yes	Yes	Yes
Robust SE	Yes	Yes	Yes	Yes	Yes	Yes

Standard errors in parentheses; Significance levels: \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

Notes: The Adult Sample is confined to males aged 15-25 in 1901 and the Child Sample is confined to sons aged 6-18 in 1901. In the Child Sample, occupation and literacy are measured for the fathers of the individuals in the sample from the 1901 census. The measures of productivity are imputed using PLU measures from the land census that were linked to the 1901 census of population. The outcome variable refers to whether an individual moved to the United States relative to all individuals in the sample who stayed in Ireland from 1901-1911. The baseline probabilities of emigration for these samples are: 0.043 (C1, C2), 0.05 (C3), 0.083 (C4, C5), 0.09 (C6).

If intergenerational mobility was high in Ireland in the early twentieth century, fathers' occupation and literacy could be misleading measures of emigrants' characteristics.<sup>30</sup> Thus, Columns 4-6 of Table 2 use the Adult Sample to estimate the probability of emigration by own occupation and literacy. These estimates are also only conditional on age. The estimated negative coefficient for literacy on emigration in Column 4 is similar in magnitude and direction to the

<sup>30</sup> There has been little study of intergenerational mobility in Ireland in the early twentieth century.

coefficient for fathers' literacy (Column 1) but is imprecisely estimated, perhaps due to the relatively small number of illiterate young men in the 1901 census. In the case of occupation, Columns 5 shows farmers (the reference group) to be significantly more likely to emigrate than urban laborers and skilled workers. Within agriculture, Column 6 shows that emigration was also higher from low productivity areas. These estimates suggest that patterns of migrant self-selection are generally similar whether they are measured based on emigrants' own characteristics or those of their fathers. The only notable discrepancy between the samples is that farm laborers were as likely to emigrate as farmers in the Adult Sample, but this likely reflects reduced emigration for men who had acquired land.

These baseline estimates suggest that emigrants were more likely to be the sons of lower educated and farming men in the least productive rural communities. As these places were relatively poor on average, these estimates lead me to conclude that emigrants tended to be drawn from the less educated and poorer tails of the Irish population, with the caveat that they were more likely to grow up in landholding households (albeit poorer ones). These estimates are generally consistent across the Adult and Child Samples. The fact that farmers' and farm laborers' sons were similarly likely to leave low productivity areas indicates that origin communities may have affected emigration decisions above and beyond fathers' occupation.

### ***Measuring selection with control variables***

Occupation is not only correlated with geography, but also a range of other individual and household characteristics. Thus, Table 3 estimates the conditional correlation of occupation and literacy with emigration while controlling for religion, Irish-speaking ability, urban origin and household characteristics.<sup>31</sup> But for the addition of these control variables, the estimates presented in Column 1 (Child Sample) and Column 2 (Adult Sample) were generated from models identical to those presented in Table 2. These models also provide a means of testing whether religion or Irish-speaking ability may have had effects on emigration that were independent of occupation and literacy.

---

<sup>31</sup> I categorized enumeration districts as urban based on whether 20 percent or less of the heads of household are farmers or farm laborers.



TABLE 3. OCCUPATION, LITERACY AND EMIGRATION WITH CONTROL VARIABLES TO THE USA 1901-1911

	Child Sample	Adult Sample	Child Sample	Adult Sample
	(1)	(2)	(3)	(4)
Outcome	Moved to USA	Moved to USA	Moved to USA	Moved to USA
Compared to:	(Stayed in Ireland)		(Moved to Irish urban area)	
Occupation (ref = Farmer)				
White-collar	-0.008 (0.006)	-0.001 (0.009)	-0.112*** (0.023)	-0.121*** (0.026)
Skilled worker	-0.010** (0.005)	0.002 (0.008)	-0.083*** (0.023)	-0.047** (0.024)
Urban laborer	-0.011* (0.005)	-0.001 (0.009)	-0.085*** (0.027)	-0.051* (0.029)
Farm laborer	-0.011** (0.005)	0.005 (0.007)	-0.067*** (0.026)	-0.033 (0.023)
Can read and write	-0.006 (0.004)	0.001 (0.010)	-0.024 (0.019)	0.003 (0.030)
Catholic	0.005* (0.003)	0.027*** (0.005)	0.002 (0.017)	0.043** (0.018)
Speaks Irish	0.019*** (0.006)	0.029*** (0.007)	0.086*** (0.024)	0.085*** (0.021)
Lived in urban area	-0.010** (0.004)	-0.024*** (0.007)	-0.014 (0.021)	-0.075*** (0.021)
Parents present	-	-0.012** (0.005)	-	0.030* (0.016)
Married	-	-0.015* (0.009)	-	-0.020 (0.034)
Constant	0.012** (0.005)	0.091*** (0.015)	0.117*** (0.027)	0.372*** (0.049)
Observations	17669	13785	3120	3644
R <sup>2</sup>	0.042	0.021	0.149	0.080
Adjusted R <sup>2</sup>	0.041	0.020	0.143	0.074
Age FE	Yes	Yes	Yes	Yes
Robust SE	Yes	Yes	Yes	Yes

Standard errors in parentheses; Significance levels: \* p < 0.10, \*\* p < 0.05, \*\*\* p < 0.01

Notes: The Adult Sample is confined to males aged 15-25 in 1901 and the Child Sample is confined to sons aged 6-18 in 1901. In the Child Sample, occupation and literacy are measured for the fathers of the individuals in the sample from the 1901 census. The outcome variable is whether an individual moved to the United States relative to all individuals in the sample who stayed in Ireland. The baseline probabilities of emigration for these samples are: 0.043 (C1), 0.083 (C2), 0.30 (C3), 0.37 (C4).

The estimates in Columns 1 and 2 of Table 3 provide a robustness test of the initial findings regarding selection on occupation and literacy. Although the main finding from the Child Sample – the high emigration rate of farmers’ sons – persists in Column 1 with these additional controls variables, there is a sizeable attenuation in the estimated differences between farmers’ sons and the sons of urban laborers, and skilled and white-collar workers. While the latter were 2 percentage points less likely to emigrate in the initial univariate models (Table 2), they are around 1 percentage point less likely to do so once the conditional controls are added. Further, the

controls account for almost the entire effect of own occupation and literacy on emigration (Column 2).

Much of this attenuation is driven by the added control variable for living in an urban area. Urban residence is associated with a 1 percentage point reduction in the probability of emigration in the Child Sample and more than 2 percentage point reduction from the Adult Sample. This is roughly commensurate with the attenuation in the occupational coefficients between Tables 2 and 3, and holds when urban residence is modeled as the only control variable (see Section 13 of the Online Appendix). Thus, around half of the association between fathers' occupation and emigration, and almost all of own occupation, can be accounted for by urban residence, providing further evidence for the role of origin location in affecting emigration decisions by occupational background.

Despite the attenuation of migrant selection by occupation in these models, Irish-speaking ability and religion are revealed to have a strong correlation with emigration. Depending on whether the Adult or Child sample is used, Catholics are from 0.5-2.7 percentage points more likely to move to the United States than non-Catholics. Likewise, Irish-speakers are around 2-3 percentage points more likely to emigrate. Irish-speakers were heavily concentrated in the Gaeltacht communities across Ireland's emigration-prone west coast. While these data do not allow me to reveal the direct source of the correlation of religion and Irish-speaking with emigration, these models show these correlations to persist even after accounting for occupational background, literacy and urban residence.

### ***Comparison of emigrants to movers to urban areas in Ireland***

High rates of rural emigration could reflect poor conditions in rural areas driving emigration or, conversely, the reluctance of individuals with highly valued skills to leave Irish towns and cities. I examine this by testing for occupational differences between emigrants and men who moved across counties to an urban area over the 1901-1911 period. If occupational differences in emigration simply reflect stronger "push" forces from rural than urban areas, we would expect high rates of rural emigration but weaker occupational selection across destinations. In contrast, if Irish towns and cities provided relatively better opportunities for

highly skilled workers, we would expect greater differentiation in the pre-migration occupations of inter-county movers and emigrants.

I make this comparison in Columns 3 and 4 of Table 3, where I compare the characteristics of emigrants to their counterparts moving to urban areas in Ireland over the same period. Column 3 shows strong differences in the occupational background of emigrants relative to movers within Ireland: the sons of farmers are significantly more likely to choose the United States than to move to Irish towns and cities. The relative probability of moving to urban areas in Ireland over the United States also appears to increase with skill. This is most pronounced for the sons of white-collar workers who are 11-12 percentage points less likely than farmers to move to the United States than to move to an urban area in Ireland. As these models control for urban origin and a range of other characteristics, these destination differences by occupation are even more striking.

The magnitude and robustness of these destination choices suggest that urban areas in Ireland were attracting particularly highly skilled workers. As the occupational differences between emigrants and inter-county movers (Column 3 and 4) are substantially larger than when made to the full population staying in Ireland (Column 1 and 2), this implies that men moving within Ireland were disproportionately drawn from the higher skilled tail of the Irish population. As highly skilled workers were generally less economically constrained in their migration decisions, the most plausible explanation for high skilled inter-county migration is the greater availability of skilled opportunities in Irish towns and cities. Articulated in reverse, the relatively high rate of emigration from lower skilled and poorer farming households likely reflects limited opportunities in home communities and in Irish urban labor markets.

## **COMMUNITY AND HOUSEHOLD COVARIATES AND MIGRANT SELECTION**

The previous analyses highlight that *who* left Ireland depended heavily on *where* they grew-up and which families they were born into. While skill-selective migration within Ireland may reflect the higher domestic demand for skilled labor, *origin* communities may have affected the balance of emigration from farming or laboring households. This section explicitly examines the role of rural conditions and access to migrant networks in accounting for emigrant self-

selection.

Table 4 describes the variation in origin community characteristics that could impact migrant selection by fathers’ occupation. In this table, I present average values of community landholding inequality, land value, and migrant network strength by fathers’ characteristics. The main pattern evident from Table 4 is that on average, the sons of farmers and illiterate men – individuals found to be more likely to emigrate – were more concentrated in unequal communities with lower land values and stronger migrant networks than their counterparts from literate, laboring or highly skilled households. This descriptive table provides a clear indication that fathers’ characteristics are correlated with origin community attributes, which could account for occupational and literacy differences in migration decisions.

TABLE 4. AVERAGE ORIGIN COMMUNITY CHARACTERISTICS BY FATHERS’ OCCUPATION AND LITERACY

Characteristics of father	Child Sample (rural only)		
	Average for origin community (PLU)		
	Inequality of landholding	Land value (per acre)	Network strength
Farmer	0.43	0.48	1.0022
Laborer	0.39	0.66	1.0016
Skilled and white-collar	0.39	0.63	1.0017
Can read and write	0.41	0.56	1.0019
Cannot read and write	0.43	0.45	1.0022

Notes: Table 4 presents the average values of PLU characteristics for sons in the rural Child Sample based on their fathers’ characteristics.

My first objective is to ascertain whether these origin community differences can account for which sons left rural areas to move to the United States. I do this by estimating baseline differences in the probability of emigration by fathers’ literacy and occupation in Column 1 of Table 5, and then individually adding community control variables. Before simultaneously modeling the relationships of landholding inequality, land value and network strength with emigration (Column 5), I estimate each relationship in isolation in Columns 2-4, as sequentially adding these variables may lead to misinterpretation in the correlation of community covariates with emigration (Gelbach 2016).

The estimates presented in Columns 2 through 4 show strong significant coefficients for

each of these community characteristics. Controlling for parents' characteristics, emigration was more likely from low productivity and more unequal communities with stronger migrant networks. Individually, a standard deviation increase in migrant network strength is associated with a 1 percentage point increase in the probability of emigration, while a similar increase in landholding inequality predicts a 0.6 of a percentage point increase (from a base probability of 4.7%). As expected, emigration decreases where land values are higher.

TABLE 5. COMMUNITY CHARACTERISTICS AND RURAL EMIGRATION TO THE USA 1901-1911

Outcome	Child Sample (rural only)				
	(1) Moved to the USA	(2) Moved to the USA	(3) Moved to the USA	(4) Moved to the USA	(5) Moved to the USA
Father's occupation (ref = Farmer)					
Laborer	-0.015*** (0.005)	-0.013*** (0.005)	-0.009* (0.005)	-0.010** (0.005)	-0.008 (0.005)
Skilled & white-collar	-0.009** (0.004)	-0.007* (0.004)	-0.005 (0.004)	-0.006 (0.004)	-0.004 (0.004)
Father can read and write	-0.011** (0.005)	-0.009* (0.005)	-0.007 (0.005)	-0.008* (0.005)	-0.007 (0.005)
<i>Community characteristics</i>					
Inequality of landholding		0.006** (0.002)			0.001 (0.003)
Land value (per acre)			-0.009*** (0.003)		-0.005* (0.003)
Network strength				0.010*** (0.003)	0.007** (0.003)
Constant	0.015*** (0.005)	0.013*** (0.005)	0.010** (0.005)	0.012*** (0.004)	0.010** (0.005)
Observations	13975	13975	13975	13975	13975
R Squared	0.043	0.044	0.045	0.045	0.046
Adjusted R Squared	0.042	0.043	0.044	0.044	0.044
Age FE	Yes	Yes	Yes	Yes	Yes
Robust SE	Yes	Yes	Yes	Yes	Yes
Clustered SE	PLU	PLU	PLU	PLU	PLU

Standard errors in parentheses; Significance levels: \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

Notes: This sample is confined to sons aged 6-18 and living in a rural area in 1901. In this sample, occupation and literacy are measured for the fathers of the individuals in the sample from the 1901 census. The outcome variable is whether an individual moved to the United States relative to all individuals in the sample who stayed in Ireland. I adjust the standard errors to account for the clustering of observations across PLUs (e.g. Moulton, 1990). The baseline probability of emigration for this samples is 0.047 (C1-C5).

When all of the local covariates are included in Column 5, network strength and land value have persisting and similarly sized coefficients. While the landholding inequality coefficient tends toward zero in this saturated model, a standard deviation increase in network strength is

associated with a significant 0.7 percentage point increase, and for land value a 0.5 point decrease, in the probability of emigration. This suggests that the economic productivity and migrant network strength of origin communities had substantial effects on emigration decisions.

These origin community characteristics also account for much of the variation in emigration across occupational categories. In the baseline model in Column 1, the sons of laborers and literate men are 1.5 and 1.1 percentage points less likely to emigrate, respectively. By controlling for the land value and network strength of communities in Columns 3 and 4, the coefficients on occupation attenuate by around a third, and when modeled together, these occupational coefficients almost halve from Column 1 and are no longer significant. These changes imply that much of the difference in the emigration of the sons of farmers and non-farmers stems from the residence of farming households in less productive communities with stronger emigrant networks.

### **Networks**

While landholding and economic conditions characteristics may affect where people emigrated from, the strength of migrant networks may also generate heterogeneity in migrant selection across communities. Access to emigrant networks may help the poor to overcome cost constraints on migration, which could in turn reduce the average skill- and wealth-level of the migration flow (see Belot and Hatton, 2012; McKenzie and Rapoport, 2010; Spitzer and Zimran, 2018; Wegge, 2008). In Table 6, I test these propositions across origin communities by interacting the key economic indicators of selection with the name-based measure of network strength for the rural subsample.

Columns 1 and 2 begin by testing for heterogeneity in selection by interacting fathers' literacy and occupation with network strength. In both cases, network strength has a strong positive main effect, suggesting that networks were associated with higher rates of emigration. There are, however, also strong negative interactions for the sons of men who can read and write or who hold highly skilled occupations, suggesting that the effect of networks in promoting emigration was relatively weaker for sons from more literate and non-farming households. That is, the sons of less skilled men are relatively *more likely* to leave communities with stronger

migrant networks.

TABLE 6. NETWORK EFFECTS AND EMIGRATION FROM RURAL AREAS TO THE USA 1901-1911

Outcome	Child Sample (rural only)				
	(1) Moved USA	(2) Moved USA	(3) Moved USA	(4) Moved USA	(5) Moved USA
<i>Father's occupation (ref = Farmer):</i>					
Laborer	-0.007 (0.005)	-0.008 (0.005)	-0.006 (0.005)	-0.006 (0.005)	-0.003 (0.006)
Skilled & white-collar	-0.004 (0.004)	-0.005 (0.005)	-0.005 (0.004)	-0.005 (0.005)	-0.003 (0.004)
Father can read and write	-0.005 (0.004)	-0.006 (0.005)	-0.005 (0.005)	-0.004 (0.004)	-0.003 (0.004)
<i>Community characteristics</i>					
Inequality of landholding	0.002 (0.003)	0.002 (0.003)	0.005* (0.003)	0.005* (0.003)	0.005* (0.003)
Land value (per acre)	-0.006* (0.003)	-0.006* (0.003)	-0.010*** (0.003)	-0.010*** (0.003)	-0.008*** (0.003)
Network strength	0.015*** (0.004)	0.009*** (0.003)	0.002 (0.003)	0.011** (0.004)	0.010** (0.005)
<i>Interactions</i>					
Network strength x father can read and write	-0.011** (0.004)			-0.009** (0.004)	-0.009** (0.004)
Network strength x laborer		-0.006 (0.005)		-0.004 (0.005)	-0.003 (0.005)
Network strength x skilled & white-collar		-0.008* (0.004)		-0.004 (0.004)	-0.004 (0.004)
Network strength x land value (per acre)			-0.008*** (0.003)	-0.007*** (0.003)	-0.006** (0.003)
Constant	0.008* (0.004)	0.009* (0.004)	0.003 (0.005)	0.002 (0.005)	-0.017 (0.017)
Observations	13975	13975	13975	13975	13975
R Squared	0.046	0.046	0.047	0.047	0.054
Adjusted R Squared	0.045	0.045	0.045	0.046	0.047
Age FE	Yes	Yes	Yes	Yes	Yes
Controls	-	-	-	-	Birth order Family size Religion Irish-speaker Father Age
Robust SE	Yes	Yes	Yes	Yes	Yes
Clustered SE	PLU	PLU	PLU	PLU	PLU

Standard errors in parentheses; Significance levels: \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

Notes: This sample is confined to sons aged 6-18 and living in a rural area in 1901. In this sample, occupation and literacy are measured for the fathers of the individuals in the sample from the 1901 census. The outcome variable is whether an individual moved to the United States relative to all individuals in the sample who stayed in Ireland. I adjust the standard errors to account for the clustering of observations across PLUs (e.g. Moulton, 1990). The baseline probabilities of emigration for this sample is 0.047 (C1-C5).

Does the variation in migrant selection with network strength reflect the alleviation of cost constraints? Column 3 tests this interpretation by interacting land value with network strength to check whether networks are associated with heightened emigration from less productive (or poorer) communities. In this regression, the main effect for land value is negative, as is the interaction between land value and network strength, implying that emigration is lower from more productive communities, with the effect of networks on emigration being relatively weaker in these contexts. Thus, consistent with hypotheses that networks eased cost constraints, networks appear to matter most for emigration from communities with fewer economic resources. These results strengthen in more saturated models (Column 4) and are robust to a suite of control variables (Column 5).

### ***Birth order and land***

The high rate of emigration among the sons of landholding farmers contrasts with historical findings from Norway showing that land acquisition opportunities discourage migration ((Abramitzky, Boustan, and Eriksson 2013). Due to traditional customs and the imposition of different inheritance laws on Catholics and non-Catholics, Ireland has a complicated history with respect to inheritance.<sup>32</sup> We know that while Ireland did not have a rigid system of primogeniture (first-born inheritance) in the early twentieth century, it was not typical to heavily subdivide land among heirs, with middle sons also being less likely to be the sole inheritors (Guinnane 1997; Ó Gráda 1980). I exploit these inheritance practices as a means of understanding whether land inheritance – as implied by birth order – affected emigration decisions.

In Table 7, I test the association between emigration and birth order across households headed by farmers (Column 1 and 2) and landless laborers (Column 3 and 4). The estimates in Column 1 show middle sons in farming households to be 2.8 percentage points more likely to emigrate than the only sons of farmers. The middle sons of farmers are also up to 1 percentage

---

<sup>32</sup> Prior to the nineteenth century, the Penal Laws and traditional Irish law enforced forms of partible inheritance in how land was apportioned among sons in Catholic families. Although impartible inheritance or “sole-heirs” was practiced in post-Famine Ireland, Ireland did not uniformly conform to a strict line of succession (Arensberg and Kimball 1940).



point more likely to emigrate than their first- and last-born counterparts, suggesting that sons less likely to inherit land were more likely to emigrate.<sup>33</sup> There may be concern that emigration and birth order are correlated because families were larger in emigration-prone areas (Fernihough 2017). Column 2, however, shows these birth order correlations to be robust to family size and other controls.

TABLE 7. BIRTH ORDER IN FARMING AND LABORING HOUSEHOLDS AND EMIGRATION TO THE USA 1901-1911

Outcome	Child Sample (rural only)			
	Father is a farmer		Father is a laborer	
	(1) Moved USA	(2) Moved USA	(3) Moved USA	(4) Moved USA
Birth order (ref = Only son):				
First son	0.017* (0.009)	0.012 (0.009)	-0.014 (0.013)	-0.009 (0.014)
Last son	0.021** (0.009)	0.016* (0.009)	-0.013 (0.014)	-0.011 (0.014)
Middle son	0.028*** (0.009)	0.020** (0.010)	-0.005 (0.013)	0.002 (0.015)
Family size		-0.002 (0.002)		0.001 (0.001)
Constant	-0.018* (0.011)	-0.031** (0.012)	0.008 (0.016)	0.012 (0.019)
Observations	9578	9578	2293	2293
R Squared	0.047	0.052	0.040	0.042
Adjusted R Squared	0.046	0.050	0.033	0.033
Age FE	Yes	Yes	Yes	Yes
Controls	-	Fathers' Age PLU controls	-	Fathers' Age PLU controls
Robust SE	Yes	Yes	Yes	Yes

Standard errors in parentheses; Significance levels: \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

Notes: This sample is confined to sons aged 6-18 and living in a rural area in 1901. The sample is split between households headed by laboring or farming fathers. The measures of birth order and family size are measured by examining the sibling composition of the households in the 1901 census. The PLU controls refer to the measures of network strength, land value and inequality in landholdings used above. The outcome variable is whether an individual moved to the United States relative to all individuals in the sample who stayed in Ireland. The baseline probabilities of emigration for these samples are 0.052 (C1-C2) and 0.035 (C3-C4).

The study of birth order and emigration within rural laboring families provides further insight. Column 3 shows no significant birth order association with emigration within households

<sup>33</sup> Birth order could also affect emigration decisions through pathways such as expectations as to which children would care for elderly parents.

headed by landless laborers, and this result persists after adding control variables (Column 4). The most likely explanation for the absence of birth order effects in laboring households is that in these families, birth order position did not seriously impact the probability of a son acquiring land through inheritance. Thus, the results from both farming and laboring families indicate that it was the *non-inheriting sons* of farmers who were most likely to emigrate. My conclusions, therefore, support prior work showing that land acquisition opportunities discouraged migration.

### **INTERPRETTING MIGRANT SELECTION FROM IRELAND**

How do these findings from Ireland square with theoretical models of selection? The dominant economic selection framework is Borjas' (1987) interpretation of the Roy model (Roy 1951), which can generate predictions of selection by comparing earnings inequality across countries. These models assume that migrants have two choices – stay in the home country or move abroad – and they then compare the relative returns to skill in each place, generating predictions of either negative or positive selection from the source population. Recent historical findings of negative selection into emigration from Norway and positive selection from Italy are consistent with these predictions (Abramitzky, Boustan and Eriksson, 2012; Spitzer and Zimran, 2018).

Although my data do not allow for a strong test of the Roy model, they do provide valuable theoretical insights. As the Irish economy was more unequal in how it rewarded skill than that of the United States (Begley, Geary, and Stark 2014; Boyer, Hatton, and O'Rourke 1994), the Roy model would predict emigrants to be lower skilled than non-migrants. This hypothesis is generally supported by the finding that emigrants and their fathers were more likely to be illiterate and lower skilled than their counterparts staying in Ireland. Ireland's retention of higher skilled workers is also consistent with the high demand for skill in Irish towns and cities (O'Malley 1981). Thus, the negative selection prediction from the Roy model is tentatively supported by these findings from Ireland.

What is perhaps most notable about Irish emigration, however, is that it does not smoothly conform to an interpretation of positive or negative selection. Emigration drew heavily from poorer agricultural communities in the west of Ireland, where illiteracy was higher and

emigrant networks were stronger, but much less so from urban areas and eastern counties which also had large laboring populations. These western communities had more pronounced recent histories of emigration to the United States, and these experiences are partly captured by the network variable. Overall, these findings suggest that regional and community differences played a key role in influencing who emigrated, and whether migrants were of farming origin. It is also plausible that emigration from eastern counties was directed more toward England, and this flow is not captured in this analysis. Overall, the Irish case supports predictions of migrant selection based on economic theory, but also underscores the importance of geography and community context in affecting migration decisions.

## **CONCLUSION**

The emigration of more than four million people from Ireland transformed the economic and social fabric of the United States. While Irish emigration fed the growing labor markets of the New World, Irish government leaders and officials frequently claimed that the economic development of Ireland was being curtailed by the emigration of highly skilled men. Focusing on the tail end of mass Irish emigration across the Atlantic, I undertake the first micro-analysis of how Irish emigrants in the early twentieth century compared to the people they left behind. My main finding – that emigrants were drawn from households headed by farmers and less literate men in low productivity agricultural communities – is, therefore, at odds with claims that Ireland was losing its “best men” to emigration (Royal Commission on Labour 1893 p. 49).

My findings underscore the role of origin communities in shaping national-scale patterns of migrant selection. Building on recent work in economic history showing differences in self-selection across migrant-sending regions (Spitzer and Zimran 2018), I find that by influencing the migration decisions of individuals with different attributes, origin communities shaped the composition of the emigrant flow. While emigrants were most likely to leave low productivity farming communities with strong emigrant networks, opportunities in Ireland also reduced high-skilled emigration. Thus, while wage differences between Ireland and the United States undoubtedly affected emigration decisions, one cannot downplay origin community conditions in affecting who decided to leave the island.

These findings have implications for both Ireland's economic development and Irish assimilation in the United States. First, my finding that lower skilled men were overrepresented in the emigration flow supports claims that mass emigration boosted average real wages in post-famine Ireland (Begley, Geary, and Stark 2014; O'Rourke and Williamson 1997). If this is correct, emigration may have been a veritable win-win for the economies of Ireland and the United States. With this said, persistently high rates of lower skilled Irish immigration also likely increased competition and depressed wages for unskilled workers in the United States. Second, it is claimed that the personal characteristics brought by Irish immigrants (and others) to the United States shaped their subsequent economic outcomes (Borjas 1994, 2014). While Irish Americans did experience high rates of upward mobility in the early twentieth century, my results indicate that the Irish were not likely to have been highly advantaged in their backgrounds.<sup>34</sup> Thus, my findings indirectly contribute to growing evidence that factors other than pre-migration skill drive post-migration outcomes (e.g. Alexander and Ward, 2018; Catron, 2017; Connor, 2017a; Ferrie, 1994).

Finally, this study illustrates the new possibilities for studying migrant selection in historical context. As concerns around the quantity and quality of immigrants continue to gain salience in the current political climate, we still have much to learn about how individuals self-select into emigration and of the interacting components affecting patterns of migrant selection within and across countries. Although there are major constraints on breaking ground in these areas for contemporary migration flows, rapid innovations in historical data and computation provide a handsome opportunity to understand these key economic and social science issues.

---

<sup>34</sup> Although Irish immigrants did possess the advantage of being native English speakers, emerging evidence suggests that the returns to English were relatively low in the early twentieth century (Ward 2018).

## REFERENCES

- Abramitzky, Ran, Leah Boustan, and Katherine Eriksson. 2012. "Europe's Tired, Poor, Huddled Masses: Self-Selection and Economic Outcomes in the Age of Mass Migration." *American Economic Review* 102 (5): 1832–56.
- . 2013. "Have the Poor Always Been Less Likely to Migrate? Evidence from Inheritance Practices during the Age of Mass Migration." *Journal of Development Economics* 102: 2–14.
- Abramitzky, Ran, Roy Mill, and Santiago Pérez. 2018. "Linking Individuals Across Historical Sources: A Fully Automated Approach." National Bureau of Economic Research.
- Ager, Philipp, and Markus Brückner. 2013. "Cultural Diversity and Economic Growth: Evidence from the US during the Age of Mass Migration." *European Economic Review* 64: 76–97.
- Alexander, Rohan, and Zachary Ward. 2018. "Age at Arrival and Assimilation during the Age of Mass Migration." *Journal of Economic History*.
- Anbinder, Tyler, and Hope McCaffrey. 2015. "Which Irish Men and Women Immigrated to the United States during the Great Famine Migration of 1846–54?" *Irish Historical Studies* 39 (156): 620–642.
- Arensberg, Conrad Maynard, and Solon Toothaker Kimball. 1940. *Family and Community in Ireland*. Harvard University Press.
- Bailey, Martha, Connor Cole, Morgan Henderson, and Catherine Massey. 2017. "How Well Do Automated Methods Perform in Historical Samples? Evidence from New Ground Truth." National Bureau of Economic Research.
- Begley, Jason, Frank Geary, and Tom Stark. 2014. "Convergence in the Pre-1914 Atlantic Economy: What Really Happened to Wages in Ireland between 1881 and 1911?" *Cambridge Journal of Economics*, beu063.
- Belot, Michèle VK, and Timothy J. Hatton. 2012. "Immigrant Selection in the OECD\*." *The Scandinavian Journal of Economics* 114 (4): 1105–1128.
- Borchers, Allison, Jennifer Ifft, and Todd Kuethe. 2014. "Linking the Price of Agricultural Land to Use Values and Amenities." *American Journal of Agricultural Economics* 96 (5): 1307–1320.
- Borjas, George J. 1987. "Self-Selection and the Earnings of Immigrants." *The American Economic Review* 77 (4): 531–53.
- . 1994. "Assimilation and Changes in Cohort Quality Revisited: What Happened to Immigrant Earnings in the 1980s?" National Bureau of Economic Research. <http://www.nber.org/papers/w4866>.
- . 2014. *Immigration Economics*. Harvard University Press.
- Boustan, Leah. 2016. *Competition in the Promised Land: Black Migrants in Northern Cities and Labor Markets*. Princeton University Press.
- Boustan, Leah, and Ran Abramitzky. 2016. "Immigration in American Economic History." *Journal of Economic Literature*.
- Boyer, George R., Timothy J. Hatton, and Kevin O'Rourke. 1994. "The Impact of Emigration on Real Wages in Ireland." *Migration and the International Labor Market, 1850-1939*, 221.
- Breathnach, Ciara. 2005. *The Congested Districts Board of Ireland, 1891-1923: Poverty and Development in the West of Ireland*. Portland, OR.

- Burt, Oscar R. 1986. "Econometric Modeling of the Capitalization Formula for Farmland Prices." *American Journal of Agricultural Economics* 68 (1): 10–26.
- Carter, C. F., G. A. Duncan, Donal Nevin, and Liam O Buachalla. 1956. "Symposium on the Report of the Commission on Emigration and Other Population Problems." *The Statistical and Social Inquiry Society of Ireland*.
- Casey, Marion, and J. J. Lee. 2007. *Making the Irish American: History and Heritage of the Irish in the United States*. New York: NYU Press.
- Catron, Peter. 2017. "The Citizenship Advantage: Immigrant Socioeconomic Attainment across Generations in the Age of Mass Migration." *UCLA Center for the Study of International Migration Working Papers*.
- Collins, William J., and Marianne H. Wanamaker. 2014. "Selection and Economic Gains in the Great Migration of African Americans: New Evidence from Linked Census Data." *American Economic Journal: Applied Economics* 6 (1): 220–52.
- . 2015. "The Great Migration in Black and White: New Evidence on the Selection and Sorting of Southern Migrants." *The Journal of Economic History* 75 (04): 947–992.
- Connor, Dylan Shane. 2017a. "Is Immigrant Social Mobility Driven by the People or the Place? The Case of Irish Americans in the Early Twentieth Century." *UCLA Center for the Study of International Migration Working Papers*.  
<http://escholarship.org/uc/item/5b79427m.pdf>.
- . 2017b. "Poverty, Religious Differences and Child Mortality in the Early 20th Century: The Case of Dublin." *Annals of the American Association of Geographers*.
- Connor, Dylan Shane, Gerald Mills, and Niamh Moore-Cherry. 2011. "The 1911 Census and Dublin City: A Spatial Analysis." *Irish Geography* 44 (2–3): 245–263.
- Crossman, Virginia. 2013. *Poverty and the Poor Law in Ireland 1850-1914*. Liverpool University Press.
- "Emigration Statistics of Ireland for the Year 1902." 1903. Cd. 1489. Dublin.  
<http://www.dippam.ac.uk/eppi/documents/20492/page/566483>.
- Featherstone, Allen M., and Timothy G. Baker. 1987. "An Examination of Farm Sector Real Asset Dynamics: 1910–85." *American Journal of Agricultural Economics* 69 (3): 532–546.
- Feigenbaum, James J. 2016a. "Automated Census Record Linking: A Machine Learning Approach." <https://jamesfeigenbaum.github.io/research/pdf/census-link-ml.pdf>.
- . 2016b. "JAROWINKLER: Stata Module to Calculate the Jaro-Winkler Distance between Strings." *Statistical Software Components*.
- Fernández-Huertas Moraga, Jesús. 2011. "New Evidence on Emigrant Selection." *The Review of Economics and Statistics* 93 (1): 72–96.
- Fernihough, Alan. 2017. "Human Capital and the Quantity–Quality Trade-off during the Demographic Transition." *Journal of Economic Growth* 22 (1): 35–65.
- Fernihough, Alan, Cormac Ó Gráda, and Brendan M. Walsh. 2015. "Intermarriage in a Divided Society: Ireland a Century Ago." *Explorations in Economic History* 56: 1–14.
- Ferrie, Joseph P. 1994. "The Wealth Accumulation of Antebellum European Immigrants to the US, 1840–60." *The Journal of Economic History* 54 (1): 1–33.
- . 1996. "A New Sample of Males Linked from the Public Use Microdata Sample of the 1850 US Federal Census of Population to the 1860 US Federal Census Manuscript Schedules." *Historical Methods: A Journal of Quantitative and Interdisciplinary History*

- 29 (4): 141–156.
- Fitzpatrick, David. 1980. "Irish Emigration in the Later Nineteenth Century." *Irish Historical Studies* 22 (86): 126–143.
- . 1984. *Irish Emigration 1801-1921*. Vol. 1. Economic and Social History Society of Ireland.
- . 1994. *Oceans of Consolation: Personal Accounts of Irish Migration to Australia*. Cork: Cork University Press.
- Gelbach, Jonah B. 2016. "When Do Covariates Matter? And Which Ones, and How Much?" *Journal of Labor Economics* 34 (2): 509–543.
- Grogger, Jeffrey, and Gordon H. Hanson. 2011. "Income Maximization and the Selection and Sorting of International Migrants." *Journal of Development Economics*, Symposium on Globalization and Brain Drain, 95 (1): 42–57.  
<https://doi.org/10.1016/j.jdeveco.2010.06.003>.
- Guinnane, Timothy. 1997. *The Vanishing Irish: Households, Migration, and the Rural Economy in Ireland, 1850-1914*. Princeton University Press.
- Harkness, D. A. 1931. "Irish Emigration." In *International Migrations, Volume II: Interpretations*, 261–282. NBER.
- Hatton, Timothy J., and Jeffrey G. Williamson. 1993. "After the Famine: Emigration from Ireland, 1850–1913." *The Journal of Economic History* 53 (03): 575–600.
- Kenny, Kevin. 2014. *The American Irish: A History*. Oxon: Routledge.
- 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 (04): 1015–1044.
- Leeuwen, Marco van, and Ineke Maas. 2011. *Hisclass: A Historical International Social Class Scheme*. Universitaire Pers Leuven.
- Massey, Catherine G. 2017. "Playing with Matches: An Assessment of Accuracy in Linked Historical Data." *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 1–15.
- Matheson, Robert. 1901. *Special Report on Surnames in Ireland: Together with Varieties and Synonymes of Surnames and Christian Names in Ireland*. Dublin: H.M. Stationery Office.
- . 1908. "A Review of the General Topographical Index of Ireland 1901." *Journal of the Statistical and Social Inquiry Society of Ireland* XII (LXXXIX): 263–94.
- 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.
- Meagher, Timothy J. 2009. "From the World to the Village and the Beginning to the End and after: Research Opportunities in Irish American History." *Journal of American Ethnic History*, 118–135.
- Miller, Kerby A. 1988. *Emigrants and Exiles: Ireland and the Irish Exodus to North America*. Oxford University Press, USA.
- Mokyr, Joel, and Cormac Ó Gráda. 1982. "Emigration and Poverty in Prefamine Ireland." *Explorations in Economic History* 19 (4): 360–84.
- Moulton, Brent R. 1990. "An Illustration of a Pitfall in Estimating the Effects of Aggregate Variables on Micro Units." *The Review of Economics and Statistics*, 334–338.
- Munshi, Kaivan. 2003. "Networks in the Modern Economy: Mexican Migrants in the US Labor

- Market." *The Quarterly Journal of Economics* 118 (2): 549–599.
- Ó Gráda, Cormac. 1975. "A Note on Nineteenth-Century Irish Emigration Statistics." *Population Studies* 29 (1): 143–149.
- . 1980. "Primogeniture and Ultimogeniture in Rural Ireland." *Journal of Interdisciplinary History* 10 (3): 491.
- . 2006. "The 1901 and 1911 Irish Censuses: A Unique Source for the Historian." *Library and Archives Canada*. <https://www.collectionscanada.gc.ca/ireland/033001-150-e.php?uid=033001-033001-1402.1-e&uidc=recKey>.
- O'Malley, Eoin. 1981. "The Decline of Irish Industry in the Nineteenth Century." *Economic and Social Review* 13 (1): 21.
- O'Rourke, Kevin. 1995. "Emigration and Living Standards in Ireland since the Famine." *Journal of Population Economics* 8 (4): 407–421.
- O'Rourke, Kevin, and Jeffrey G. Williamson. 1997. "Around the European Periphery 1870–1913: Globalization, Schooling and Growth." *European Review of Economic History* 1 (02): 153–190.
- Rodríguez-Pose, Andrés, and Viola von Berlepsch. 2014. "When Migrants Rule: The Legacy of Mass Migration on Economic Development in the United States." *Annals of the Association of American Geographers* 104 (3): 628–51.
- Roy, Andrew Donald. 1951. "Some Thoughts on the Distribution of Earnings." *Oxford Economic Papers* 3 (2): 135–146.
- Royal Commission on Labour. 1893. *The Agricultural Labourer, Assistant Commissioners' Reports on the Agricultural Labourer*. Vol. IV. London: H.M. Stationery Office.
- Ruggles, Steven, Catherine Fitch, and Evan Roberts. 2017. "Historical Census Record Linkage." *Minnesota Population Center Working Paper No. 2017-3*.
- Ruggles, Steven, Katie Genadek, Ronald Goeken, Josiah Grover, and Matthew Sobek. 2017. *Integrated Public Use Microdata Series: Version 7.0 [Machine-Readable Database]*. Minneapolis: University of Minnesota. <https://doi.org/10.18128/D010.V7.0>.
- 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*.
- Turner, Michael. 2002. *After the Famine: Irish Agriculture, 1850-1914*. Cambridge University Press.
- Ward, Zachary. 2018. "Have Language Skills Always Been so Valuable? The Low Return to English Fluency during the Age of Mass Migration." <https://drive.google.com/file/d/0B77COWld2HzSMVJvdjI5MmRiajA/view>.
- Wegge, Simone A. 1998. "Chain Migration and Information Networks: Evidence from Nineteenth-Century Hesse-Cassel." *The Journal of Economic History* 58 (4): 957–986.
- . 2008. "Network Strategies of Nineteenth Century Hesse-Cassel Emigrants." *The History of the Family* 13 (3): 296–314.
- Wegge, Simone A., Tyler Anbinder, and Cormac Ó Gráda. 2017. "Immigrants and Savers: A Rich New Database on the Irish in 1850s New York." *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 1–12.
- Wolf, Nicholas M. 2014. *An Irish-Speaking Island: State, Religion, Community, and the Linguistic Landscape in Ireland, 1770–1870*. Madison: University of Wisconsin Press.